



## **Masterarbeit**

### **Generative AI in Techno Music Production: Evaluating Audio Content and Sound Design**

Zur Erlangung des akademischen Grades

Master of Engineering

im Studiengang  
Audiovisuelle Medien

vorgelegt von

Roland Ernst  
Matr.-Nr.: 42677

am 16. August 2024  
an der Hochschule der Medien Stuttgart

Erstprüfer/in: Prof. Dr. Andreas Koch  
Zweitprüfer/in: Prof. Oliver Curdt

## Eidesstattliche Erklärung

Hiermit versichere ich, Roland Ernst, ehrenwörtlich, dass ich die vorliegende Masterarbeit mit dem Titel "Generative AI in Techno Music Production: Evaluating Audio Content and Sound Design" selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Ebenso sind alle Stellen, die mit Hilfe eines KI-basierten Schreibwerkzeugs erstellt oder überarbeitet wurden, kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§ 23 Abs. 2 Master-SPO (Vollzeit)) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

Ort, Datum: 16.08.2024

Unterschrift:

A handwritten signature in black ink, appearing to read 'Roland Ernst', with a long horizontal stroke extending to the right.

### Hilfsmittel

Die KI-basierten Werkzeuge "Chat-GPT" und "DeepL Translator" wurden eingesetzt, um bei sprachlichen Formulierungen zu unterstützen. Da sprachliche Anpassungen früh im iterativen Schreibprozess passiert sind, können sie nicht mehr auf einzelne Stellen zurückgeführt werden. Der Hinweis gilt daher für den gesamten Text. KI-basierte Werkzeuge wurden ausdrücklich nicht für Textgenerierung oder zu Recherchezwecken verwendet.

## Kurzfassung

Künstliche Intelligenz (KI) hat in den vergangenen Jahren Einzug in diverse Bereiche des menschlichen Lebens gehalten, darunter auch die Musikproduktion. Dabei ist Techno ein Genre, das aus dem Fortschritt der Technologie geboren wurde und sich dadurch stetig weiterentwickelt hat. Diese Demokratisierung von Musikproduktion wird nun durch generative KI-Modelle expandiert.

In dieser Arbeit werden die aktuellen Möglichkeiten der künstlichen Intelligenz im Bereich der Musikproduktion anhand des Genres Techno untersucht. Hierbei liegt der Fokus auf der Analyse der generierten Audioinhalte durch KI und deren Potenzial in praktischen Anwendungsfällen. Dazu werden nach eingängiger Literaturrecherche Experteninterviews durchgeführt und aus den daraus gewonnenen Informationen Kriterien zur Evaluation von KI-Applikationen extrahiert. Dafür werden die modernsten und fortschrittlichsten Modelle und Applikationen zur Generierung von Musik und Klängen ausführlich getestet und bewertet. Ein weiterer Fokus liegt auf der Replikation von Sound Design durch darauf spezialisierte KI-Anwendungen und deren Vergleich zu den Originalen.

Durch theoretische Analysen und praktische Anwendung leistet diese Forschungsarbeit einen Beitrag zur sich stetig weiterentwickelnden Landschaft der KI in der Musikindustrie und bietet zusammenfassend eine Bewertung, Einblicke und Empfehlungen für zukünftige Entwicklungen in diesem interdisziplinären Bereich.

**Schlagwörter:** Künstliche Intelligenz, Musikproduktion, elektronische Musik

## Abstract

Artificial Intelligence (AI) has made its way into various aspects of human life in recent years, including music production. Techno, a genre born from technological progress, has continuously evolved alongside it. This democratisation of music production is now being expanded by generative AI models.

This work investigates the current capabilities of artificial intelligence in music production, focusing on the genre of techno. The analysis centres on the generated audio content by AI and its potential in practical applications. Following a comprehensive literature review, expert interviews are conducted, from which criteria for evaluating AI applications are extracted. The most advanced models and applications for music and sound generation are extensively tested and evaluated. Another focus lies on the replication of sound design by specialised AI applications and the comparison of their results to the source.

Through theoretical analyses and practical application, this research contributes to the ever-evolving landscape of AI in the music industry, offering an assessment, insights, and recommendations for future developments in this interdisciplinary field.

**Keywords:** Artificial Intelligence, Music Production, Electronic Music

# Table of Contents

Eidesstattliche Erklärung.....	II
Kurzfassung.....	III
Abstract.....	III
Table of Contents .....	IV
List of Figures.....	VII
List of Tables.....	VIII
List of Abbreviations .....	IX
Introduction.....	1
<b>1 Goals and Methodology .....</b>	<b>3</b>
<b>2 Current State of Technology and Foundations.....</b>	<b>4</b>
2.1 Artificial Intelligence .....	4
2.1.1 Definition Artificial Intelligence .....	4
2.1.2 Development and Current State.....	5
2.2 Techno and Production .....	6
2.2.1 Techno .....	6
2.2.2 Production of Techno.....	6
2.2.3 Sampling .....	7
2.2.4 Drums.....	7
2.2.5 Synthesisers.....	8
2.3 AI in the Music Industry.....	8
<b>3 Expert Interviews .....</b>	<b>10</b>
3.1 Qualitative Research.....	10
3.1.1 Approach with Expert Interviewees.....	10
3.1.2 Data Collection .....	11
3.1.3 Questionnaire .....	11
3.1.4 Conducting the Interviews.....	14
3.1.5 Interpretation .....	14
3.1.6 Interview Evaluation.....	17
<b>4 Generative AI Applications .....</b>	<b>20</b>
4.1 Prompting Methodology .....	20
4.1.1 Prompts .....	20
4.1.2 Evaluation Methodology .....	21
4.2 Udio.....	22
4.2.1 Testing.....	24
4.2.2 Evaluation.....	25

---

4.3	Suno .....	26
4.3.1	Testing.....	28
4.3.2	Evaluation.....	32
4.4	MusicFX.....	32
4.4.1	Testing.....	33
4.4.2	Evaluation.....	34
4.5	Riffusion.....	35
4.5.1	Testing.....	36
4.5.2	Evaluation.....	36
4.6	SoundVerse AI.....	37
4.6.1	Testing.....	38
4.6.2	Evaluation.....	40
4.7	Stable Audio .....	41
4.7.1	Testing.....	42
4.7.2	Evaluation.....	43
4.8	MusicGen/LimeWire Studio AI .....	44
4.8.1	Testing.....	44
4.8.2	Evaluation.....	45
4.9	Eleven Labs .....	46
4.9.1	Testing.....	47
4.9.2	Evaluation.....	54
4.10	Other Applications.....	55
4.10.1	Audiobox .....	55
4.10.2	Sonauto AI.....	55
4.10.3	Rightsify.....	55
4.10.4	Snowpixel AI.....	56
4.11	Mubert .....	56
4.11.1	AIVA .....	56
<b>5</b>	<b>Sound Replicating Plug-Ins.....</b>	<b>57</b>
5.1	Synplant 2.....	57
5.1.1	Evaluation.....	60
5.2	MicroMusic.....	60
5.2.1	Evaluation.....	61
<b>6</b>	<b>Application of AI in Music Production.....</b>	<b>63</b>
6.1	Replicating Presets.....	63
6.1.1	Acid Sound with Delay.....	63
6.1.2	Acid Sound Classic, Squelched .....	64
6.1.3	Classic Warm Synth Pad .....	64
6.1.4	Ominous Pad.....	65
6.1.5	Dystopian Horn Sound.....	65
6.1.6	Electric Pad .....	66
6.1.7	Open Filter Bass Pluck .....	66

---

6.1.8	Dissonant Screech Sound .....	67
6.1.9	Dark Plucky Stab .....	67
6.1.10	Complex Space FX.....	68
6.1.11	Dissonant Synth Stab .....	68
6.1.12	Summary / Evaluation.....	69
6.2	Replacing Samples with AI-Generated Samples.....	70
6.2.1	Evaluation.....	71
6.2.2	Conclusion.....	72
6.3	Producing with AI-Generated Samples .....	73
6.3.1	Production 1 .....	73
6.3.2	Production 2 .....	74
<b>7</b>	<b>Conclusion and Recommendation .....</b>	<b>76</b>
	<b>References .....</b>	<b>79</b>
	<b>Appendix .....</b>	<b>84</b>
	Digital Appendix .....	84
	Transcription Expert Interviews .....	85
	Interview 1.....	85
	Interview 2.....	98
	Interview 3.....	111

## List of Figures

Figure 1: Udio Prompt Interface .....	22
Figure 2: Created Clip and Description .....	23
Figure 3: Custom Prompt Menu in Suno .....	27
Figure 4: A Generation of Clips .....	28
Figure 5: Music FX Prompt Menu.....	33
Figure 6: Riffusion Prompt Menu.....	35
Figure 7: SoundVerse Interface .....	37
Figure 8: SoundVerse Editor Studio.....	38
Figure 9: SoundVerse Chatbot not Responding .....	40
Figure 10: Stable Audio Interface.....	41
Figure 11: LimeWire Studio AI Chat Window .....	44
Figure 12: ElevenLabs Speech Synthesis Menu .....	46
Figure 13: ElevenLabs Sound Effects Prompt Menu .....	47
Figure 14: Synplant 2 Interface .....	58
Figure 15: Genopatch .....	59
Figure 16: MicroMusic Menu.....	61
Figure 17: Visual Comparison between Space FX Samples .....	69
Figure 18: Comparison Kick Samples; above Original, below Synplant Replica...	73

## List of Tables

Table 1: Interview Questionnaire .....	14
Table 2: Data Analysis Interviews .....	17



## List of Abbreviations

ADSR	attack decay sustain release
AGI	artificial general intelligence
AI	artificial intelligence
AIFF	audio interchange file format
API	application programming interface
BPM	beats per minute
GPT	generative pre-trained transformer
CPU	central processing unit
DAW	digital audio workstation
DIY	do-it-yourself
e.g.	exempli gratia
FM	frequency modulation
i.e.	id est
Kbps	kilobytes per second
LFO	low-frequency oscillation
LLM	large language model
LUFS	loudness units full scale
MIDI	musical instrument digital interface
SACEM	society of authors, composers and publishers of music
VPN	virtual private network
VST	virtual studio technology
WAV	waveform audio file format

## Introduction

*I dream of instruments obedient to my thought and which with their contribution of a whole new world of unsuspected sounds, will lend themselves to the exigencies of my inner rhythm.*

—Edgard Varèse (A Brief History of Computer Music (1994) – Eric Kuehnl 2024.000Z)

Technology is ubiquitous and has consistently been a driving force in the evolution of music, particularly over the last century. It has altered how music is transmitted, preserved, heard, performed, and composed (Hua 2018). The journey from drums and handmade string instruments, meticulously crafted by specialists, to the industrialisation and mass production of any instrument has significantly lowered the entry-barrier for music creation. Music, once the domain of trained musicians, that began by transcribing music by hand, can now be created by virtually anyone with access to a computer and a few minutes to spare (Zantal-wiener 2017). Technically, the requirements for making music are lower than ever. With the advent of digital audio workstations (DAWs) and plug-ins, physical instruments are no longer a necessity to record and mix high quality audio in home studios. Through this, a wide variety of production styles and musical genres have emerged through the limitations of the systems they were made with (Venutti 2024). Today, even these digital tools may become optional, as it is now possible to produce sounds and music by simply typing text into the menu of an artificial intelligence application.

Arguably, technology has shaped every piece of music we hear today. This evolution began with the capability to access music at any time through recordings and has progressed over the creation of fully electronically generated genres to the creation of new pieces through mere textual descriptions. When we hear a piano, it's increasingly uncertain whether the sound is taken from a physical recording, a sampler, a synthesiser, or, as of recently, generated by artificial intelligence. The progression to making music mobile, portable, and entirely digital was transformative; now, the creation of sounds and complete compositions by AI represent the next frontier.

As stated, a variety of musical genres merely exist because of technological progress, above all techno, which embodies its origins within its name. Without electronic technology, there would be no techno music. Today, it is the third most listened genre in Germany in 2024 (Statista 2024.000Z). The rapid development of artificial intelligence has impacted numerous fields of creative work, including music production. With this context, the question arises: to what extent is this novel technology capable of recreating and influencing the production of techno music itself? Is AI capable of generating authentic

---

sounds and compositions? Can it replicate existing characteristic sounds of techno music, and how does the quality of the results compare to that of traditional approaches? This study is to be understood as snapshot of the capabilities and limitations of AI-generated sounds and music in the field of techno.

# 1 **Goals and Methodology**

The goal of this study is to analyse and evaluate audio content generated by artificial intelligence in the style of techno. It investigates the extent to which AI can already generate audio content and tries to explore how it can be utilised in the production of techno.

To achieve this, the current state of artificial intelligence and techno music production techniques are reviewed. Then, expertise knowledge and opinions on the topic are collected through the conduction of expert interviews, giving important results to base further procedure on. In the following chapters, relevant AI-music-generation applications are analysed and evaluated based upon the extracted key knowledge of the interviews. Afterwards, experiments within an electronic music production environment are conducted, to explore the implementation and possible practical applications of AI-generated content. In the final chapter, the results of this study are discussed, summarised and an outlook on the future of the topic is given.

Thus, this work contributes to the debate on the topic of AI in music production within the techno genre and aims to provide a better overview of the current capabilities of AI. Additionally, it offers an insight on the current state of AI in music production and an outlook on future developments.

The focus lies on the generation, analysis and application of AI-generated content. Therefore, mixing, mastering and vocal production – except for spoken word - are not included in this work, due to them resembling their own aspect of production.

## 2 **Current State of Technology and Foundations**

The following section introduces the current state of AI, electronic music production techniques for the genre techno, and the current state of AI in the music industry. It aims to provide an understanding of the thematic relevance to the topic.

### 2.1 **Artificial Intelligence**

#### 2.1.1 **Definition Artificial Intelligence**

Artificial Intelligence, or in short: AI, is a term that used to belong to the studies of computer science but has permeated various fields of science in the recent years. Though, its definition is not universally consistent. While the specifics vary across different AI techniques, the core principle revolves around data. AI systems are trained and learn on vast amounts of data. Through this, they recognise and identify patterns and relationships within a short period of time relative to humans (Google Cloud 2024.000Z). By doing so, they are able to perform tasks typically associated with human intelligence, like understanding natural language, image recognition, decision making and problem solving (Schümann 2024, p. 15).

The problem is that human consciousness and intelligence are yet to be fully understood, including the human brain. Here, neural networks, that aim to replicate the human brain in a technical matter, is where they reach their limits. AI attains its “intelligence” from processing large amounts of data and mathematical algorithms that exceed the capabilities of the human brain. This machine learning is a central part of AI, but rather to be considered an evolution of digitalisation. (Mockenhaupt and Schlagenhauf 2024, p. 54). The specialised part of it is deep learning, which is employed for most commercial applications using AI. It is a process in which knowledge is deducted from experiences instead of predetermined rules. Therefore, it is not to be considered storage and accessing data, but an internal process of finding similarities and relationships (Klawonn 2023, pp. 36–37).

To further distinguish the capabilities of AI, it is often divided into two systems: strong AI and weak AI. Weak AI, also known as narrow AI, focuses on solving very specific application problems based on algorithms and methods from computer science and is capable of self-optimisation. Aspects of human intelligence are simulated and formally described, and systems for human thinking constructed.

Strong AI on the other hand is described as having equally intellectual capabilities to a human or even exceed them. This type of AI is only existing in research environments and is not to be expected to appear in the general public or applications in the near future (Mockenhaupt and Schlagenhauf 2024, pp. 60–61). A form of AI that is considered fully developed and indistinguishable from human intelligence is called AGI (Artificial General Intelligence) (Fabian J. G. Westerheide 2017). This science-fiction-like version of

artificial intelligence would introduce a multitude of ethical, philosophical and practical problems (Mucci 2024).

Although the term AI is frequently used, the distinction between deep learning and AI must be made, as AI pertains more than machine learning. To improve reading comprehension, when referred to as “AI”, this work indeed speaks of deep learning algorithms.

### **2.1.2 Development and Current State**

AI has shown a rapid advancement in in the past years; however, it is not an entirely new field of study nor interest. Creating intelligence or life from inanimate matter has been a theme since ancient times, as stories of golems have existed for thousands of years up until the famous novella Frankenstein from 1818. This deeply engraved fascination arises from the desire to explore and understand our own human consciousness (Schümann 2024, p. 15).

The more modern understanding of AI arose with Alan Turing and John McCarthy, the latter introducing the term “Artificial Intelligence” during a project proposal on 31 August 1955. One year later, McCarthy invited a wide array of scientists to a conference on AI, where they agreed on the definition of AI. (Mockenhaupt and Schlagenhauf 2024, p. 70). To them, AI was “the science and engineering of making intelligent machines” (Teneo.ai 2024).

In the following decade, interest and research in AI fluctuated due to false assumptions and unrealistic expectations about the progression within the field, which lead to an AI winter (mebis Magazin 2024). Then, in 1970, expert systems were introduced, marking the first successful model of AI. These systems solve complex tasks and aid with decision making by deducting knowledge from large knowledge bases of particular domains such as science and medicine (jvatpoint 2024).

In the following years, AI had not taken a desired path and so, during the 1980s, a second AI winter began. Then, in 1997, IBM’s chess computer Deep Blue beat the then reigning world chess champion Garry Kasparov in a match of chess, therefore being the first AI to exceed human capabilities in an expert field (IBM 2024).

Since then, through the further development of processing power and speed, worldwide networking, including broad access to the internet, and ubiquitous computing since the introduction of smartphones, AI has found its way into society (Mockenhaupt and Schlagenhauf 2024, p. 71). Now, neural networks and deep learning enable LLMs (Large Language Models) like Chat-GPT to flexibly generate complex content such as text, images, and music (Schümann 2024, p. 19). Despite the progression in the development and the integration of AI into a multitude of business branches and common technology, the future remains unclear. Regulations relating to ethics, data protection and an understanding of the impact of AI on society are desired by scientists, artists and businesses alike.

## 2.2 Techno and Production

### 2.2.1 Techno

Techno emerged in the 1980s as a more radical variant of dance music reminiscing the groove and danceability of disco from the 1970s, though presenting itself in a more experimental nature whilst staying “underground”. By the end of the decade, it had become its own genre with an attachment of sub-genres, and together with rap was often proclaimed to be the only innovative musical genre by the end of the 20<sup>th</sup> century (Phleps 2015, p. 87). Together with the distribution and availability of newly created electronic instruments, namely drum machines such as the TR-808 and TR-909 by the Roland Company, whose influence on dance music and hip-hop was compared to that of the Stratocaster’s influence on rock music, techno had found its musical identity (McKee and Jamie Grierson 2017). These drum machines, alongside the TB-303 synthesiser, inspired and influenced the development of entire genres within dance music, such as house and acid.

Today, techno sounds different from its origins in the 1980s. However, the sounds created by these outdated machines, are still prevalent in modern sound design and are employed to cater to these auditory habits, although often through samples or VST plugins (Virtual Studio Technology) (Vetter 2018). In addition to the use of these sounds, techno exhibits a range of characteristics. The tempo ranges from 120 to over 160 BPM (Beats Per Minute), making a classification solely on this measurement impossible. However, other characteristics such as the utilisation of 4/4 synthesised kick drums, as well as repetitive percussion loops, and the aforementioned synthesiser TB-303 or manipulated samples, make it recognisable. Typically, techno is associated with a darker feel and sonic elements compared to other genres, although it is not always present. Compared to other genres of dance music, techno has a higher focus on percussive elements and is known for creating moments of tension and release through its arrangement structure (SIX AM 2024a).

### 2.2.2 Production of Techno

Recently, dance music, including techno, has seen a rise in popularity over the past decade. In 2021, it was the fifth most popular music genre in Germany (Bundesverband Musikindustrie e. V. 2022, p. 27). This reflects the level of establishment it has achieved in recent years. Techno is particularly prevalent at festivals (Festival Fire 2022, p. 8). What originated in the 1980s from affordable drum machines and synthesisers can now be digitally replicated, as these instruments have not been manufactured for decades. Since its inception, a DIY (Do It Yourself) approach has been integral to techno's culture. With its experimental character and electronic nature, techno offers a wide range of creative production techniques. Today, powerful and affordable computers, DAWs, and VST plugins allow for the recreation of early techno sounds from home. Most producers do not have fully equipped studios and are thus categorised as “bedroom producers”

(Gensch et al. 2009, p. 263). This democratisation enables the general public to produce professional sounding music from virtually anywhere.

In the techno genre, producers typically take on the tasks of composing, recording, and arranging. Often, they also handle mixing and mastering themselves, as the necessary tools are readily available within DAWs (Keeley 2018). Nevertheless, analogue drum machines, synthesisers, and hardware controllers continue to be used, as their tactile interface provides a more intimate connection to the music. Thus, a combination of analogue and digital elements offers a satisfying compromise.

### **2.2.3 Sampling**

This process is, of course, not limited to techno, but has become increasingly established since the introduction of DAWs. It involves extracting sound patterns, noises, voice clips, or other acoustic elements from existing works and media or field recordings and adapting them contextually for production. It is also common to use entire sample packs and libraries, which are now often integrated into DAWs via cloud services and frequently operate on a subscription model (Avid 2024).

Samples are divided into two types: one shots and loops. One shots are, as the name suggests, singular sounds that are played once. Generally, they are employed to create custom patterns or to enhance arrangements, thus allow for more flexibility. Loops on the other hand offer pre-arranged patterns, that can be manipulated to fit into projects and enhance their overall sound quality (Soundtrap 2024).

### **2.2.4 Drums**

The foundational element of most electronic music genres is rhythm, typically established by 4/4 kick drums and percussion. In the early days of techno, these were exclusively produced by analogue drum machines. Nowadays, kicks are often synthesised from scratch or constructed from samples, with multiple samples frequently layered together. These layers typically undergo extensive processing with effects such as equalisation and compression. In techno, where sixteenth note sub-basslines, called "rumbles", are common, kicks often serve as both sub-bass and bass (Open Music Academy 2024). For example, short delay and reverb effects are applied to create a driving, deep rhythm. Alternatively, kick samples can be cut and arranged in staccato-like sixteenth notes in the second bar. In harder, more aggressive sub-genres, these kicks are often layered with noise samples or hi-hats with boosted mid frequencies to add rhythmic elements and texture. Particular emphasis is placed on balanced low frequencies, which should be summed to mono below a threshold of around 80 Hz in the mix to avoid phase issues (Horea 2023).

The distinctive low frequencies of the kick are often complemented by an off-beat hi-hat. Cymbals are typically sampled from drum machines, most notably the Roland TR-909 in techno, and are layered into repetitive pattern loops using different hi-hats and rides. Percussion elements also originate from drum machines, providing greater creative



freedom during production. When hi-hats, percussion, and sometimes kicks are sequenced and processed together, this is referred to as a top loop, commonly utilised in faster genres. The rhythmic section is often enhanced with snares and claps on the second beat. Pattern variations and continuous development build tension and release throughout the arrangement.

### 2.2.5 Synthesisers

Depending on the sub-genre, synthesisers play a significant role with varying presence. They typically handle melodic parts, although not all types of techno feature prominent melodic elements. Often, "warm" sounds, reminiscent of the 1980s and 1990s sound aesthetics, are employed. However, due to the significant financial investment required for vintage synthesisers, digital alternatives are often relied on. In general, there are no limitations or requirements regarding sound creation. Only the acid techno genre is distinguished by the intensive use of the Roland TB-303 bass synthesiser or its replicas, known for its distinct squelching sound and "barking" effect created by opening the resonance filter. Through sound design, it is possible to replicate this analogue technique in modern digital synthesisers (Baetz 2023). The importance of this synthesiser is based on the fact that it commercially and sonically failed upon release to be utilised as a bass guitar synthesiser, but inevitably revolutionised dance music through its unique sound and unintuitive programming methodology (Voigts et al. 2024, p. 234).

## 2.3 AI in the Music Industry

First known attempts at creating music with artificial aid date back to 1650 and Athanasius Kircher, who built the composition machine "Arca Musarithmica", to help musical amateurs compose music for ecclesiastical texts (Klawonn 2023, p. 31). More recently, with the emerge of LLMs, AI can generate music and lyrics with merely a few text prompts. Due to a shift from home-made scripts to APIs (Application Programming Interface) and web-based applications, the integration of AI into the music creation process has become more playful rather than technical (Voigts et al. 2024, p. 63).

The design of such usable AI music systems can be roughly divided into two categories: tools targeting music producers and others that are already familiar with production techniques; and tools and applications suitable for non-musicians in need – or want – of musical services. This includes film or video content producers, amateurs and businesses (Brown 2021, p. 6). Both systems have brought forward questions about originality and ownership of intellectual property regarding AI-generated content.

In 2023, the song "Heart on My Sleeve" created by the artist ghostwriter977 was submitted to the annual Grammy consideration as best rap song. The song featured deep-faked voices of Drake and The Weeknd without either artist's involvement or permission (Hight 2024). Following the reveal, Universal Music Group, home to both artists, removed the song from streaming platforms. Later that year, what is considered to be "the last Beatles song", Paul McCartney released a new Beatles song called "Now and Then", an old

demo featuring John Lennon's voice, in which the producer employed AI tools to extract and refurbish Lennon's voice (Sun 2023).

In March 2024, the US state Tennessee passed a legislation to protect music artists, e.g. performers, songwriters and other professionals against potential abuse of AI technology. One month later, an open letter organised by campaign group Artists' Rights Alliance was posted to the long-form writing site Medium (Artist Rights Alliance 2024). It was signed by over 200 artists, claiming that AI will "infringe upon our rights and devalue the rights of human artists", demanding for legislations and a possible stop in development of AI applications (McMahon 2024). In June 2024, music labels including Sony Music Entertainment, Universal Music Group Recordings and Warner Records sued the developers of AI music generators Suno and Udio over copyright infringement, further fuelling the discussion about AI-music (The Guardian 2024). It is important to recognise that currently neither German nor European copyright law contains any explicit provisions for the protection of AI products (Klawonn 2023, p. 64).

## 3 Expert Interviews

This chapter consists of information research, organisation and evaluation through the execution of expert interviews. The following will describe the structure, theoretical concept and its practical approach.

### 3.1 Qualitative Research

With the planning and implementation of qualitative expert interviews, the research has already reached the point at which independent information gathering and data collection can be considered. Everything begins with the development of an interview guide and ends with the evaluation of the data collected from the experts (Kaiser 2021, p. 63)

There are three different types of expert interviews:

1. Explorative expert interviews: they serve either general information collection in a yet little explored field of study, the preparation of systematic main research or the “opening of a field” in a way of identifying relevant experts (Kaiser 2021, p. 40)
2. Guideline-based expert interviews: they are more strictly structured forms of questioning with the aim to obtain hard facts that are limited in other resources or not available at all. By means of the interview guide, the survey is conducted with the clear objective of gathering specific knowledge that finds purpose in the answering of a precise research question (Kaiser 2021, p. 41)
3. Plausibility interviews: they can be held after the completion of the empirical research and are used either for gathering “practical recommendations” or to receive information about the presentation of the research results (Kaiser 2021, p. 41).

During this research, the focus lies on the explorative expert interviews mixed with the guideline-based expert interviews, as they deliver a much more open approach at the time of research. It is crucial to explicitly gather information on the research topic rather than relying merely on the reviewed research at this point.

#### 3.1.1 Approach with Expert Interviewees

The research topic of this thesis is divided into the following three subject areas:

1. AI in techno music production
2. Capability of current AI models and applications
3. Human-machine interaction and implementation

Acquiring interview partners is naturally a practical hurdle. The importance of this step lies in the influence that the interviewees have on the overall quality of the research. Qualitative interviews do not aim at producing results that are generally applicable on the research topic but rather understanding the topic on a basis of systematic analysis

(Kaiser 2021, p. 84). Considering this, choosing the right interviewees is crucial when conducting an analytical approach.

Interview partners were put into a target group and were chosen accordingly. This group consisted of experts in different subject fields with experience in music production, AI implementation, music industry, sound engineering and sound design. Three interview partners were chosen with varying degrees of knowledge in the respective fields of the target group to present a more varied point of view on the overall picture. The interview partners were partly known prior to the beginning of the conduction to ensure their expertise in their respective field and an open comfortable communication.

### 3.1.2 Data Collection

It is crucial to adapt existing interview guidelines to the current interviewee and their specific role in the target group. Interview partners can differentiate in knowledge and expertise in the chosen target subject fields, which makes an adaption of the questionnaire according to their respective expertise fields important. A core guideline is created, which is then transformed into individual guidelines through adding, subtracting or altering questions. To ensure that all three interviews and their results can be compared, most of the questions stay unaltered. The questions range from broad to precise, with the broader ones being more open and generally more interchangeable in place during the interview. This approach provides the interviewees with an opportunity to answer in a more holistic way.

### 3.1.3 Questionnaire

#### I. Introduction

- **Greeting and introduction to interviewee**
- **Small talk and acknowledgement**
- **Short introduction to the topic**
- **Privacy policy and consent to the recording of the interview**

AI is rapidly advancing in all kinds of areas, including, of course, the creative fields and media in general. This thesis deals with the current state of both generative and supporting AI in the field of music production and sound design, especially electronic music and its production techniques. The aim is to shed light on the extent to which AI can already replace and work with the humans behind the production. In addition to the creative aspects, technical aspects will also be considered to gain a comprehensive understanding of the effects of AI on music and its production. This thesis will analyse and compare the extent to which music is already being shaped by AI and

whether it has created new approaches in the music production process to a certain extent.

To get topic specific results, qualitative expert interviews are conducted. The experts have been chosen on basis of their knowledge, insight and experience in their respective fields.

**Any uncertainties on the interviewee’s part are clarified followed by the recording of the interview.**

**II. Introductory Questions**

- **What is your name?**
- **What is your job/position?**
- **How did you get to your position?**

**III. Main Part: Questionnaire**

1	Historical Development and current Status	Intention
	<ul style="list-style-type: none"> <li>• How has AI in music developed in recent years?</li> <li>• Which new developments and technologies in the field of AI and music are currently the most promising?</li> <li>• How relevant is AI currently for the music industry?</li> </ul>	Assessment/Expertise
2	Artistic Aspects	
	<ul style="list-style-type: none"> <li>• To what extent does AI influence creativity and artistic intuition in music?</li> <li>• To what extent can AI contribute to the development or expansion of (new) musical genres or styles?</li> </ul>	Experience/Assessment
3	Interaction between Human and Machine	

	<ul style="list-style-type: none"> <li>• How does the collaboration between you as a musician/computer scientist and AI work?</li> <li>• In which area does AI support you the most?</li> <li>• Would you describe AI as a partner?</li> <li>• What role does human input and your intuition play in the production of music with AI?</li> <li>• What impact does AI have on the way you work in music composition and production?</li> <li>• Does AI also help you with sound design or other areas?</li> </ul>	Experience/Expertise
<b>4</b>	<b>Ethical Considerations</b>	
	<ul style="list-style-type: none"> <li>• What ethical issues does AI-generated music raise?</li> <li>• How is copyright handled for music created by an AI?</li> <li>• Do you think AI can recognise cultural diversity and individual expression in music?</li> </ul>	Assessment/Expertise
<b>5</b>	<b>Prediction about the Future</b>	
	<ul style="list-style-type: none"> <li>• Do you think that music generated by AI could develop its own cultural identity in the future - or even already has one?</li> <li>• How do you see the future of AI in the music industry? (production, distribution, ...)</li> <li>• What impact do you see?</li> <li>• Will there be more "average" music? Will the gap between studios and bedroom producers widen because of the tools?</li> <li>• Do you see this as positive or negative?</li> <li>• How do you see the future of AI in terms of consumption?</li> </ul>	Assessment
<b>6</b>	<b>Personal Opinion</b>	
	<ul style="list-style-type: none"> <li>• Do you think that music generated by AI can have an artistic value similar to music made by humans?</li> <li>• Difference between generative AI and using tools?</li> </ul>	Expertise/Assessment

	<ul style="list-style-type: none"> <li>• What should we keep an eye on for the future?</li> </ul>	
<b>IV. Closing</b>		
<ul style="list-style-type: none"> <li>• <b>End of recording</b></li> <li>• <b>Quick evaluation</b></li> <li>• <b>Information about processing the collected data and in which way to expect an outcome</b></li> <li>• <b>Acknowledgements and farewell</b></li> </ul>		

Table 1: Interview Questionnaire

### 3.1.4 Conducting the Interviews

The interviews were held either on the online video platform Zoom or in person, after the interview partners confirmed their participation. Whether online or in person depended solely on the preference of the interviewees. In advance to the interview, the interviewees were provided with background information on the topic at hand, including reasoning as for why they were chosen to participate and the academic purpose and technical circumstances (Kaiser 2021, p. 92). Furthermore, they were informed about the recording of the interview and the process of data collection. The possibility to anonymise their names was given but none of them opted for it.

Every interview started with a greeting and an informal conversation about the topic. The interview structure was explained to them and after all questions were answered, the recording and the interview were started. The questions were asked in their respective categories, when possible, but because of the nature of an explorative interview, some questions were brought forward when appropriate. This flexibility to switch between topics within the guideline is necessary on the part of the researcher (Kaiser 2021, p. 95). The duration of the interviews ranged from 26 to 50 minutes. Immediately after the interview, a protocol was created to capture small annotations, the atmosphere during the conversation and reactions to questions. These protocols are not part of the transcription but rather function as an additional aid for later interpretations of the collected data (Kaiser 2021, p. 102). Subsequently, the recordings were transcribed, employing a smooth verbatim protocol (Mayring 2014, p. 45).

### 3.1.5 Interpretation

To interpret and draw information from the gathered data, the structuring technique was utilised. This technique's goal is to filter out specific elements of the material, and to categorise them in pre-defined structures. This ensures that the data can be evaluated

under determined criteria (Mayring 2023, p. 98). These categorised structures are deducted from the collected data and backed up by theoretical considerations about the goal of the analysis (Mayring 2023, p. 99). Through subsumption, categories are worked out from the text. Potential content is then subsumed and put into main categories and more nuanced subcategories.

Category	Subcategory	Interpretation
<b>AI in Music Industry</b>	Current State of Technology	<ul style="list-style-type: none"> <li>• Text-to-Music most interesting and fascinating</li> <li>• Rapid development</li> <li>• Quality is improving</li> <li>• Inflexibility</li> <li>• Agentification next step</li> </ul>
	Use of AI-generated Music	<ul style="list-style-type: none"> <li>• Stock music will be AI-generated</li> <li>• Background music for Videos, as content music already utilised</li> <li>• Jingles and advertising music will be non-human</li> <li>• Cost and time effective</li> </ul>
<b>AI in Music Production</b>	Use of AI Applications	<ul style="list-style-type: none"> <li>• Workflow optimisation</li> <li>• Not used for sound design yet</li> <li>• Comparison to sampling</li> <li>• Analysing tool</li> <li>• Used like an instrument</li> <li>• Stem separation</li> <li>• Creativity sparking</li> </ul>
	Knowledge	<ul style="list-style-type: none"> <li>• Curating is the essential aspect of production</li> <li>• Prior knowledge about music production required</li> <li>• AI cannot compensate completely for a lack of knowledge</li> </ul>



		<ul style="list-style-type: none"> <li>• Enabler for non-professionals</li> <li>• Inspiration for professionals</li> </ul>
	Quality	<ul style="list-style-type: none"> <li>• For AI-generated content too low, prompt adherence not high enough</li> <li>• Further processing required</li> <li>• Production quality increased due to workflow optimisation</li> <li>• AI-generated content does not improve perceived production quality</li> <li>• High quality music will increase even further in quality</li> <li>• Low quality music will increase in quality</li> <li>• Overall music quantity will increase significantly</li> <li>• Overall generic music quantity will increase significantly</li> </ul>
	Technology	<ul style="list-style-type: none"> <li>• Openness towards AI in music production</li> <li>• General progress of technology always existed</li> <li>• Democratisation happened already through DAWs, sampling and VSTs</li> </ul>
<b>Influence on Artistry</b>	Authenticity	<ul style="list-style-type: none"> <li>• AI-generated Music lacks authenticity</li> <li>• AI alone is entirely devoid of creativity or inspiration</li> <li>• AI can be used as a tool like any sample or instrument</li> <li>• AI can influence existing creativity</li> </ul>

		<ul style="list-style-type: none"> <li>• Artists need to discern themselves from AI</li> </ul>
	Threshold of Originality	<ul style="list-style-type: none"> <li>• A transfer of the medium is required</li> <li>• Distinction between generated by AI or with AI</li> <li>• Acceptable to use AI like samples</li> <li>• AI-generated music is not original</li> </ul>
<b>Ethical Concerns</b>	Identity	<ul style="list-style-type: none"> <li>• AI cannot represent subculture on its own</li> <li>• AI is missing an identity</li> <li>• AI may develop its own niche culture</li> </ul>
	Human Factor	<ul style="list-style-type: none"> <li>• Human behind the music more important than the music</li> <li>• Human itself is part of the music</li> <li>• Creativity is already not valued enough</li> </ul>
	Regulations	<ul style="list-style-type: none"> <li>• Licensing needed for AI, especially generated Music</li> <li>• Datasets important</li> <li>• Artists need to be protected</li> </ul>

Table 2: Data Analysis Interviews

### 3.1.6 Interview Evaluation

The evaluation is based on the collected data within the previous chapter “Interpretation”. While these represent summarised statements, the complete transcriptions of the interviews can be found in the appendix Transcription Expert Interviews in full detail. The following summarises all interviews based on the deduced categories.

Within the music industry, text-to-music was found to be the most promising current development in AI technology. Generally, the rapid progression of AI technologies was stated throughout the conversations. The significant improvement in quality of AI-generated content was highlighted particularly. Nevertheless, doubts were expressed regarding the quality and true prompt adherence, which were noted to be lacking in the current

applications. Furthermore, it was observed that content generation applications lack flexibility, especially regarding manipulation of select parts within the generated audio. For instance, it is currently not possible to manipulate or regenerate individual stems of generated songs. The progression towards further agentification of AI was highlighted as a future prospect, as well as an improvement in sound quality. Simultaneously, the future of "passive music"—music for video content background by companies and platforms, as well as corporate music—was attributed to AI-generated music. It was noted that content, such as advertisements, are already backed by AI-generated music. Additionally, the formulaic structure of existing stock music was mentioned and compared to AI-generated content. This is primarily due to the significant cost savings for companies, as they do not need to compensate studio musicians when using AI.

When it comes to the use of AI in music production, all interviewees agreed that their workflow and the quality of their work had improved using AI. However, this improvement does not rely on the use of AI-generated material and excludes sound design. None of the interviewees had used AI tools to assist them with sound design, primarily due to a lack of available tools for this purpose. Instead, AI tools were more commonly used for mix analysis, MIDI note generation, and stem separation. All these applications sparked creativity among the interviewees. It was repeatedly emphasised by all participants that their role was akin to that of a curator when working with AI, a role they also assumed in the regular process of music production with the use of samples and design choices. It was also noted that prior knowledge in the field of music production is still necessary to produce high-quality music. For individuals lacking this knowledge, AI acts as an "enabler," making music production possible without prior knowledge in music or producing. Regarding quality, all interview participants agreed that the quality of AI-generated audio has significantly improved over time and is now difficult to distinguish for untrained ears. Furthermore, there was consensus that there will be more average and unoriginal music in the future and that the quantity of music will continue to increase overall. Nonetheless, openness to advancing technology is important, as its progress is inevitable and has always existed. The democratisation of music production was generally viewed positively and compared to the advent of DAWs and VSTs.

When talking about AI-generated music, there was consensus that it lacks authenticity and an identity when compared to regularly made music. It was determined that AI lacks any form of inspiration, which represents a significant aspect of music. Music created without any human involvement is therefore considered uninspired and unoriginal. However, using AI as a tool is deemed acceptable. The interviewees emphasised the importance of achieving a threshold of originality for the final work to be perceived as authentic and inspired. It is particularly necessary to distinguish between AI-generated music and music produced "with AI". It was also noted, sometimes with positive connotation, that artists must differentiate themselves from AI, which is another factor that fosters creativity.

Finally, it was agreed that AI is not capable of representing musical cultures. However, it was suggested that AI could potentially develop its own culture. It was repeatedly

---

emphasised that the human element behind the music remains the most important factor in music production. Consequently, the overall presentation of the music and the artist will become even more significant in the future than it already is. However, concerns were also raised that creativity is not sufficiently appreciated as it stands, and this situation could deteriorate further with the advancement of AI. All interviewees agreed that artists should be better protected and that regulations regarding the use of AI-generated music need to be implemented now.

## 4 Generative AI Applications

In the following section, various generative AI applications are described, analysed, and evaluated for their potential current sound design capabilities regarding techno music. The following testing criteria were established to enhance comparability among individual evaluations:

**Consistency:** The same prompts are used as much as possible with the same settings to ensure that they are comparable and consistent.

**Efficiency:** The examples are generated in succession to make them easier to organise.

**Immediate Analysis:** The examples are listened to immediately, and the first impressions are noted down.

**Adaptability:** The number of prompts and repetitions, as well as the length of the prompts, are adapted to the limitations of the individual model and its capabilities.

**Cost-Benefit Analysis:** If the quality of the first results is considered subpar, the number of prompts and repetitions is decreased to save resources.

### 4.1 Prompting Methodology

It is ensured that prompts are optimised and adapted for different models, as not all models can process the same inputs or specifications. If the quality is recognisably poor or there are significant deviations from the desired result, the number of prompts is reduced to avoid unnecessary effort. This methodology ensures efficient and targeted use of prompts when working with different AI models and helps to optimise the use of available resources.

The prompts were chosen based around hard acid techno. Though not representative for the whole scene, the acid sound made by the Roland TB-303 remains one of the most distinguishable sounds commonly heard throughout techno music (SIX AM 2024b). The BPM prompt was chosen because it fits into the range of hard techno generally with 150 BPM on average, while 140 comes closer to other genres of techno to give the generative models more freedom and test their ability to distinguish between them. As the origin of the sound comes from a specific instrument, its name is stated, even though it can be generated with modern synthesisers as well.

#### 4.1.1 Prompts

- **Prompt 1:** A hard techno song featuring a Roland TB-303 synth bassline used as lead, heavily distorted percussion and a rumble bass. Dark mood and entrancing, very driving rhythm. The BPM should be around 150, so it is fast. The key is G minor.

Reference: [Prompt 1](#)

- **Prompt 2:** A hard techno song, dark ambient mood, aggressive Roland TB-303 synthesizers in sixteenth-notes. The kick is punchy and distorted with rumble. Heavily distorted Percussion, driving rhythm. BPM 150, key Gm:
  - Reference: [Prompt 2](#)
  - **Prompt 3:** Tempo 140 BPM, key Gm. Acid techno with a heavy rumble kick, driving but groovy rhythm, gated pads in a trance-like style, hypnotic and driving.
- Reference: [Prompt 3](#)

The results of the individual prompts are reviewed summaratively.

### 4.1.2 Evaluation Methodology

The AI applications and models are evaluated according to the following criteria:

#### **Performance (quality of sound, complexity and structure, prompt adherence, reliability)**

- How good is the quality of the generated sound on a technical level?
- How good is the quality of the generated sound on a perceived level?
- How complex are the generated sounds?
- How well structured are the generated sounds?
- How high is the prompt adherence?
- Can the results be repeated?

#### **Usability (ease of use, integration into workflow)**

- How easy is the use of the application?
- How good is the user experience?
- How easy is the further processing of the generated sounds?

#### **Creativity and Flexibility (creative outcome, parameter control)**

- How creative are the generated sounds?
- How dynamic are the generated sounds?
- What parameters can be adjusted before and after the generation?
- Does the application promote or spark creativity?

#### **Conclusion:**

- Is it worth the effort?
- Is the sound design of high quality?

- Can it be used as a tool and if yes, how?

After testing and analysing the applications, a summative conclusion on the individual model is given at the end.

All files can be accessed here.

## 4.2 Udio

Udio is a generative AI model and application that produces a broad variety of music in various styles. Users can generate clips or entire songs extended from said clips of up to 4 minutes and 22 seconds by only inputting a type, genre, or topic of music. Udio produces instrumental or lyrical clips with user generated lyrics. Alternatively, it can generate them as well using ChatGPT-4o. Manually added lyrics offer a broader range of expression than AI-generated lyrics.

Additionally, Udio can extend uploaded audio up to 15 minutes. It is only applicable for audio content to which the user owns full usage and distribution rights, and these extended generations are not publicly available. Udio uses a credit-based system for its generations.

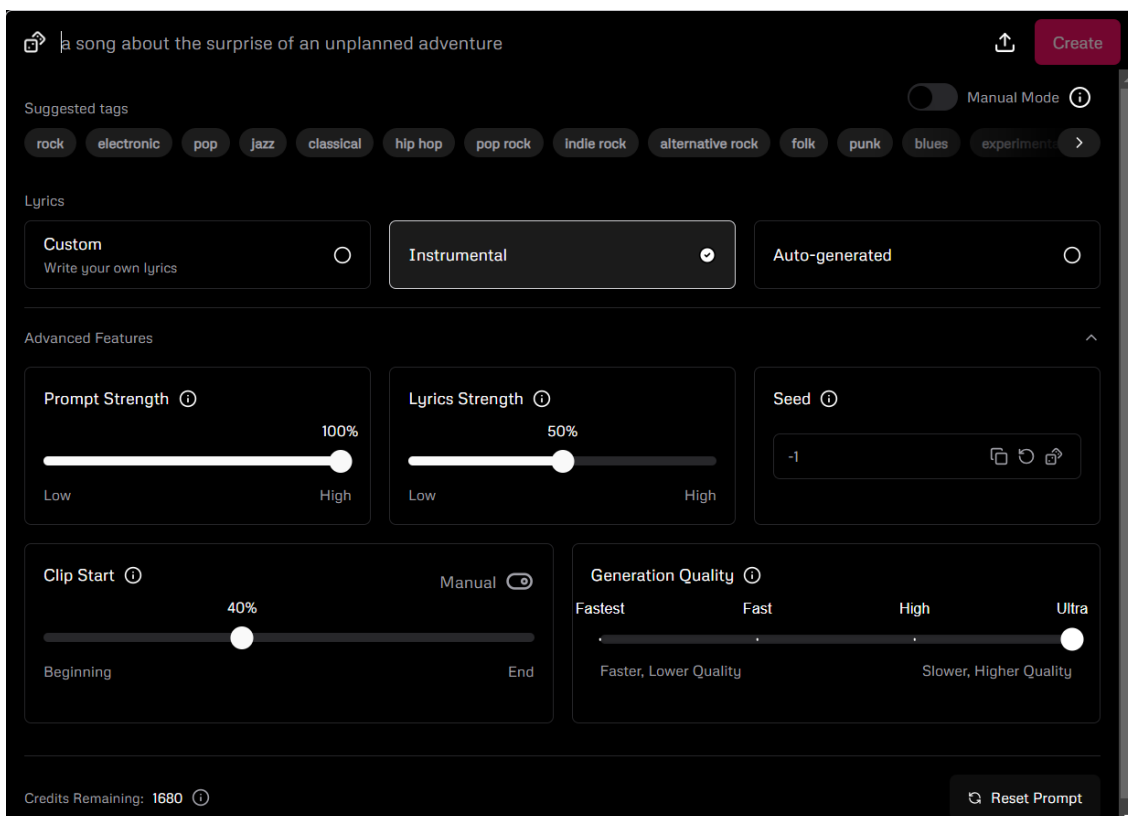


Figure 1: Udio Prompt Interface

To create a clip, the description of the desired output can be prompted (Fig. 1). Udio then gives suggestions for tags to use while creating, though they are not necessary to

choose. By default, Udio will rewrite the prompts for optimal output creation. If specific characteristics are desired, this optimisation process can be bypassed by enabling manual mode (Udio 2024)

It is possible to randomise the output but also to use suggested tags, while the manual mode is useful for precise and non-enhanced prompting. In its current beta state, Udio allows for up to 600 monthly generations in the free version, while premium users can generate up to 1200 songs per month. The unused credit tokens from free version are kept after subscribing to premium.

Random seed can be selected to try and make output songs more reproducible when manual mode is enabled. When the same seed is used, but the prompts are changed, the outcome could vary without explicitly changing the prompts for features in the song. Furthermore, prompt and lyrics strength can be set to a percentage, controlling the influence on the output. Higher values in prompt strength should lead to a greater adherence to the prompt, while potentially affecting the naturality of the outcome. In addition to that, the potential beginning of the outcome clip within the song can be set. For example, when set to 0 %, the clip will be an intro, while 90 % would be close to the end of the potential song. This, together with the extension feature, can assure that the outcome will be only the desired parts of the song. Lastly, the quality of the output clip can be controlled through a slider, which trades in speed-of-generation for higher quality.

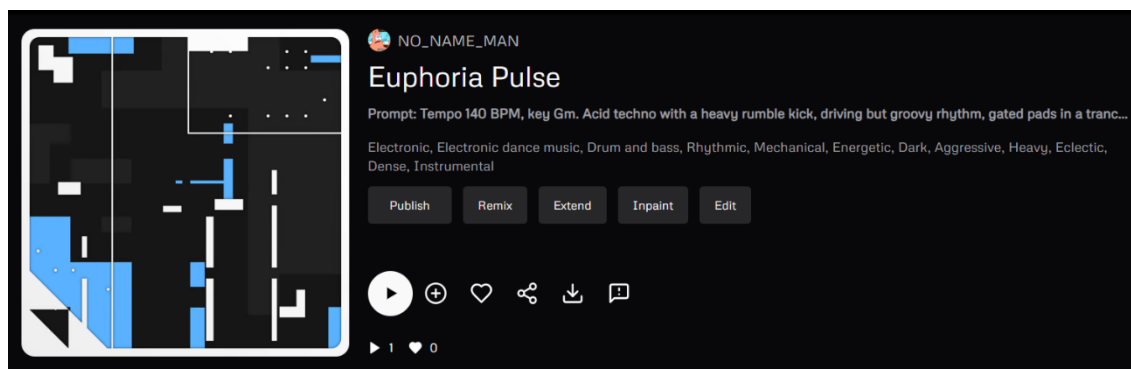


Figure 2: Created Clip and Description

Udio always generates two clips at once at 32 seconds length each. The clips are given a cover featuring geometrical shapes in three colours: magenta, light blue, and light green. In addition, the clips receive a title that is representing the prompts given (Fig. 2). There seems to be only limited variation between the covers and the titles, as the same names and covers occurred multiple times during testing. Once the clips are generated, it is possible to either publish, remix, extend, inpaint or edit them. Udio tags the clip with what it thinks to be fitting descriptions that are taken partly from the prompts, even though they do not apply to the generated clip.

A remix will redo the selected clip, but it is only possible with 32-second-long clips. The same parameters as in the regular create menu can be altered, additionally the variation to the original clip can be altered with a slider ranging from 0 to 1. An extension is to be



chosen if the length of the generated clip is desired to be extended. Through the extension menu, it is either possible to add an intro, an outro, a section before or after the original clip, thus giving the opportunity to build a desired song structure when repeated multiple times. Additionally, a context length slider can be set in the advanced features, giving control over how much of the original clip should be considered for the extension.

Within the standard tier of the subscription model, the feature of uploading and extending song snippets is also enabled. Through this feature, snippets of a few seconds or even minute-long clips can be analysed and extended, with giving additional prompting and quality features like in the extension menu. The additional extension is again 32 seconds long. Also hidden within the paid subscription tier is the ability to use the inpaint feature. Inpainting lets the user change up to 4 selected parts of the generated clip and re-generate the clip with the selected regions altered. This process re-generates only the selected regions.

## 4.2.1 Testing

### 4.2.1.1 Generating Clips

The prompts were not altered for Udio. Prompt strength was set to 100 %, clip start to 40 %, and generation quality to 100 % for all generations. The extension feature of uploaded audio was tested with a snippet from the self-produced song “Sonder” and a very basic loop of a kick and an acid bass on top to test how the AI responds to different lengths and complexities. Prompt strength was set to 50 %.

#### **Prompt 1:**

The loudness of the clips ranged from -10 to -14 LUFS integrated. Most were out of key or dissonant, making it difficult to determine the prompted key, which is not necessarily to be seen as a problem in techno. The tempo was slower than prompted for, at around 130 to 135 BPM instead of closer to 150 BPM. The mixes were at times poorly balanced, all of them with low stereo information and noticeable audio artifacts, which was to be expected. None of the clips matched the prompted genre, and they often sounded weird or uncanny rather than catchy and danceable. Only one clip featured an acid lead. Generally, the rhythm was not driving, and percussion was minimal and undistorted. Overall, the mood and adherence to prompts were significantly lacking.

Highest prompt adherence: [Clip 6](#)

Lowest prompt adherence: [Clip 5](#)

#### **Prompt 2:**

The loudness of the clips ranged from -11 to -14 LUFS integrated. They were not as dissonant as the clips from prompt 1. The tempo was again slower than prompted for, ranging from 130 BPM to 167 BPM even instead of 150 BPM as prompted. The mixes were at times poorly balanced, all of them with narrow

stereo width and noticeable audio artifacts. Three of the ten clips featured an acid lead. Generally, they were closer to the prompt than the examples from prompt 1. There was one exception, clip 10 turned out to be more drum and bass demonstrating significantly better sound quality. Overall, the mood and adherence to prompts were still lacking.

Highest prompt adherence: [Clip 6](#)

Lowest prompt adherence: [Clip 3](#)

### **Prompt 3:**

The loudness of the clips ranged from -5 to -14 LUFS integrated. They were again dissonant. The clips matched the prompted BPM of 140 more but did not get them right. Four of the ten missed the genre completely. What stood out was that the clips that failed the genre had an overall better quality and were much louder, closer to the genre specific average values. There was one exception, clip 10 turned again out to be more drum and bass with significantly better quality. In the end, the mood and adherence to prompts were lacking.

Highest prompt adherence: [Clip 1](#)

Lowest prompt adherence: [Clip 2](#)

#### 4.2.1.2 Extending Audio

The extension feature of uploaded audio was tested with a snippet from the self-produced song “Sonder” and a very basic loop of a kick and an acid bass on top to test how the AI responds to different lengths and complexities. Prompt strength was set to 50 %.

Snippet Sonder, length 3:17, prompt: “a part that changes up the melody”

The generated extensions did not come close to the original song and felt severely separated from the rest of it. The prompt adherence turned out to be very low.

Example: [Extension 3](#)

Snippet Acid test bar loop, length 0:13, prompt: “add more percussion”

The generated extensions of this short snippet followed the prompts much more adherently compared to a longer clip. It seemed that the lower complexity of the snippet also helped. Once, it added in something completely out of place; a piano. Overall, it works well for generating simple additions with shorter snippets.

Example: [Extension 3](#)

#### 4.2.2 Evaluation

**Performance:** Udio is at the time of writing one of the most developed AI models on the market. While the generated snippets vary in quality, they generally maintain a high

standard for AI-generated music. Clips can be downloaded as mp3-files. That said, the prompt adherence is rather low when examined more strictly in regards of genre. It appears that the prompted genre was not part of the data set used to train this AI model, as it switches to between genres or generates lower quality clips. Furthermore, achieving the desired outcome it is not easily repeatable. The extension feature only works with short snippets.

**Usability:** The user interface appears intuitive and clean initially. However, it gets confusing and disorienting when multiple clips are generated from the same prompts. Variation between the covers and the titles of the clips are significantly lacking, complicating the distinction between clips. Additionally, the downloading feature appears to be notably inconsistent and unresponsive.

**Creativity and Flexibility:** Udio's output is generally to be seen as dynamic. Nonetheless, when it comes to electronic genres, it is lacking in the more intense genres that require a more drastic approach during sound design and arrangement. The possibility to change selected parts of clips afterwards with inpainting gives plenty of flexibility to the user. Extensions offer a feature that can nourish creativity in the user, if they know how to manifest the idea themselves.

**Conclusion:** Udio offers high quality clips and even complete songs through multiple extensions and detailed prompting options. However, it is lacking in prompt adherence when it comes to electronic genres, indicating that further training data in these genres are required. The extension feature, when used with short snippets, can act as a source of inspiration.

### 4.3 Suno

Suno is a generative AI model and application that can create clips and songs on a prompt-based input in any genre of up to 4 minutes within the paid subscription tier. It also offers a continuation feature that can extend any uploaded audio with AI assistance. Much like its competitor Udio, Suno will only offer this feature if you own all the rights to the uploaded audio (Suno 2024). Like most AI applications, Suno makes use of a credit-based system for its generations. Additionally, Suno is part of Microsoft's Copilot AI (Ward 2023).

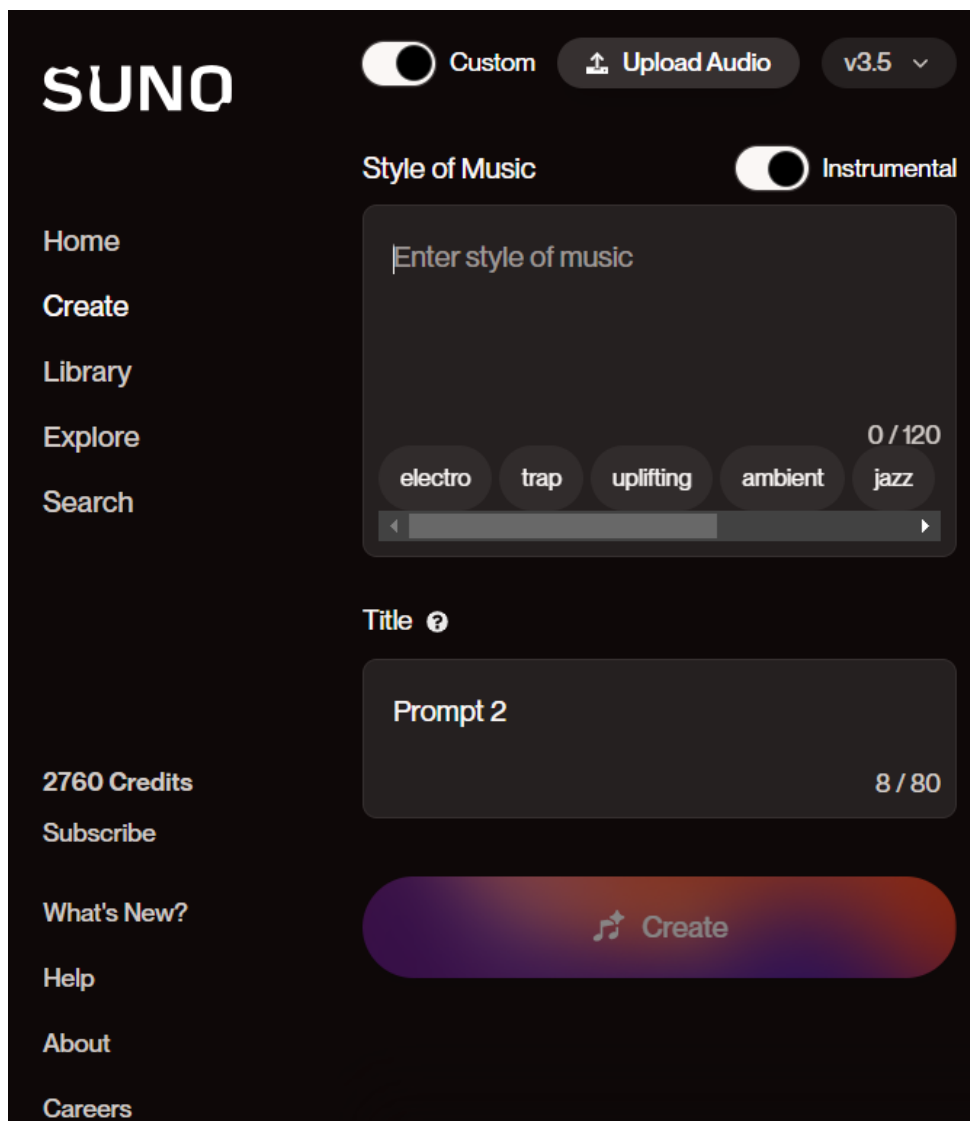


Figure 3: Custom Prompt Menu in Suno

Suno lets you input prompts of a length of up to 200 characters in simple mode and only 120 characters in custom mode (Fig. 3). It can create AI-generated lyrics, alternatively user lyrics can be input using custom mode, and even randomised lyrics can be generated. Furthermore, it can be set to instrumental only. Suno suggests optional prompts to use for the description of the style of music. If desired, older models of Suno can be selected for different outcomes or comparison. Titles can only be given manually in custom mode.

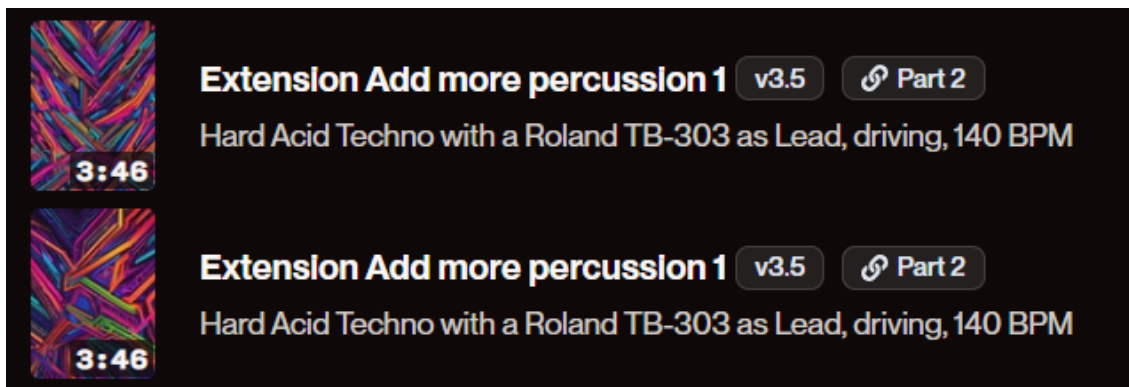


Figure 4: A Generation of Clips

Suno always creates 2 clips at once (Fig. 4). When in regular mode, complete songs of up to 4 minutes are generated. When in custom mode, clip length may vary. The clips are given names with relation to the lyrics or prompts if not specified in custom mode. The two clips are given AI-generated covers that share the same theme and represent the title of the clips. The clips can then be rated as good or bad, be published, shared or extended. The extension feature allows Suno to create clips that are longer than 4 minutes, which is the limit length for a generation. When extending AI-generated clips, the same prompt menu as in custom mode will open, only supplemented by an option to choose the section from the original clip that it should extend from. The newly generated extension is then given a tag that marks it as second part of the original clip, additionally acting as a link to the original.

Suno's continuation feature allows for uploading user-owned audio. This can be done directly through a microphone or by uploading a file (Suno 2024). Once recorded or uploaded, the clip is found in the library tab and can be extended from there using either simple or custom mode. The uploaded clips should range from 6 to 60 seconds in length. When a longer clip-length is desired, it can be extended to its "full length" by clicking on "Get Whole Song" within the more options menu. If done so, the generated clip is tagged as "Full Song".

### 4.3.1 Testing

#### 4.3.1.1 Generating Clips

The prompts had to be shortened for Suno, as it only allows prompts of a length of up to 200 characters. All important information was kept, while conjunctions were cut out. All prompts were generated in simple mode, instrumental enabled. As stated in prompt guide, only genres and vibes were used.

#### Prompt 1:

The loudness of the clips ranged around -14 LUFS integrated for all generated clips. They were neither continuously adherent regarding the prompted key nor the BPM, but generally close to it. The lowest BPM generated was 138 with the

highest being 170. The stereo information was above average in comparison to other AI models, not demonstrating the feel of a mono mix. The mixes were balanced and the quality very high for AI-generated music. Genres switched throughout, from psy-trance to trance and schranz, without any acid sounds used. The melodies and arrangement were very human-like and catchy, making use of more sophisticated song structure like breaks, build-ups and drops. However, the 4-minute length limit of the clips often ended the clips abruptly. Percussion was not as prominent within the songs as they were prompted. Overall, while generating high-quality clips with good structure, the prompt adherence was lacking.

Highest prompt adherence: [Clip 10](#)

Lowest prompt adherence: [Clip 1](#)

### **Prompt 2:**

The loudness of the clips ranged again around -14 LUFS integrated for all generated clips. They were neither continuously adherent regarding the prompted key nor the BPM, but generally closer to it than prompt 1. The lowest BPM generated was 144, the highest difficult to describe, as it kept rising throughout the song, but reached at least 155. The stereo information was again above average in comparison to other AI models, not giving the feel of a mono mix. The mixes were balanced and the quality very high for AI-generated music. Genres switched throughout, from melodic hard techno, to schranz, to modern hard techno and trance. Again, no noticeable acid sounds were found. Distorted percussion were more present when compared to prompt 1. The melodies and song structures were remarkable and catchy throughout, even more human-like than in prompt 1. Songs were cut off at the 4-minute-mark again. Overall, closer prompt adherence except for the absence of acid sounds.

Highest prompt adherence: [Clip 3](#)

Lowest prompt adherence: [Clip 10](#)

### **Prompt 3:**

The loudness of the clips ranged around -15 LUFS integrated for all the generated clips. They were neither continuously adherent regarding the prompted key nor the BPM, but with one exception close to it. The lowest BPM generated was 100, the highest around 166. Both outliers were out of genre. The stereo information was again above average in comparison to other AI models, not giving the feel of a mono mix. The mixes were balanced and the quality very high for AI-generated music. Genres were often trance-like, focusing on one feature of the prompt only. Cyberpunk, trance and drum and bass were found to be the most common genres. However, two of the clips featured an acid-like sound with no noticeable squelching. As prompted, the rhythm had more groove overall. The melodies and song structures were again remarkable and catchy throughout, to the extent that they sounded too familiar, though in the wrong genre. Most of the

songs were cut off at the 4-minute-mark again, but 3 of them had an intended ending. Overall, prompt adherence was low when considering the genre but higher in BPM and the trance-like pads, while the quality remained uncannily high.

Highest prompt adherence: [Clip 8](#)

Lowest prompt adherence: [Clip 6](#)

#### 4.3.1.2 Extending Audio

The extension feature was tested with a snippet from the self-produced song “Sonder” and very basic loop of a kick and an acid bass on top to test how the AI responds to different lengths and complexities. As Suno can turn short snippets into full songs, more testing was carried out, but the snippet for Sonder was shortened, as only audio of up to 60 seconds is allowed for uploading. Additionally, Suno allows microphone input as well, therefore a short recording emulating an acid lead was uploaded.

Snippet Sonder, length 0:13, prompt: “a part that changes up the melody”

Suno does not truly extend the original, it weaves it into a completely new song from the start. The added percussion was fitting throughout the generated clips. The melody was changed in every single output. Throughout the arrangements, the original snippet was referenced and used. The arrangements evolved the original snippet in an exciting way. On the downside, the BPM were speeding up and slowing down again in different parts in every generated clip, and the original snippet fades out during the process, inciting the feel of two separate songs being mixed into each other. Some of the created melodies felt completely out of genre considering the original.

Example: [Extension 6](#)

Snippet Sonder, length 0:13, prompt: “add more percussion”

Suno fails to adhere to the simple prompt. The added percussion was fitting throughout the generated clips, but additional melodies were added every time. When comparing with prompt 1, it is not possible to discern them from the generated clips alone. Throughout the arrangements, the original snippet was referenced and used in exciting ways. The arrangements evolved again into different BPM ranges and poppy melodies, giving them an unnatural feel.

Example: [Extension 4](#)

Snippet Acid test bar loop, length 0:13, prompt: “Hard Acid techno with a Roland TB-303 as Lead, driving, 140 BPM”

The generated extensions of this short snippet all followed a similar formula as the other tested clips. While they all added fitting percussion, the songs evolved into different genres with melodies lacking in originality. Additionally, the BPM

ranged from slow to very fast, while fading into new melodies, inciting the feel of a DJ mixing two songs into each other. One song added vocals. None of the clips added acid sounds nor stayed within the prompted BPM consistently.

Example: [Extension 1](#)

Snippet Acid test bar loop, length 0:13, prompt: “add more percussion”

None of the generated extensions of this short snippet followed the prompt. They all turned into clichéd pop songs. Some of them picked up the melody from the original snippet and evolved them, before switching to a different melody altogether. The BPM were mostly stable. The generated songs did not keep the original melody for long.

Example: [Extension 6](#)

Snippet Acid test bar loop, length 0:13, prompt: “add a squelching acid lead melody”

None of the generated extensions of this short snippet followed the prompt. They all turned into happy electro. Most of them picked up the melody from the original snippet and evolved them, but no acid sounds were added. The BPM changed throughout the songs in an unnatural way. The generated songs did not keep the original melody for long. Overall, Coherence to the original snippet was lacking.

Example: [Extension 6](#)

Snippet microphone recording, length 0:09, prompt: “dark acid techno with heavy rumble kick, driving rhythmic percussion, 140 BPM”

Of the six generated extensions, solely one was adherent to the genre of the prompt, and only so in the beginning. The BPM was adherent to the prompt. The use of the original sample was overall well placed and creatively used but gave the impression of interchangeability. Nonetheless, Suno delivered impressive results.

Example: [Extension 6](#)

Snippet acid with delay, length 0:06, prompt: “dark acid techno with heavy rumble kick, driving rhythmic percussion, 150 BPM”

The generated clips made good use of the sample and stayed in key. They also stayed in the BPM range of the original snippet, 140, even though 150 BPM were prompted. None of the generated clips were adherent to the prompted genre and none of them demonstrated a rumble kick. One of the clips utilised a delay on the snippet, giving it effectively an acid bass line. The stereo width improved in comparison to other generated tracks, as the width of the original sample was kept. The rest of the generated songs rather belonged in the universe of atmospheric electronic or rap instrumental. While very high in overall quality, the prompt adherence was lacking.

Highest prompt adherence: [Extension 2](#)



Most coherent example: [Extension 1](#)

### 4.3.2 Evaluation

**Performance:** Suno is at the time of writing one of the most developed AI models for generative music on the market. The generated snippets do not show much variation in mixing quality and demonstrate convincing sound quality for AI-music, containing more stereo information than other models. Clips can be downloaded as mp3- or WAV-files, making them easy to process. Generation-time is short. The arrangement and structure of the generated clips are exciting and can build tension and release. The generated clips do not exhibit all the prompted features but are often close to them.

**Usability:** The user interface appears intuitive and clean. However, it can get confusing and disorienting when multiple clips are generated from the same prompts. After creating many clips with similar prompts, the cover arts can look similar. Further processing of generated clips is straightforward.

**Creativity and Flexibility:** Suno's output is impressively arranged and detailed for AI-generated music to the point of feeling human-like in places, although its range in electronic music genres is lacking. The created melodies appear to be pleasing and catchy. However, the custom mode does not feature a lot of options, simultaneously making it less confusing but also giving fewer options to get the desired outcome. Extensions take a snippet and turn it into a complete song but the options to control the outcome are lacking, thus making it an interesting feature that potentially promotes creativity in the user.

**Conclusion:** Suno offers longer than average high-quality clips and complete songs with only a few clicks. The melodies are catchy and the arrangements impressive, but genre variety is lacking, indicating the need for further expansion of training data. The extension feature can show what possibilities the user could have at hand with only short audio snippets.

## 4.4 MusicFX

MusicFX is an AI-application made by Google as part of their AI Test Kitchen. MusicFX is powered by Google's MusicLM model and uses Google DeepMind's novel watermarking technology, SynthID, to embed a digital watermark in the generated outputs (MusicFX 2024). MusicFX is non-commercial and to be treated as experimental. As part of the AI Test Kitchen, its purpose is to enhance AI in cooperation with the community. Certain prompts, such as specific artists or anything including vocals, cannot be generated due to Google's desire to make AI safer regarding copyrighted material. Contrary to most AI applications, MusicFX does not use a token-system, but enforces a limitation of generations when in high demand. Additionally, the AI Test Kitchen is restricted to some regions of the world, excluding Germany, making it at the time of writing only accessible via a VPN connection.

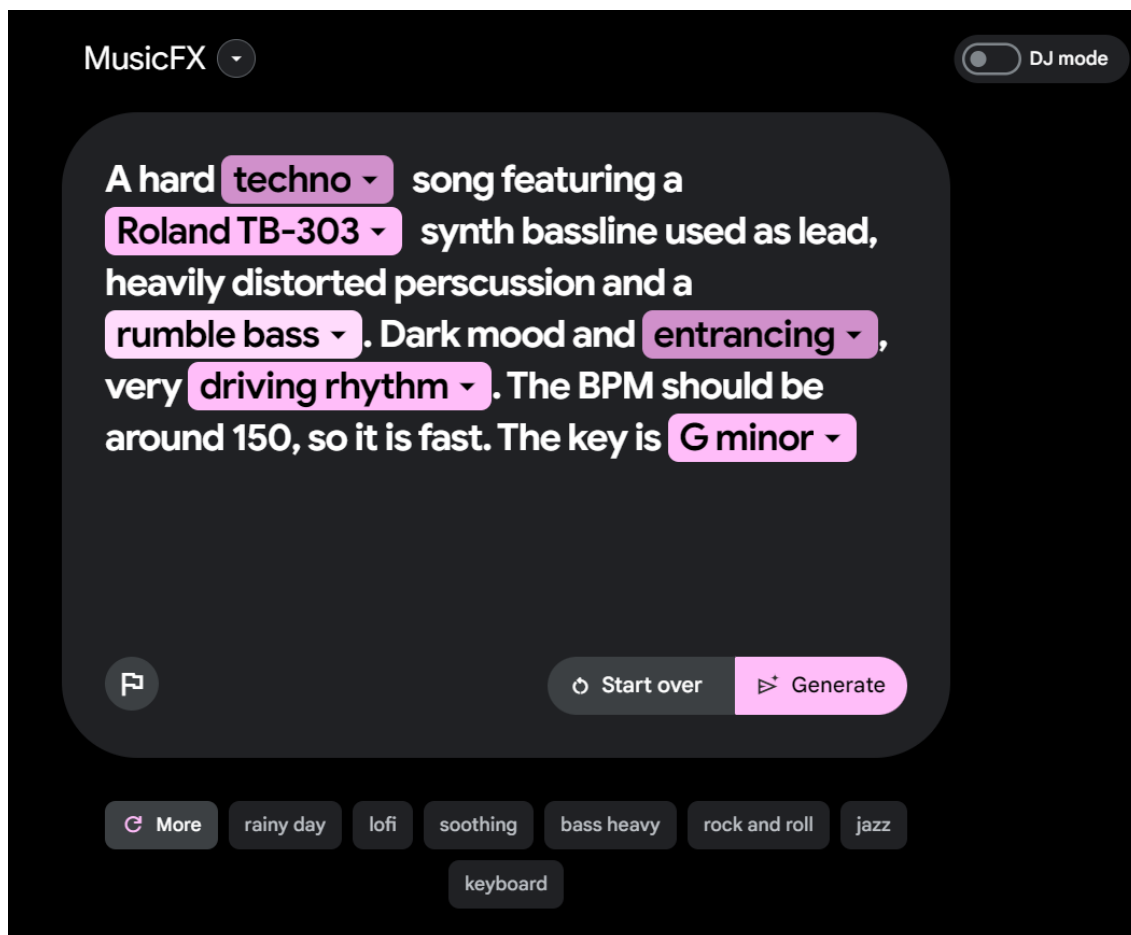


Figure 5: Music FX Prompt Menu

To the left-hand side of the screen, a large prompting menu is found (Fig. 5). Here, the desired song features can be described. MusicFX optimises the prompts automatically, making key characteristics of the prompts exchangeable through a simple drop-down menu. It also suggests optional moods, genres and instruments at the bottom. Additionally, prompts can be randomised. The clip length can be set to 30, 50 or 70 seconds in total. There is no option to extend generated clips, so the maximum length stays at 70 seconds. Additionally, the seed can be adjusted, and the output can be enabled to be loopable. Once, created, the clips can be shared or downloaded as WAV-files.

## 4.4.1 Testing

### 4.4.1.1 Generating Clips

The prompts were accepted by MusicFX and did not have to be altered.

#### Prompt 1:

The loudness of the clips ranged around -14 LUFS integrated for all the generated clips. They were all ranged around 150 BPM as prompted and stayed adherent to the key as well. The mixes were balanced and the quality very high for AI-generated music, apart from missing stereo information. None of them

featured acid sounds. Apart from that, they all demonstrated driving, rhythmic percussion and could all be considered hard techno. Some of them displayed some form of arrangement and tension development, but giving the short length of 70 seconds, there was not much to be expected to begin with. The sounds created were, apart from the missing stereo information, of very high quality. One clip showed inconsistencies with its low-end.

Highest prompt adherence: [Clip 5](#)

Lowest prompt adherence: [Clip 6](#)

### **Prompt 2:**

The loudness of the clips ranged around -14 LUFS integrated for all the generated clips. They were all ranged below 150 BPM but stayed adherent to the key. The mixes were balanced and the quality very high for AI-generated music, apart from one clip that failed to create a kick drum. None of them featured acid sounds. They would be more considered dark techno or raw techno. Apart from that, they all featured driving, rhythmic percussion. Arrangement development stayed minimal. The sounds created were, apart from the missing stereo information, of very high quality. Overall, lower prompt adherence than the first generations.

Highest prompt adherence: [Clip 5](#)

Lowest prompt adherence: [Clip 5](#)

### **Prompt 3:**

The loudness of the clips ranged around -14 LUFS integrated for all the generated clips. They were all ranged around 140 BPM but did not stay as adherent to the prompt as the previous generations. The mixes were balanced and the quality very high for AI-generated music. None of them featured squelching acid sounds nor gated Trance-like pads. Again, they would be considered raw techno or hypnotic techno instead of acid techno. Arrangement development stayed even more minimal, giving them the feeling of a loop. The sounds created were, apart from the missing stereo information, of very high quality. Overall, demonstrating the lowest prompt adherence.

Highest prompt adherence: [Clip 5](#)

Lowest prompt adherence: [Clip 2](#)

## **4.4.2 Evaluation**

**Performance:** The underlying model behind MusicFX, MusicLM, generates short clips of very high quality. The generations all fall within an acceptable range of genre, nothing differs completely from the prompted aspects. The generated clips can be downloaded as WAV-files, making them easy to process. On a related note, they are watermarked and not to be processed and published commercially. Furthermore, the stereo

information is very lacking, as it is only possible to generate audio clips in mono and without any vocals. With their short maximum length of 70 seconds, they do not provide full length songs on purpose. The prompt adherence is, when excluding the non-existing acid sounds, considerably high.

**Usability:** The user interface is exemplary intuitive and clean. While missing cover art or titles, the prompts are always legible and saved in a library in addition to that.

**Creativity and Flexibility:** MusicFX generates impressive high-quality music clips of short length, featuring catchy rhythms and driving beats. A lack of vocals is not problematic for the prompted genres. On the downside, not many controllable parameters are featured and there is no option to further process clips within the application itself.

**Conclusion:** MusicFX shows an impressive demo version of the AI model behind the application. While not adherent to all the prompted characteristics, the generated clips came very close to the identic sound of techno music. Furthermore, being a research tool, the exemplary handling of copyright issues must be praised.

## 4.5 Riffusion

Riffusion is an AI application based on stable diffusion made by Seth Forsgren and Hayk Martiros (Coldewey 2022). The currently available version is labelled as a demo version at the time of writing. In comparison to other AI applications, Riffusion offers only two prompt menus, one for inputting lyrics and one for the music description (Fig. 6). Lyrics can be generated randomly, as can the sound of the desired music clips.

Lyrics  
Soon I am done with my thesis!

Describe the sound  
Add genre, vocal style, or vibes...

✦ Prompt genius

Riff

Figure 6: Riffusion Prompt Menu

Riffusion always generates three clips for one prompt, including an individual cover visualising the given prompt, as well as a title. Each clip, or “riff”, is always 12 seconds long and saved in a library called “my riffs”. It is not possible to create longer clips. However, created clips can be remixed. They can be shared, downloaded as mp3-files or videos. Additionally, Riffusion offers a stem splitting feature, allowing to separate the clips into vocals, drums, bass and other track information. Riffusion does not employ a token or credit system.

## 4.5.1 Testing

### 4.5.1.1 Generating Clips

The prompts did not have to be changed for Riffusion. Each prompt was tested twice, accumulating a total of 18 clips. Due to the low prompt adherence in general, all prompts were evaluated summatively.

#### **Prompt 1:**

The loudness of the clips ranged around -14 LUFS integrated for all the generated clips. None of them generated the prompted genre nor the key. The BPM ranged around 140. The clips showed no stereo information, making them mono mixes. Generally, the demonstrated quality of the clips was low in comparison to other AI-generated music. Artefacts could be heard, and melodies were inconsistent and off-wandering. It was hard to discern specific genres, but they tended to stay closer to house. The brevity of the clips ruled out the possibility of song development from the outset. Overall, the quality was lacking and the prompt adherence low.

Highest prompt adherence: [Clip 2](#)

Lowest prompt adherence: [Clip 5](#)

## 4.5.2 Evaluation

**Performance:** Riffusion is at the time of writing one of the AI applications with a longer time on the market, and it shows. Clips in the tested genre all failed to adhere to the features prompted and contained noticeable artefacts. Additionally, the application does not offer stereo information. As generated clips fluctuated in genre so widely, replicability proved to be low.

**Usability:** The user interface is simple and intuitive. There is not much to prompt or parameters to control. Further processing of generated clips is straightforward, and the possibility to separate the stems from the application directly proves useful.

**Creativity and Flexibility:** The generated clips from Riffusion showed interesting but inhuman patterns and sounds. The range in electronic genres is severely lacking. The prompt menu does not offer a variety of parameters to change. However, inspiration can

be found within anything, and the option to download the stems directly could incite creativity.

**Conclusion:** Riffusion offers short clips from simple prompts and can be used to create samples or incite song ideas. However, due to the nature of their length, the generated clips seem to strive towards light entertainment rather than a useful tool for sound design or music production.

## 4.6 SoundVerse AI

SoundVerse is a generative AI model that features a chatbot assistant functioning as user interface (Fig. 7). According to the creators, “the more you speak to it, the more it starts understanding you and your goals” (Soundverse 2024). Additionally, what sets it apart from other AI applications based on generative AI-models is the studio editor, which enables the user to edit the generated clips within the application. This feature works alone or in cooperation with other users, once invited. SoundVerse’s chatbot assistant also features a voice output. Furthermore, it is possible to separate stems within the application and extend audio clips and auto-complete songs, including uploads by the user. However, vocal generation is not possible at the time of writing.

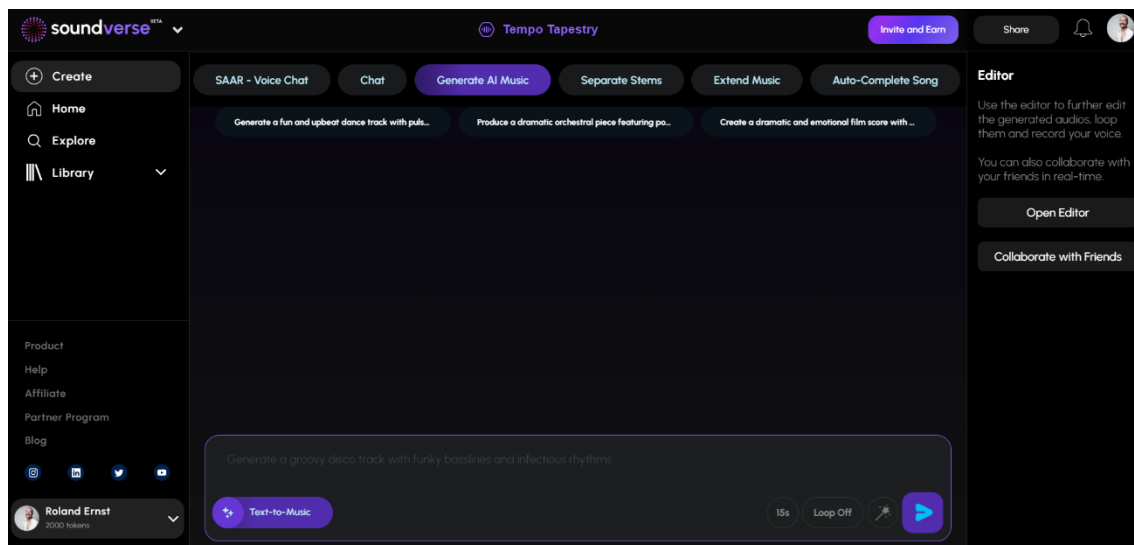


Figure 7: SoundVerse Interface

SoundVerse always generates two clips that are announced through the chatbot assistant, featuring the prompt and cover art that is supposed to represent the former. The options to extend them, or to separate the stems, are shown intuitively. Furthermore, the supposed BPM of the generated clips are displayed below them. From the chat window, the generated clips can be dragged into the editor to process them further (Fig. 8). There, the individual clips can be arranged in a simple style. The clips can be set to be repeated and can be renamed. Additionally, the BPM of the project can be set manually, and the user has the possibility to record their own voice and overdub it over the project.

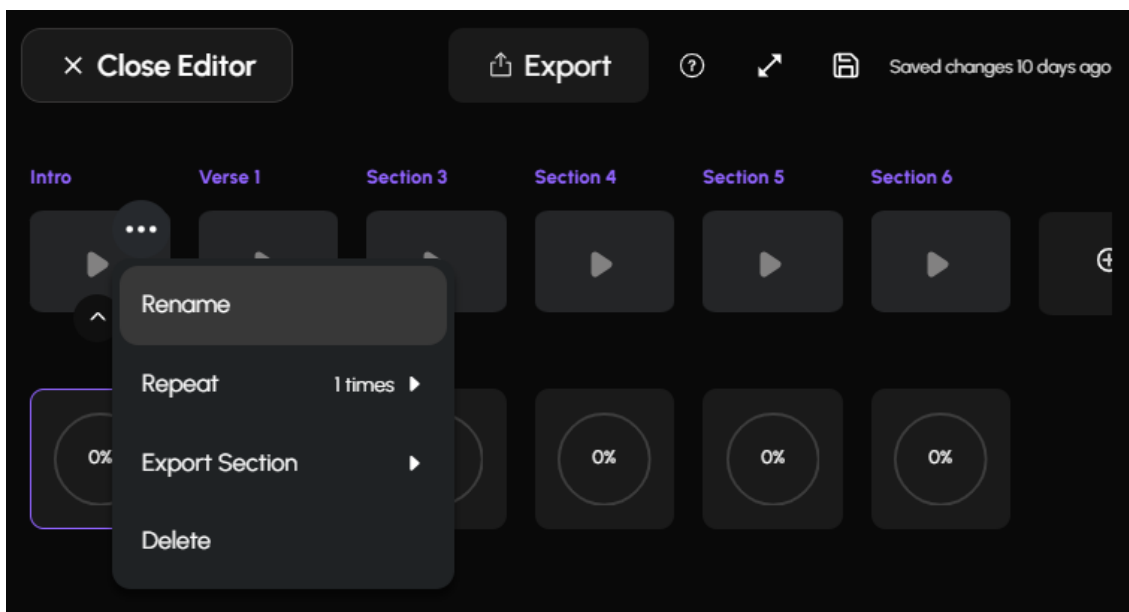


Figure 8: SoundVerse Editor Studio

## 4.6.1 Testing

### 4.6.1.1 Generating Clips

The prompts did not have to be altered for SoundVerse.

#### **Prompt 1:**

The loudness of the clips ranged from -14 LUFS to -17 LUFS integrated for all the generated clips, making them quiet. They were generally adherent to the prompted BPM but only half of them were correct to the BPM shown by the application itself. The key was adherent to the prompt. The stereo information was substantially wider than expected for AI-generated music, giving the mixes a wide feeling and depth. On the downside, the mixes were not balanced and often lacking in low-end information. The general quality could be described as average, due to the low-end being the most important part of the prompted genre. Acid sounds were lacking, and the genres were far from hard techno. The melodies ranged from catchy to generic. Attempts of arrangement developments could be heard, but none were fully executed. Percussion was not a standout-feature of the clips as prompted. Overall, the quality of the clips was acceptable, while lacking in prompt adherence.

Highest prompt adherence: [Clip 6](#)

Lowest prompt adherence: [Clip 2](#)

#### **Prompt 2:**

The loudness of the clips ranged again from -14 LUFS to -17 LUFS integrated for all the generated clips, making them quiet. They were generally not adherent

to the prompted BPM, while every second clip had a BPM in the range of 90 to 100, making them off-prompt by a large margin. The clips stayed adherent to the key prompt. The stereo information was substantially higher than expected for AI-generated music, giving the mixes a wide feeling and depth. On the downside, the mixes were even less balanced and often lacking in low-end information. No rumble kicks nor acid sounds could be heard, genre adherence was lacking substantially. The dark mood was captured adequately, unrelating to the missing features prompted. Arrangement development was found in some clips, but it was not clear if they were made on purpose, as they were found to be simple transitions from sections with a beat to sections without an underlying beat. Overall, quality and prompt adherence were lacking.

Highest prompt adherence: [Clip 4](#)

Lowest prompt adherence: [Clip 2](#)

### **Prompt 3:**

The loudness of the clips ranged around -14 LUFS to -17 LUFS integrated for all the generated clips. Half of them were close to the prompted BPM, while every second clip had a BPM outside of the genre average by a large margin. The clips key stayed adherent to the prompt. The stereo information was, again, substantially higher than expected for AI-generated music, giving the mixes a wide feeling and depth. The mixes were averagely balanced but often lacking in low-end information and consistency. Neither rumble kicks nor acid sounds were featured, and no trance-like pads, making the genre adherence lacking. Some of the clips featured interesting sound design, when ignoring the genre aspect. Arrangement development was kept to a minimum. Overall, the quality of the clips was lacking, but the sound design aspect was well enough done, even though general prompt adherence was not achieved.

Highest prompt adherence: [Clip 6](#)

Lowest prompt adherence: [Clip 1](#)

#### 4.6.1.2 Extending Audio

SoundVerse offers two features that provide the feature of extending audio clips, either uploaded or generated by the application. When extending audio clips through the extension feature, it is not possible to alter the given default prompt “Extend the provided audio clip by generating an additional 45 seconds of music. Maintain the style and mood of the original piece while seamlessly continuing the composition.” (Soundverse 2024). The audio clip will always be extended by 45 seconds. The extension feature of uploaded audio was tested with a short snippet from the self-produced song “Sonder” and a very basic loop of a kick and an acid bass on top to test how the AI responds to different lengths and complexities. Unfortunately, the extension feature, as well as the auto completion feature, did not function properly, and was stuck infinitely in a waiting request on



multiple attempts (Fig. 9). Therefore, not testing was completed with the auto-complete feature, while solely two clips could be generated from extensions.

Snippet Acid test bar loop, length 0:13, default prompt:

The generated extensions of this short snippet failed to continue the melody and the simple beat. A transition from the original audio clip to the added AI generation happened notably. Furthermore, the extensions lacked instruments and coherency to the original clip. Overall, the resulting clips were of inferior quality.

Example 1: [Extension 1](#)

Example 2: [Extension 2](#)

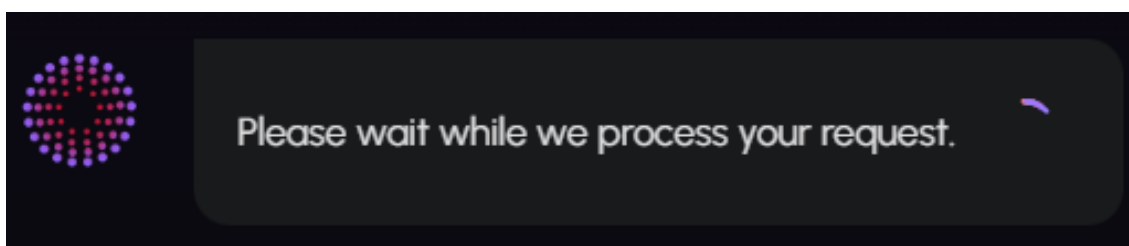


Figure 9: SoundVerse Chatbot not Responding

#### 4.6.2 Evaluation

**Performance:** SoundVerse generates short clips that are often not adherent to the prompt and demonstrating lacking low-end information. Unbalanced mixes with considerably more stereo width compared to other AI-generated music. Results cannot be repeated consistently. There is almost no arrangement happening during the clips. Overall, prompt adherence is low, with average quality.

**Usability:** The user interface appears intuitive on first sight but quickly becomes unintuitive when looking for further features. The screen space available is not used to full potential and some of the features are only found by using the arrow keys on the keyboard, which, again, proves itself unintuitive. Furthermore, the generation-speed ranges from slow to over 10 minutes. Multiple times, the extension feature was not working at all. The auto-complete feature did not work once. The studio, while not disruptive per se, appears unnecessary. While the download button for WAV-files is clickable, it only downloads unplayable files, which renders them unnecessary as well. Additionally, while interesting at first, the chatbot assistant evokes irritation over time and serves no beneficial purpose. Overall, SoundVerse is significantly lacking in its usability aspects.

**Creativity and Flexibility:** SoundVerse's output is, considered the low prompt adherence, acceptable. While nothing is notably positive, only few clips stand out negatively. However, not much arrangement development is found in the generated clips. The range in electronic music genres is considerably lacking, and melodies were not fully developed. While the application offers a variety of features, not all of them work consistently.

The studio however, while simple in nature, could promote creativity further in non-professional users.

**Conclusion:** SoundVerse introduces an assistant chatbot and a studio to edit generated clips to the formula of generative music applications. However, with a subpar user experience and a lack in prompt adherence in addition to unbalanced mixes, its most prominent positive aspect remains the stereo width of the mixes. Extension and autocomplete features lacked functionality and reliability.

## 4.7 Stable Audio

Stable Audio is a music generation application developed by Stability AI based on a stable-diffusion model (Stable Audio - FAQs 2024). It features an advanced prompt menu, text-to-audio generation and audio-to-audio generation. Stable Audio works with a token system that lets the user create up to 20 clips per month. As advanced prompting is seen as an important factor by the application, it provides a sophisticated guide on optimising input. Through this, it allows the creation of stems when following the guide. In addition to the prompt field, the application offers a prompt library sorted by genre, to help improving the desired outcome (Fig. 10). The time for generated clips default at 3 minutes but can be set shorter down to 0 seconds. When creating clips from audio, the duration is set to the time of the input audio clip.

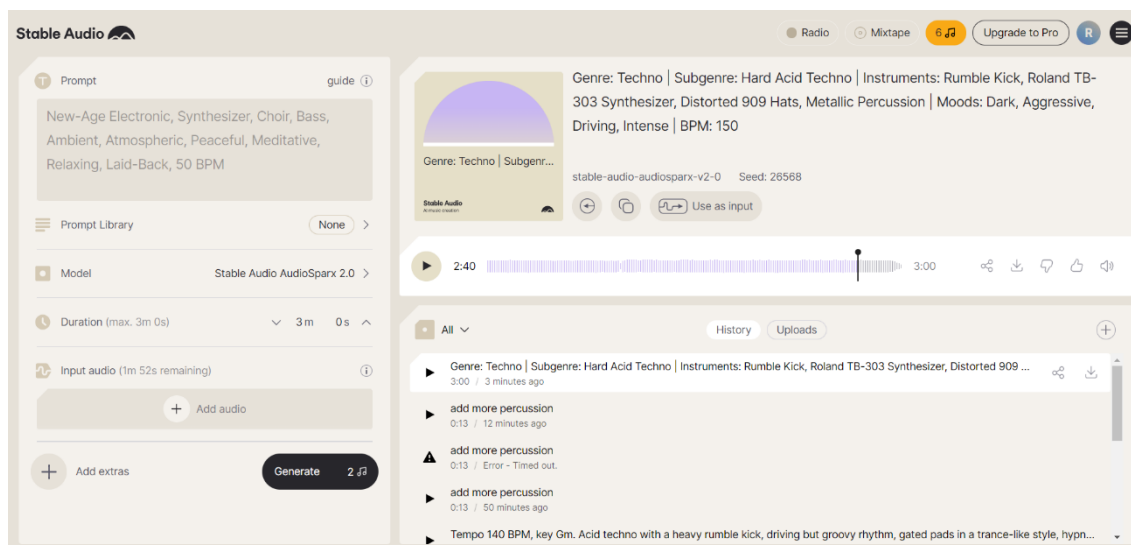


Figure 10: Stable Audio Interface

Steps, seed and prompt strength can be adjusted in the extras sub-menu. In the free tier of the subscription model, Stable Audio will generate a single audio clip. This can be altered solely in the paid subscription models. The generated clips do not feature special titles but rather present the used prompt as a title. Additionally, a random colour is assigned to the cover art of each clip. The clips can be reviewed, shared, downloaded and

used as input from the clip menu. A history tab saves all generated clips and songs by date.

## 4.7.1 Testing

### 4.7.1.1 Text-To-Text

The prompts were not altered for the use of Stable AI for most of the testing. Additional testing was carried out with optimised prompts following the prompt guide provided by the application. Due to early impressions of the generated clips and a considerably strict limitation of generations because of a low free token count, testing was carried out in a limited approach with only 2 generated prompts. Therefore, prompt evaluation is to be seen in summary. Prompt strength was set to 80 %.

#### **Prompt 1-3:**

The loudness of the clips ranged from -11 LUFS to -14 LUFS integrated. Generally, the BPM of the clips kept very close to the prompt. The notes fell all within the key, but most songs did not adhere to the prompt in this aspect. The soundstage ranged from narrow, almost mono, to wide. Mixes were balanced, although lacking in precision and direction in the low-end. With regards to arrangement, none of the songs exhibited any notable developments apart from filter sweeps, making the overall structure lacking in tension and relief. The sound quality was comparable to other current AI-models. What distinguishes the songs from the generations of other current models was the nature of repetition within the songs. The structure felt very loop-based, with the loops being of short nature. Melodies, while present, were lacking originality. Acid sounds were only present in one song and very static. The output genres could be described as hard techno. When prompted in accordance with the prompt guide provided by Stable Audio, the melody variety increased, as well as the arrangement, while the mix balance suffered simultaneously. Overall, prompt adherence was lacking in some respects. Arrangement and melody variety proved to be the major flaw.

Highest prompt adherence: [Prompt 3](#)

Lowest prompt adherence: [Prompt 2](#)

Prompted according to guide: [Example 1](#)

### 4.7.1.2 Audio-To-Audio

Stable Audio offers audio-to-audio generation, where an uploaded audio clip or a previously generated clip can be used as a base. The feature was tested with a snippet from the self-produced song “Sonder” and a very basic loop of a kick and an acid bass on top, to test how the AI responds to different lengths and complexities. Because of low free token availability, limited testing was carried out. Prompt strength was set to 75 %.

Snippet Acid test bar loop, length 0:13, prompt: “add a squelching acid lead melody”

The generated clips failed to add the prompted lead melody. Instead, the existent melody was altered sonically by adding effects.

Example 1: [Extension 1](#)

Example 2: [Extension 2](#)

Snippet Sonder, length 0:13, prompt: “add more percussion”

The generated clips added additional percussion as prompted. Additionally, they altered the pitch of the original audio and created a deviant rhythm from the original. The complexity of the generated songs varies and is generally considered low. Prompt adherence proves to be high with an exception to specific features of the prompt such as acid sounds.

Example 1: [Extension 1](#)

Example 2: [Extension 2](#)

#### 4.7.2 Evaluation

**Performance:** Stable Audio offers high-quality clips with a sample rate of 44.1 kHz. The genre variety is wide but lacking in sub-genres such as the prompted. Mixes are missing low-end information and clarity, while the stereo information ranges from very narrow to wide. The generated clips can be downloaded as mp3-files and are named according to the prompts, making them easy to process further if desired. The results are repeatable, though not adherent to prompt.

**Usability:** The user interface is simple and intuitive. However, prompting can get very complicated when trying to create a specific outcome. The application offers a guide on complex prompting. While this feature enables knowledgeable user’s potential to increase prompt adherence, it holds the possibility of decreasing the general output for inexperienced users.

**Creativity and Flexibility:** Stable Audio’s generated songs are often lacking in tension and dynamics as a result of minimal arrangement development. The created sounds, while not of low quality per se, are solely used in melodies low in originality. Because of the absence of a remix feature, generated clips can only be altered by using them for audio-to-audio generation. These clips may alter the melody in addition to the prompts, so caution is advised. By altering the original audio clips, this feature could provide the possibility to evoke creativity. The detailed prompting offers potential for high flexibility.

**Conclusion:** Stable Audio offers high-quality generated clips and a complex prompting system that even allows for creating stems when prompted correctly. However, the generated songs inhibit a sense of looped short clips and lack arrangement development, indicating an artificial sound.

## 4.8 MusicGen/LimeWire Studio AI

MusicGen is a single Language Model (LM) by Meta (MusicGen - Advanced AI Music Generation 2024). Currently, it is available on the LimeWire AI Studio by the same company known for their peer-to-peer-sharing file sharing client (LimeWire 2024). The prompt menu features an assistant, image creation, and music creation. Apart from optional prompt suggestions, no parameters can be controlled in the prompt menu. Once a clip is created, a chat window opens. Here, it is possible to set the duration of the clips in increments of 5 seconds from 5 of up to 45 seconds maximum (Fig. 11). It employs a token system as currency. The free tier provides 10 tokens per day, while a single 45-second clip generation costs 3 tokens. Generated clips can be heard from the chat directly, and variations can be created additionally.

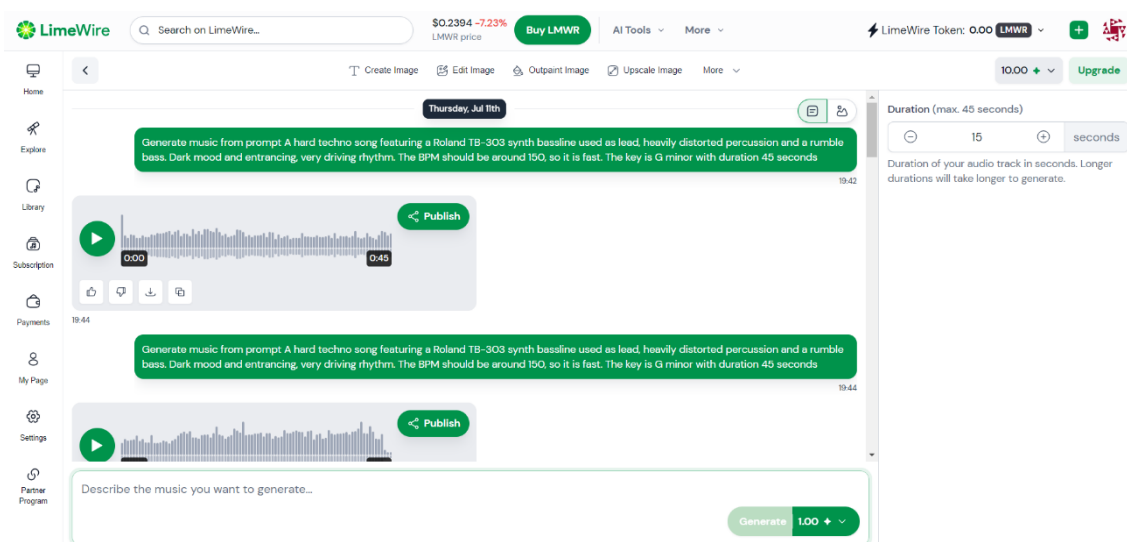


Figure 11: LimeWire Studio AI Chat Window

### 4.8.1 Testing

#### 4.8.1.1 Generating Clips

The prompts were accepted by MusicGen and did not have to be altered.

#### Prompt 1:

The loudness of the clips ranged around -14 LUFS integrated for all generated clips. They were true to the prompted BPM and stayed adherent to the key as well. Surprisingly, no songs were generated, but rather stems for the synthesiser alone. The clips were all in stereo and the quality high. Unexpectedly, they featured squelching acid sounds and even development during the arrangement. Apart from missing elements from the song, prompt adherence proved to be considerably high.

Highest prompt adherence: [Clip 1.2](#)

**Prompt 2:**

The loudness of the clips ranged around -12 LUFS integrated for all generated clips. They were true to the prompted BPM and stayed adherent to the key as well. Again, no songs were generated, but rather stems for the synthesiser alone. The clips were all in stereo and the quality high. Acid sounds were featured and again, development during the arrangement reminiscent of opening a filter. Apart from missing elements from the song, prompt adherence proved to be high.

Highest prompt adherence: [Clip 2.1](#)

**Prompt 3:**

The loudness of the clips ranged around -11 LUFS integrated for all the generated clips. They were true to the prompted BPM and stayed adherent to the key as well. Again, no songs were generated, but rather stems for the synthesiser alone. The clips were all in stereo and the quality high. Acid sounds were featured and mixed with pads, adherent to the prompt, and development during the arrangement reminiscent of opening a filter. Apart from missing elements from the song, prompt adherence proved to be high.

Highest prompt adherence: [Clip 3.3](#)

## 4.8.2 Evaluation

**Performance:** The underlying model behind LimeWire Studio AI, MusicGen, generates only short clips of up to 45 seconds. The quality is considerably high, including stereo information and development during the arrangement. Unfortunately, no song clips were created. This, however, can be considered practical for music production purposes, as it delivers isolated stems. The clips can be downloaded as WAV-files. Overall, the prompt adherence is, apart from missing instruments, remarkably high.

**Usability:** The user interface lacks organisation and is not intuitive from the beginning regarding parameters. With no cover arts and only a chat window functioning as history collection, the usability shows room for improvement.

**Creativity and Flexibility:** MusicGen is distinguished through the generation of actual acid-like sounds, including sonic development. However, a lack of controllable parameters and further instrumentation apart from the synthesiser proves it to be not generally suited for song generation.

**Conclusion:** MusicGen promises powerful and prompt adherent clip creation through its host LimeWire Studio AI, when considering the main instrument. However, it fails to provide a holistic solution to clip generation, and with limited generations, it only serves a purpose for sample or stem generation.

## 4.9 Eleven Labs

ElevenLabs is an application made by the homonymous research and deployment company (ElevenLabs 2024). The browser-based application is primarily known for its capabilities of cloning real-life voices and generating audio samples by means of text-to-speech. Additionally, by the end of May 2024, a sound effect generation feature was added to the website. This feature allows the generation of natural sound effects, speech, and musical clips of up to 22 seconds. According to ElevenLabs, the AI-model was trained on shutterstock’s audio library (Mehta 2024). Similar to many AI-applications, it employs a credit-token system.

The prompt menu for creating AI-generated speech can be used in simple or advanced mode. It is capable of processing 5000 characters at once for text-to-speech generation. (Fig. 12) Alternatively, audio files of up to 50 mb can be uploaded or directly recorded via microphone for speech-to-speech processing. For generating speech, a voice must be chosen. These voices are provided by ElevenLabs and make use of a tag-system to help identify a desired speech. Additionally, it is possible to upload and clone the user’s voice for generating speech. When using the advanced mode or by navigating to the settings button, a variety of models can be set, as well as stability, similarity and style exaggeration. By enabling “speaker boost”, the synthesised speech will exhibit more similarities to the original speech sample.

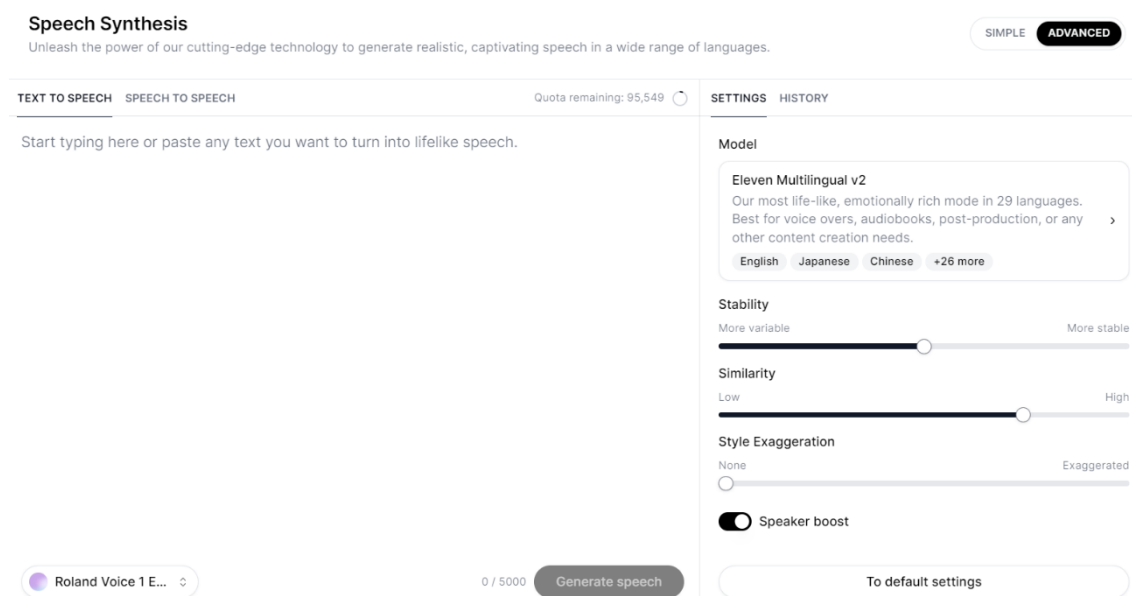


Figure 12: ElevenLabs Speech Synthesis Menu

The primary focus of this thesis is on ElevenLab’s sound effects generator. The prompt menu allows for a character count of 120 when describing a desired sound. Optionally, suggested prompts can be selected (Fig. 13). The settings menu allows for control over the generated clip length in seconds or provides the option to let the application automatically determine the clip duration. Furthermore, the prompt influence can be adjusted

to either adhere to the prompt more strictly or to let the model work more creatively. The application will always generate four clips that share the same icon of an Erlenmeyer flask when it represents a sound effect or an icon of a speech bubble when it represents a speech clip. All instances are saved within a history tab.

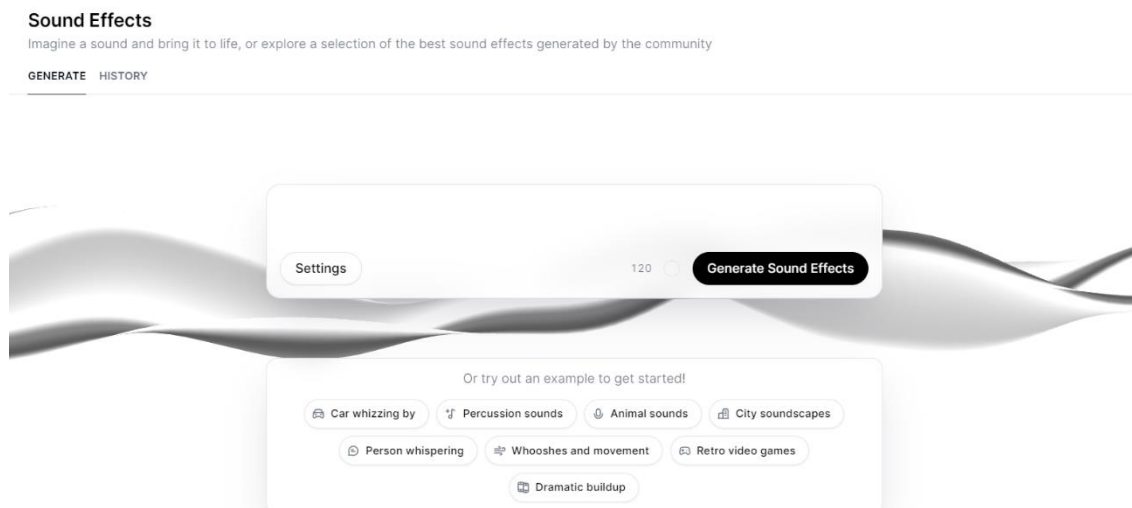


Figure 13: ElevenLabs Sound Effects Prompt Menu

## 4.9.1 Testing

### 4.9.1.1 Generating Clips

As this feature allows for the creation of one-shots and stems, extensive testing was carried out. A multitude of different sound types was chosen, each with three different prompts with increasing aspects, except for hi-hats, as a single prompt was found to be sufficient. Finally, the third prompt was developed to prompt for a loop of the sound type. While the clip length for individual sounds varied in accordance with a reference sound, the loops were all set to 3,2 seconds and prompted for 150 BPM, to ensure that the clips fit into 2 bars of a project set to 150 BPM. The loops were each generated four times. Prompt strength was set to 69 %. To ensure a better comparability, reference sounds for each sound type are provided. Sound types are evaluated in summary.

#### **Kick:**

Reference Sound: [Kick](#)

Prompts:

- a hard techno kick with heavy rumble, punchy and distorted mid frequencies
- a hard techno kick with a rolling bass rumble, punchy and distorted
- a techno kick with a rolling bass rumble, punchy and saturated, groovy



Loop:

- a techno kick with a rolling bass rumble, punchy and saturated, looped at 150 BPM

Evaluation:

The generated kicks were, considering the low bitrate, of high quality. Of the twenty-five generated kicks, solely five of them resulted in unwanted loops. The remaining samples achieved complete prompt adherence. The prompted loops appeared to be influenced by psy-trance but were generally prompt adherent as well. Solely for loops, because of their sonic complexity, artefacts were more noticeable. Overall, remarkably high prompt adherence.

Highest prompt adherence: [Kick 2.2](#)

Lowest prompt adherence: [Kick 3.4](#)

Example loop: [Loop 10](#)

### **Hi-Hat:**

Reference Sound: [Hi-Hat](#)

Prompt:

- a hi-hat for electronic music, slightly distorted and high-passed

Loop:

- a hi-hat for electronic music, slightly distorted and high-passed, looped at 150 BPM

Evaluation:

The generated hi-hats were, considering the low bitrate, of high quality. Of the twenty-four generated hi-hats, all proved prompt adherent and true to genre. Solely for loops, because of their sonic complexity, artefacts were more noticeable, which does not present a problem for this sound type within the genre. Of the twenty-four generated loops, only two failed to meet the prompted BPM. Overall, the results showed exceptionally high prompt adherence.

Example Hi-Hat: [Hi-Hat 10](#)

Example loop: [Loop 17](#)

### **Synth Stab:**

Reference Sound: [Synth Stab](#)

Prompts:

- a techno synth stab in g-minor
- a techno synth stab in a g-minor chord
- a short and plucky techno synth stab in a g-minor chord

Loop:

- a short and plucky techno synth stab in a g-minor chord, looped at 150 BPM

Evaluation:

The generated synth stabs were, considering the low bitrate, of generally higher quality. Because of the general description of the prompts, the generated samples varied in timbre and style, with some of them taking a sonic profile of an electric guitar processed with distortion. All samples achieved prompt adherence regarding the key. The loops achieved complete prompt adherence. Artefacts were, solely for loops, because of their sonic complexity, more noticeable. Overall, remarkably high prompt adherence.

Highest prompt adherence: [Stab 3.4](#)

Lowest prompt adherence: [Stab 1.3](#)

Example loop: [Loop 4](#)

### **Top-Loop:**

Reference Sound: [Top-Loop](#)

Loop:

- hard techno top loop at 150 BPM, made of various hi-hats, a high-passed kick, percussion and all of it side-chained to a 4/4 beat

Evaluation:

The generated top loops were of exceptional quality. Apart from the missing high-pass filter on the kick, all loops were adherent to prompt. Artefacts were more noticeable, but do not present a hindrance, as this type of sound is usually further processed. Overall, remarkably high prompt adherence.

Example Top-Loop [Top-Loop 7](#)

### **Spoken Word:**

Prompts:

- Ad-lib aggressive male shouting "what?!"
- ad-lib female for electronic music, calmly saying "welcome to the future"
- male english radio dj announcing "buckingham palace has announced the death of his royal highness the duke of edinburgh"

Evaluation:

The generated vocal samples were of average quality with noticeable artefacts. The first prompt was remarkably high in prompt adherence, with the lower quality of the audio even enhancing the authenticity. The second prompt showed, apart from one sample, high adherence to the prompt. However, the employed reverb decreased the overall impression. The third prompt demonstrated that more complex structures pushed the model to its boundaries, as none of them achieved

prompt adherence, mispronouncing the word “palace” in every instance. Additionally, some of the samples did not finish the prompted sentence, while some rearranged the structure. Overall, the results proved that complexity decreases prompt adherence.

Example prompt 1: [What 8](#)

Example prompt 3: [Future 5](#)

Example prompt 2: [Duke of Edinburgh 3](#)

### **Sweeps and Risers:**

Reference Sound: [Riser](#)

Prompts:

- build-up riser effect for electronic music
- white noise exhaust for electronic music, drenched in reverb, stereo
- short sweep-up sound for electronic music

Evaluation:

The generated effects were, given the simple prompt, only partly adherent to prompt. The build-up risers were all almost identical in length and pitch rise. While discernible as risers, the atmosphere was found to be closer to horror risers as a result of the low quality and high resonance. Of the eight created white noise exhausts, merely a quarter of them were considered white noise, and solely one to be considered an exhaust. The remaining clips consisted of silence or low-passed song clips. The sweep-up sounds showed the highest prompt adherence, but given the broad description of the prompt, turned out rather unoriginal sounding. Overall, the generated samples demonstrated unexpected unoriginality.

Highest prompt adherence: [Exhaust 4](#)

Highest prompt adherence: [Riser 4](#)

Highest prompt adherence: [Sweep-up 2](#)

### **Screech Stabs:**

Reference Sound: [Screech](#)

Prompts:

- hard techno screech stab, dissonant, key of g-minor
- hard techno screech stab, short and plucky, dissonant, key of g-minor
- techno screech stab in g-minor, plucky with short release-time

Loop:

- techno screech stab in g-minor, plucky with short release-time, looped at 150 BPM

Evaluation:

The generated screeches could be described as such. Predominantly, the clips were true to the key prompt of g-minor and exhibited a very harsh sound profile. Some of the supposed screeches represented growls, and some electric guitar licks. Several clips contained detailed stereo information. As prompted, around half of them exhibit a plucky sound, though often pertained to an unprompted loop. The short release-time given in the third prompt was not present. Interestingly, the loops are closer in sonic profile to acid sounds than to screeches. Overall, the prompt adherence could be considered high with some exceptions.

Highest prompt adherence: [Screech 1.4](#)

Lowest prompt adherence: [Screech 2.9](#)

Example loop: [Loop 9](#)

### **Acid Stabs:**

Reference Sound: [Acid Stab](#)

Prompts:

- acid synth stab in g-minor
- acid synth stab in g-minor, plucky with open resonance filter
- low acid synth stab in g-minor, plucky with open resonance filter, short release-time

Loop:

- low acid synth stab in g-minor, plucky with open resonance filter, short release-time, looped at 150 BPM
- low Roland TB-303 synth in g-minor, plucky with open resonance filter and mostly closed cutoff filter, short release-time, sixteenth notes, driving rhythm, looped at 150 BPM

Evaluation:

The generated clips were, considering the prompts, mostly short sequences rather than one-shots. The clips were true to the prompted key. Regarding sound, some of them were similar to acid. More adherent sounding clips were generated during the loop testing. Although having a low-passed sound, they generally came closer to the desired Roland TB-303 sound when prompted so. Overall, the prompt adherence stayed low, but it must be highlighted that some clips came close to the sound.

Highest prompt adherence: [Acid Stab 3.6](#)

Lowest prompt adherence: [Acid Stab 3.4](#)

Example loop: [Loop 9](#)

### **Percussion:**

Reference Sound: [Percussion Hit](#)

**Prompts:**

- snappy percussion hit
- thumpy percussion hit, dry
- distorted percussion hit, dry

**Loop:**

- distorted percussion hit rhythm, snappy, dry, looped at 150 BPM

**Evaluation:**

The generated clips proved to be of exceptionally high quality and prompt adherence. The primary prompted percussion hits were snappy and included reverb, as they were not prompted to be dry. The secondary percussion hits were lower in frequency and dry, and the distorted percussion hits exhibited high amounts of distortion. The loops were all true to prompt regarding BPM and featured rhythmic groove. All clips showed high stereo information. Overall, they achieved true prompt adherence.

Example prompt 1: [Perc 1.7](#)

Example prompt 2: [Perc 2.3](#)

Example prompt 3: [Perc 3.4](#)

Example loop: [Loop 8](#)

#### 4.9.1.2 Voice Cloning

Two clones of the user's voice were created with ElevenLabs' instant voice cloning feature. In this simplified feature, audio samples of 1 minute or longer are sufficient, whereas professional voice cloning requires at least 30 minutes of professionally recorded sample input. The recommended total sample length however is 3 hours for this feature.

Both voice models were trained on a 3-minute recording of the user reading out loud an article on Wikipedia in English (Wikipedia 2024). For voice clone 1, natural inhaled during speech were kept. In contrast, they were manually edited out of the same recording for voice clone 2, lending it a more commercial tone. For speech-to-speech generation, the spoken sentence "Hello! My name is Roland, and I'm a student at Stuttgart Media University." was repeated five times with a different timbre, pronunciation and intonation, to test whether the AI could analyse and replicate the differences. Additionally, the instruction for listening to AI generated presets, as used in chapter Replicating Presets was also cloned. These clips were then generated once each with the two different voice clones. For text-to-speech generation, the prompt sentences were generated five times. Settings used; stability 50 %, similarity 100 %, style exaggeration 50 %, speaker boost enabled.

All clips can be found [here](#).

### **Speech-to-speech:**

#### Voice 1:

All clips were high-passed and demonstrated a thinner sound compared to the originals. However, the intonation followed that of the original closely. Interestingly, while the original prompt 5 utilised a noticeably different pronunciation, the replica clip did not alter its pronunciation to this extent. Furthermore, when comparing the introduction audio clip, which employed a more monotonous intonation, the replica appeared to pronounce words marginally more enthusiastic. These differences are interpreted as a result of the training data. Overall, when excluding the filtered equalisation, the replicated voice clips presented a high degree of similarity and naturality.

Original: [Prompt 5](#)

Voice 1: [Prompt 5](#)

#### Voice 2:

This voice clone possessed a noticeably deeper timbre compared to the original user voice and demonstrated a less convincing sounding tone. While the intonation followed the that of the original closely, prompt 5 exhibited an unnatural sound and rhythm. This unnatural pronunciation appeared significantly more noticeable during the introduction clip, where the word “twice” possessed unnatural sounding slides. Overall, while convincingly enhancing the user’s original voice, the generated voice clips demonstrated a higher degree of unnatural intonation.

Original: [Original Introduction](#)

Voice 1: [Voice 2 Introduction](#)

### **Text-to-speech:**

#### Voice 1:

The tonal qualities of the voice remained unchanged from the speech-to-speech generations. However, due to the lack of an original audio clip to follow, the intonation changed noticeably. The spoken accent changed slightly towards a more British pronunciation. This could be heard during the introduction clips, as “Synplant” was pronounced differently (‘sɪn.plænt versus ‘sɪn.plɑ:nt). Furthermore, the letter “t” was pronounced more softly. The intonation, while often akin to a human melody of speech, sporadically changed velocity in an unnatural way. Overall, the generated voice clips sounded more monotonous and less human, while still maintaining a remarkably high degree of similarity to the original voice.

Highest similarity: [Student 1](#)

Lowest similarity: [Introduction 3](#)

### Voice 2:

The tonal qualities remained unchanged. However, the intonation changed noticeably, as well as the pronunciation. The cloned voice presents itself in a more British manner, resulting in a voice that exhibits a blend of qualities reminiscent of a British radio DJ and those inherent to the user. Additionally, some words were pronounced with less natural emotion. Furthermore, the British-sounding AI voice sporadically became discernible, e.g. with the word “twice”. Overall, while the quality of the generated voice clips remained consistent, the exhibited approximate resemblance to the original speaker induced feelings of uneasiness.

Highest similarity: [Introduction 2](#)

Lowest similarity: [Student 1](#)

## 4.9.2 Evaluation

**Performance:** ElevenLabs proves through its sound effect generation feature that true prompt adherence is possible. While not the entirety of the generated clips adheres to the prompts, a majority does. Percussive elements and loops demonstrated a remarkably high quality regarding sound design and arrangement. Furthermore, the generated audio clips stay remarkably true to key and BPM. With percussion sounds, the lower sample rate is less noticeable to lead sounds or voice snippets. On a technical side, the quality of the samples is lacking for a use in general music production. The generation time is very short. Furthermore, the voice cloning feature provides high-quality voice clips with a high degree of resemblance to the original speaker.

**Usability:** The user interface is perceived as intuitive and clean. Settings and the history of generated audio can be found without hindrance. However, as with many AI applications, it can get confusing when creating multiple clips from the same prompts.

**Creativity and Flexibility:** ElevenLab’s clips display highly advanced sound design and detailed sounds, to the point that they can be difficult to discern from regular samples. Remarkably, the AI seems to have been trained on many genres of electronic music. Interestingly, when counterchecking with shutterstock’s audio library, on which the AI model was allegedly trained on, no similar samples could be found (Exclusive Royalty Free Stock Music | Shutterstock 2024). Regarding flexibility, optional parameters are lacking. However, the duration setting in conjunction with prompted BPM result in clips with nearly fitting durations. Regarding voice cloning, the number of settings is adequate.

**Conclusion:** ElevenLabs distinguishes itself from other AI music applications by providing a sound effect feature that can be used to generate stems or one shot samples. The generated audio clips are remarkably prompt adherent and are often only discernible by the low sample rate and occasional artefacts. Additionally, the voice cloning feature provides mostly convincing voice clips.

## 4.10 Other Applications

A wider array of applications was tested, but due to their lacking output quality during testing, they did not appear to be of adequate quality for this study to be examined individually and extensively. Therefore, they are evaluated on a smaller scope in this composite chapter.

### 4.10.1 Audiobox

Audiobox by Meta is a research demo of a generative AI model that can produce sound effects and voices from text or audio prompts, similar to ElvenLabs. However, they implemented safety guardrails such as prompt filters for race, gender, sex, and religion, harmful associations, violence, and other sensitive topics (Audiobox | Meta FAIR 2024). Additionally, watermarks are added to the generated audio. Because of these safety measurements, none of the prompts utilised in testing other AI-applications were accepted and repeatedly resulted in a processing error. Therefore, the application is not suitable for further research with respect to this study.

### 4.10.2 Sonauto AI

Sonauto AI is another generative AI-music application that leans towards the creation of songs using celebrity voices such as Frank Sinatra or Johnny Cash. It offers a prompting menu for text-to-music generation similar to other applications and an audio upload feature for references. However, apart from an unintuitive user interface, the generated audio clips demonstrated sub-par prompt adherence and remarkably low arrangement development, even the absence thereof overall. It was clear that the AI model's data set did not contain enough training data to create adequate quality for electronic music. Despite this, the implementation of a stem separation tool sets Sonauto AI apart from other generative AI applications.

Example: [Prompt 1](#)

### 4.10.3 Rightsify

Rightsify is a global music licensing agency providing content music for businesses. Since December 2023, it is possible to use their generative AI music model Hydra to create music with a text-to-music approach, offering more flexibility to the desired content (Rightsify 2018). However, due to the corporate nature of the licensed music, the application did not demonstrate the capabilities to create adequate prompt adherence with the provided inputs. Additionally, offering a very limited free token amount and prices aimed at a business clientele, no further testing was conducted on Hydra outside of the free subscription.

Example: [Prompt 1](#)



#### 4.10.4 Snowpixel AI

Snowpixel AI is a media platform offering generative image, video, animation, 3D objects, and audio AI models based on text-to-media. Based on topic of this study, nothing outside the audio generation model was experimented with. Snowpixel employs a credit system, with no free tier available currently (Snowpixel 2024). For research purposes, upon request, the user was granted a total of two credits. Due to this limited number of available credits, no extensive testing was conducted. The generated audio clips demonstrated unnaturally high, inconsistent stereo information, panning, and noticeable artefacts.

Example: [Prompt 1](#)

#### 4.11 Mubert

Mubert is a generative AI model that creates songs via text-to-music. It is, according to the developers, one of the pioneers of commercial AI music (Mubert 2024). The application features a prompting menu similar to other generative AI music applications with possibilities to set the length and type of output, i.e. track, loop, mix, and jingle. In addition to using the prompting field, Mubert offers a feature to choose tags from predetermined genres, moods and activities. Because niche genres such as acid techno and hard techno were eligible, they were added during testing.

When examining the generated tracks, they appeared to clearly adhere to the genre tag given and exhibited exceptionally high sound quality. However, none of the manually stated prompt features, such as BPM or key, were featured adherently in the resulting audio clips. This conveys the impression that Mubert works by merely interpreting the genre tags rather than the user prompts. Furthermore, its data set is composed of specifically designed musical data created and tagged by humans, leading to the assumption that it is essentially a composing tool for pre-designed samples, rather than focusing on generative sound design. Therefore, it is not considered relevant to the subject of this study.

##### 4.11.1 AIVA

AIVA is a generative AI music generation application and artificial composer by the homonymous AI model (Artificial Intelligence Virtual Artist). It has been officially recognised as a composer by SACEM (Société des Auteurs, Compositeurs et Éditeurs de Musique), the author's right society in France and Luxembourg (Lauder 2017). While this marks a significant milestone in the AI industry, it is also a subject of concern for human artists. However, similar to Mubert, AIVA has no capability to design sounds itself, making it a mere compositional tool. Furthermore, no prompting capabilities are exhibited outside of key, duration and clip number, after a genre is chosen from its predetermined selection. As a result of this lack of controllable parameters and due to its predetermined framework, its relevance regarding this study is neglectable.

## 5 Sound Replicating Plug-Ins

This chapter analyses two of the most promising AI applications for sound design that are employable directly within a DAW. These plug-ins are the only AI-based applications on the market capable of replicating or cloning audio clips. To ensure coherence in comparability, evaluation criteria from chapter Evaluation Methodology are applied for these applications.

### 5.1 Synplant 2

Synplant 2 is a software synthesiser plug-in developed by NuEdge Development that was created by Markus and Frederik Lidström and released in October 2023 in its second iteration (Magnus and Lidström 2024). It represents a more experimental approach towards sound design through planting “seeds” and evolving them into “branches organically”. By doing so, it is possible to create personal synth patches. By evolving the “branches”, all patch settings are randomised. Branches hold unique sound properties each, with an exception for macro sliders like atonality, tuning, effect, volume, and release. The mode switch stays unaffected from branch evolvment as well. The application introduces an AI powered patch generator called genopatch, that can analyse, and transfer given audio samples into manipulatable patches (presets). The resulting patches, or presets, exhibit the main characteristics of the analysed samples when working correctly. However, not all sounds can be replicated, and alternative interpretations can be generated as result of this.

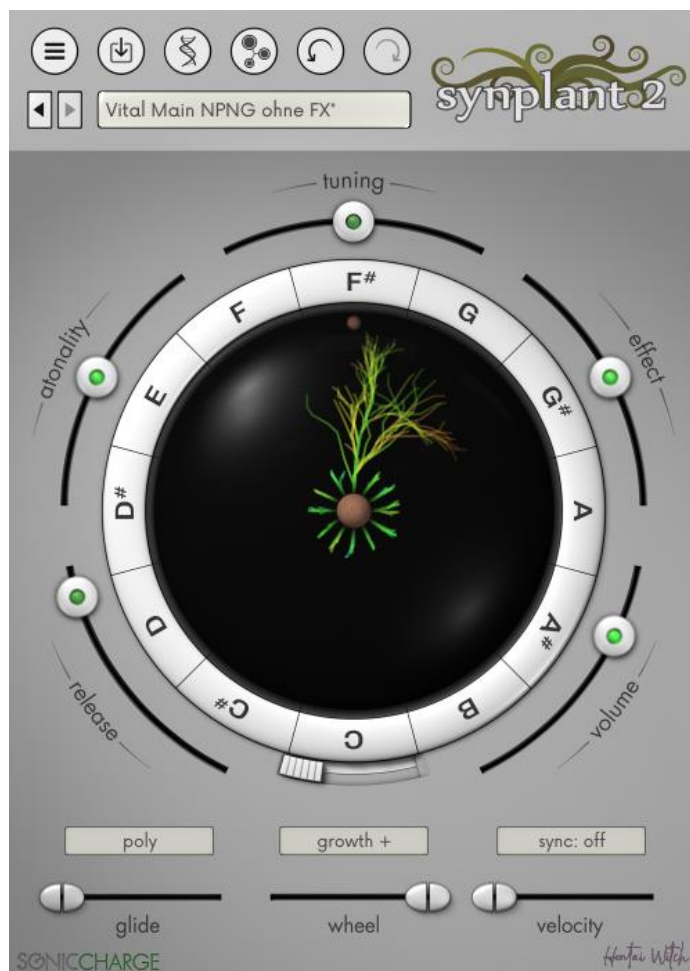


Figure 14: Synplant 2 Interface

The main interface presents a circular bulb containing a visual representation of a seed called “DNA Editor” (Fig. 14). From the seed, branches can be dragged out towards the key indicators. Each branch is connected differently to synthesiser controls under the “gene manipulation tab” such as envelope and LFO, oscillators, and filter and effect modulations. Through this design, Synplant offers a more organic and random sound design approach compared to other synthesisers. However, the focus of the conducted tests lies on its genopatch feature, which allows for replication of audio samples by transferring them into patches. These patches can be manipulated and shaped with synthesiser controls, allowing for adjustments and dynamic development.

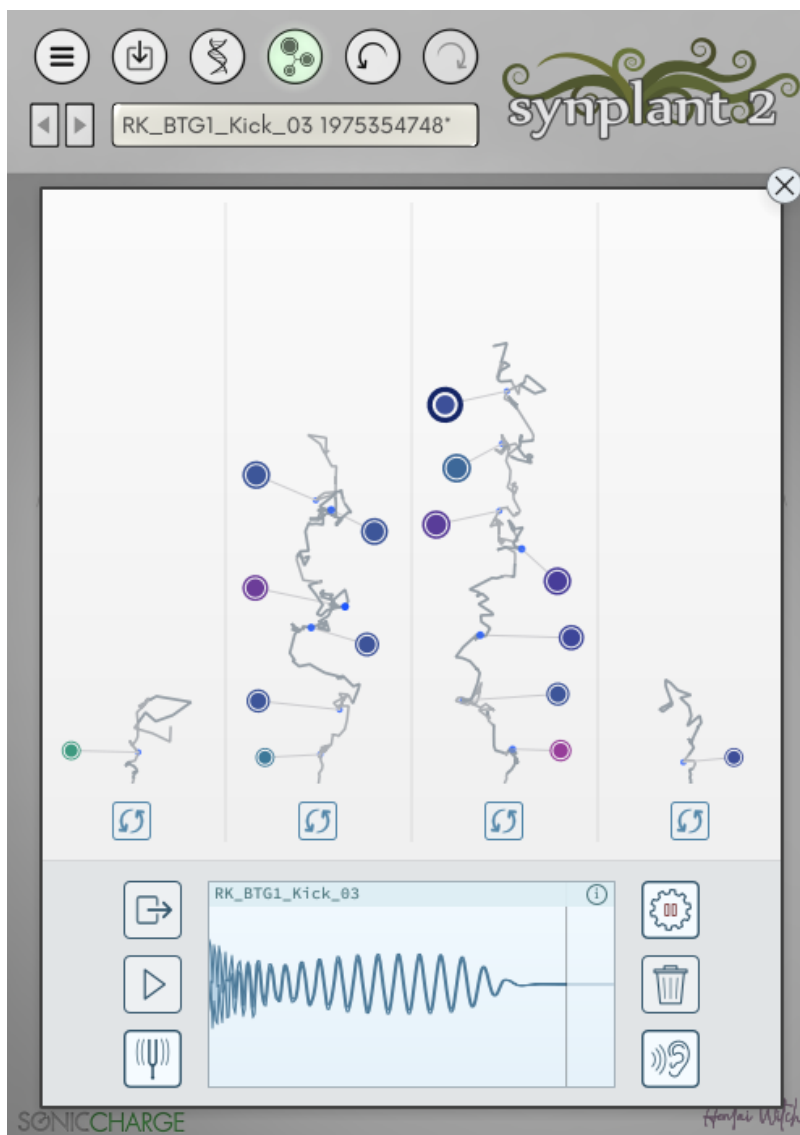


Figure 15: Genopatch

Replication begins by inputting a WAV or AIFF audio clip of any sample rate and bit resolution. Then, the part of the sample that is desired for replication can be selected. Up to 2 seconds can be highlighted. By clicking on the cogwheel, the application will commence.

From there, four different strains will visually rise, generating preset patches from the analysed sound (Fig. 15). These continue to grow over time, with each node representing a new patch. The strain height indicates the similarity to the analysed sample. Each of the four strains focuses on a different aspect of the original sample. Once they reach the top, they stop creating new patches, as they have reached maximum similarity. These patches either sound very close or even identical to the original or vastly different, depending on the sample. The computing is performed locally on the CPU rather than a cloud server.

Profound testing is carried out in Application of AI in Music Production.

### 5.1.1 Evaluation

**Performance:** Synplant 2 is one of two applications at the time of writing that are available for replicating sound design using AI algorithms. The patches created by genopatch range from nearly identical replica to indistinguishable, chaotic textures. For more simplistic sounds, the results are adequate. With complex or longer sounds, the outcome is often something undesired. Furthermore, the effects are lacking in quality and variety. The results showed repeatability.

**Usability:** While unconventional, Synplant 2's interface appears unintuitive on first impression. The main interface is unique and promotes an alternative approach to altering presets rather than a traditionally technical. Within the gene manipulation tab, the latter is possible as well, though not intuitively, as control parameters are partly animated, and the names of the control parameters require prior knowledge with synthesisers. The genopatch menu is presented in a simple and intuitive design. On a positive note, the plug-in works completely offline.

**Creativity and Flexibility:** Synplant 2's generated patches offer a variety of notably creative sound potential. Early generations of replicated sounds within genopatch often offer a midway between the original sample and the desired replica, featuring interesting and unique sonic possibilities. Once created, the patches can be further processed and manipulated with an array of synthesiser controls or even, if desired, sampled and put into genopatch again. Overall, the creative output, as well as the potential for creativity, are exceptional.

**Conclusion:** Synplant 2 proves to be a unique approach towards AI sound design and is one of the few plug-ins that use machine learning for creating innovative sounds. The genopatch feature shows immense potential, with lacking quality in effects and filters.

## 5.2 MicroMusic

MicroMusic is a VST plug-in created for a student project by students at the University of Waterloo. It promises to analyse samples, and create presets based on the input for Vital, a common and powerful wavetable synthesiser VST available for free (MicroMusic Inc. 2024). The interface of the application is solely utilised for inputting samples and choosing from a variation of the inherited AI model (Fig. 16). MicroMusic accepts mp3 and WAV-files. After the audio clip is dropped, a sound category can be chosen, i.e. "EDM Synth". Once an audio clip containing the desired sounds is prompted, the application will analyse it and create up to ten presets in a dedicated folder.

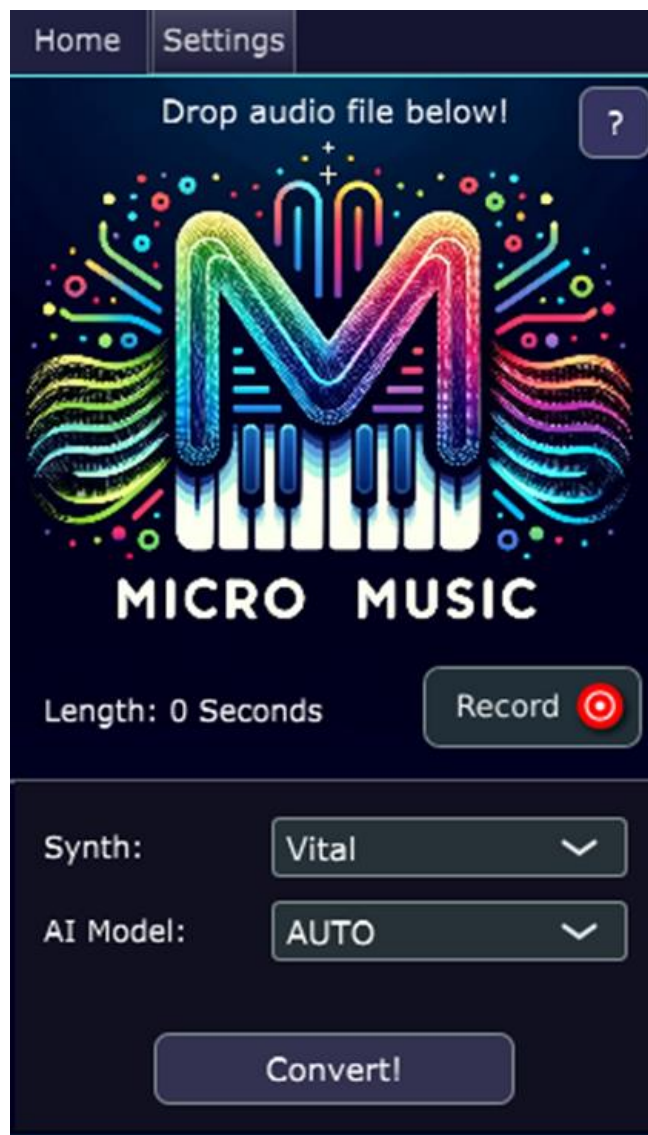


Figure 16: MicroMusic Menu

From this point forward, the sound design is performed within Vital. MicroMusic tries to replicate the sound design as similar as possible to the original sound.

Profound testing is carried out in Replicating Presets.

### 5.2.1 Evaluation

**Performance:** MicroMusic is one of two VST plug-ins that utilise AI algorithms to imitate sound design available on the market at the time of writing. It distinguishes itself from Synplant 2 by only carrying out the preset generating feature, while the rest of the sound design is done within an external synthesiser. This is advantageous, as plentiful resources on the external synthesiser exist. However, the generated presets delivered sub-par results. Repeatedly, presets generated similar sounding sounds from different source clips. Furthermore, considering the capabilities of Vital, the output presets appear severely lacking in details and prompt adherence. Merely simplistic sounds can be

imitated. Overall, while offering a unique approach towards sound design, the results deliver sub-par quality.

**Usability:** The user interface is intuitive and simple. Due to the lack of features apart from cloning, this was to be expected. However, the sheer number of presets created can get confusing and disorganised. Additionally, extra care should be put into naming conventions, as the possibility of overwriting existing presets in Vital arises.

**Creativity and Flexibility:** MicroMusic itself solely provides a limited number of settings, such as sound categories or different AI models. The generated presets lack the depth and design of the original audio clips. When changing settings, no significant differences could be noted. Overall, the presets borrow their flexibility and sound design capabilities from Vital but underutilise them.

**Conclusion:** MicroMusic appears to be revolutionising initially, until the generated output is examined. Nearly the entirety of generated presets sounded unoriginal until further designing was performed. The use of such technology arises the question whether the creation of AI-generated presets increases efficiency when compared to regular presets generated by more knowledgeable users or professionals. Learning the fundamentals of sound design remains a requirement to achieve a desired outcome.

## 6 Application of AI in Music Production

The previous chapters delved into the possibilities of AI-generated musical content and evaluated several applications capable of creating music for musicians and non-musicians equally. This chapter tries to experiment with the replication features of Synplant 2 and MicroMusic more profoundly, as well as the generated sample clips generated by the sound effects feature of ElevenLabs. The experiments focus on sound design replication and the employment of AI-generated audio content within a practical music production environment.

### 6.1 Replicating Presets

This experiment aims to derive synthesiser presets from sampled audio clips with specific characteristics commonly used in hard techno. All sampled sounds were created and specifically designed for this experiment by the author using the free wavetable oscillator synthesiser plug-in Vital. The BPM of the project was set to 140, and each sound type was played at the note C4. The sounds were recorded to WAV-files. To generate replica presets, Synplant 2 and MicroMusic were utilised. The generated patches were then curated, recorded, and loudness-normalised, to ensure comparability between the original and the replica. For a more comfortable listening experience, all samples were compiled into a demonstration file to be heard sequentially. Additionally, all presets can be found below and inspected or employed.

[AI Sound Replication Demonstration](#)

[Reference Presets](#)

[Synplant 2 Presets](#)

[MicroMusic Presets](#)

The use of headphones or a stereo speaker setup is recommended.

#### 6.1.1 Acid Sound with Delay

This preset resembles a modern low-passed acid sound, characterised by a squelching saw wave with an open resonant filter. It specifically utilises a ping pong delay, with the frequency set to  $\frac{1}{4}$  tempo for the left channel and  $\frac{1}{8}$  tempo for the right channel, to create a sense of wideness. Additionally, a longer attack time is employed to impart a softer feeling. Due to its filter-shaping nature, the initial pluck is slightly detuned to D#.

Synplant 2 Version:

- Detuning: It is also slightly detuned to D#.
- Metallic Quality: The sound exhibits a very metallic quality due to the reverb.
- Soundstage Characteristics: The sound itself is narrow, but not mono, with the reverb providing a wider spatial effect.
- Absence of Ping Pong Delay: There is no ping pong delay, resulting in a narrow soundstage.



- Squelching Quality: The squelching effect is present, but not clean due to the reverb.

Conclusion: The resulting sound is similar, but missing depth.

#### MicroMusic Version:

- Sound Quality: The sound is very plain.
- Soundstage Characteristics: The sound is wide.
- Absence of Ping Pong Delay: There is no delay present.
- Lack of Squelching: There is no squelching effect at all.
- Acid Characteristics: The sound is not readily recognisable as acid.

Conclusion: The sound quality is subpar and deviates significantly from the original.

### **6.1.2 Acid Sound Classic, Squelched**

This preset resembles a more classic band-passed acid sound, characterised by increased squelching with an open resonant filter. The attack time is set to 0 milliseconds to ensure a snappy transient. This rendition of the classic sound does not incorporate delay, yet it still maintains a wide spatial effect.

#### Synplant 2 Version:

- Sound Accuracy: The sound itself is depicted accurately.
- Soundstage Characteristics: The soundstage is very narrow, almost mono.
- Reverb Positioning: The reverb is positioned too far to the right.

Conclusion: The sound is recognisable and replicated accurately with lacking soundstage characteristics.

#### MicroMusic Version:

- Sound Accuracy: The sound is not recognisable as an acid sound.
- Channel Imbalance: The left channel is too loud, while the right channel is too quiet.
- Asymmetry: The sound is very uneven and asymmetrical.
- Simplicity: The waveform is a simple saw wave, resulting in a plain sound.
- Reverb Positioning: The reverb is positioned too far to the right.

Conclusion: The sound deviates significantly from the original and is of poor quality.

### **6.1.3 Classic Warm Synth Pad**

This preset resembles a warm, analogue-like synth pad/lead. Created from a single saw wave with added brown noise. Character is imparted through the use of chorus, a ping pong delay, and saturation, culminating in reverb. The sound is very wide, primarily due to the delay.

#### Synplant 2 Version:

- Sound Accuracy: The sound is very recognizable.
- Soundstage Width: The sound is wide, though not as wide as the original.
- Delay Functionality: The delay does not function as intended.

Conclusion: Without the delay, the sound would be usable and remains very recognisable.

MicroMusic Version:

- Sound Accuracy: The sound is too different from the original.
- Soundstage Characteristics: The sound is not as wide and imbalanced.
- Absence of Delay: There is no delay present.
- Warmth: The sound is not as warm.

Conclusion: While it functions as a pad, it does not serve well as a replica.

#### 6.1.4 Ominous Pad

This preset resembles a dreamy, dark pad characterised by a moving and evolving feel, achieved through the use of chorus, flanger, and phaser effects. The sound is rendered very wide by the application of ping pong delay. Saturation and noise contribute to a powerful and warm sound.

Synplant 2 Version:

- Sound Accuracy: The sound is generally recognisable and accurate.
- Duration Limitation: The sample is too short, due to the 2-second limitation for samples in genopatch.
- Soundstage Width: The sound is slightly less narrow.
- Absence of Delay: The delay is missing.
- Reverb Intensity: The reverb is too strong.

Conclusion: The result is close to the original but lacks depth.

MicroMusic Version:

- Sound Accuracy: The sound resembles that of a flute.
- Lack of Dynamics and Width: The sound is not dynamic and wide enough.
- Simplicity: The pad is very basic.
- Absence of Delay: There is no delay present.

Conclusion: The sound does not closely resemble the intended original.

#### 6.1.5 Dystopian Horn Sound

This preset resembles a dark and menacing cinematic horn sound. It features a pitch-up effect and is warmly distorted. The sound incorporates chorus, saturation, and reverb to enhance its character.

Synplant 2 Version:

- Sound Accuracy: While not identical, the sound is very recognisable.
- Soundstage Characteristics: The soundstage is not as wide as the original sample.
- Sustain Length: The sound has a longer sustain than the original sample.

Conclusion: Although not identical, it is a competent replica.

MicroMusic Version:

- Sound Accuracy: The sound is too muffled and plain.
- Absence of Pitch-Up Effect: The pitch-up effect, which is crucial for the horn sound, is missing.
- Soundstage Width: The sound is sufficiently wide.
- Design Choice: The application of LFO on pitch is a positive enhancement.

Conclusion: The sound is not identical and lacks the menacing quality of the original.

### 6.1.6 Electric Pad

This preset resembles a modern pad with an electric feel. It features a detuned, FM-modulated sine wave combined with brown noise and white noise, processed through chorus, delay, reverb, and compression. The result is a wide-spread and dynamic atmosphere.

Synplant 2 Version:

- Sound Accuracy: The sound is more diffuse and noisier.
- Soundstage Characteristics: The sound is almost as widespread as the original.
- Reverb Quality: The reverb is too metallic.
- Envelope Characteristics: The decay is longer, but the sustain is shorter.
- Absence of Delay: There is no delay present.

Conclusion: A competent replica, but it lacks depth.

MicroMusic Version:

- Sound Accuracy: The sound is missing the electric feel entirely and resembles a horn more than the intended pad.
- Insufficient Dynamics: The sound is not dynamic enough.
- Absence of Delay: There is no delay present.
- Lack of Depth: The sound lacks depth.

Conclusion: The sound deviates too far from the original.

### 6.1.7 Open Filter Bass Pluck

This preset resembles a heavy, explosive bass pluck that transitions from very wide to narrow asymmetrically. It features a distorted, modern metallic sound with asymmetrical stereo width achieved by chorus and delay.

Synplant 2 Version:

- Sound Accuracy: The sound is close to the original but has a more pronounced low-end, resulting in a boomy quality.
- Soundstage Characteristics: The sound is wider than the original, but symmetrical.
- Hollow Sound: The reverb contributes to a hollow sound.

Conclusion: The sound is recognisable but lacks the original's impact.

MicroMusic Version:

- Sound Accuracy: The sound is a basic bass pluck, very plain, and missing the metallic quality. It lacks the explosive feeling
- Soundstage Characteristics: It is missing the rapid stereo width change.
- Absence of Distortion: The sound is missing distortion.

Conclusion: The sound does not demonstrate low quality per se, but appears too basic to be considered a replica.

### 6.1.8 Dissonant Screech Sound

This preset resembles a modern screech lead sound, typical for Hard and Industrial techno. It features a heavily detuned saw wave with pitch modulation, processed through chorus, distortion, delay, reverb, and compression.

Synplant 2 Version:

- Sound Accuracy: The sound is easily recognisable.
- Stereo Characteristics: The soundstage is very narrow.
- Reverb Quality: There is more white noise present in the reverb.

Conclusion: The sound is less nuanced but nonetheless a competent replica.

MicroMusic Version:

- Sound Accuracy: The sound is not sufficiently detuned.
- Volume Level: The sound is far too quiet.
- Absence of Distortion: There is no distortion present.
- Potential: It serves as a good starting point if the user knows how to make adjustments.

Conclusion: The preset is inadequate in its default state but can be improved with appropriate modifications.

### 6.1.9 Dark Plucky Stab

This preset resembles a modern, plucky, menacing techno stab, characterised by ping pong delay and reverb. It is created using an FM-modulated saw wave combined with brown noise, processed through saturation and compression.

Synplant 2 Version:

- Sound Accuracy: The sound is very blurry.
- Soundstage Characteristics: The soundstage is narrower.
- Reverb Quality: The reverb is wide but metallic.
- Percussive Quality: The sound is more percussive and lower in pitch.
- Absence of Ping Pong Delay: There is no ping pong delay present.

Conclusion: The sound deviates significantly from the original and lacks precision.

MicroMusic Version:

- Sound Accuracy: The sound is a very basic pluck, too plain.
- Percussive Quality: The sound is too soft and not percussive enough.
- Envelope Characteristics: The sustain is too long.

- Absence of Delay: There is no delay present.

Conclusion: The sound resembles a pad and is not recognisable as the intended preset.

### 6.1.10 Complex Space FX

This preset resembles a modern trippy effect with complex modulation, designed to test the limits of genopatch. It features a detuned wavetable FM-modulated by a saw wave, processed through chorus, compression, and an LFO-modulated delay. Additionally, it incorporates an LFO-modulated comb filter and another LFO-modulated low pass filter. The result is a spacey movement effect with a very wide and asymmetrical soundstage.

#### Synplant 2 Version:

- Sound Accuracy: The sound is significantly simpler, resembling a riser more than a movement effect.
- Stereo Characteristics: The soundstage is narrower in comparison.
- Reverb Quality: The reverb is too metallic.
- Absence of Delay: The sound lacks delay, resulting in missing depth.
- Excessive Distortion: There is too much distortion present.

Conclusion: The sound is not nearly as complex as the original and has the wrong reverb.

#### MicroMusic Version:

- Sound Accuracy: The sound is completely different and not comparable.
- Sound Type: The result is not an effect, but rather a synth lead.

Conclusion: The sound was too complex to be replicated accurately, resulting in a sub-par attempt of a replica.

### 6.1.11 Dissonant Synth Stab

This preset resembles an extremely dissonant plucky synth sound. It features a detuned saw wave FM-modulated by pink noise, processed through chorus, delay, and saturation, culminating in reverb. This preset is used as the main synth in the reference song "NO PAIN NO PAIN".

#### Synplant 2 Version:

- Sound Accuracy: The sound is very weak and dry, making it unrecognisable.
- Stereo Characteristics: The soundstage is very narrow in comparison.
- Absence of Delay: There is no delay present.

Conclusion: The sound is not well replicated.

#### MicroMusic Version:

- Sound Accuracy: The sound is bland, resembling a basic pluck sound.
- Sound Type: The sound is closer to a bass pluck.
- Absence of Delay: There is no delay present.

Conclusion: The sound is too bland to be recognisable as a replica.

### 6.1.12 Summary / Evaluation

When examining all three of the different samples per category, it is evident that the AI-generated versions almost always lack depth and clarity compared to the original sound design. For Synplant 2, this deficiency is often attributed to the lack of adequate delay and reverb, as well as the narrower stereo characteristics demonstrated. Generally, the sounds created by Synplant 2 are richer in detail than those produced by MicroMusic, and often achieve a quality closer to the original. However, the delay was never replicated sufficiently due to the absence of a ping pong delay feature within Synplant 2. While sound characteristics such as delay and reverb can be added post-production, this approach disregards the importance of their position within the effect chain. Furthermore, the reverb effect employed by Synplant 2 exhibits excessively metallic characteristics, resulting in a hollow sound quality for many of the replicated sounds. Additionally, the analysis of static samples eliminates the possibility of manipulating the original sounds throughout the process. Nonetheless, due to its branch system and the multitude of generated versions within genopatch, Synplant 2 merits its own place in the realm of sound design.

In stark contrast lie the results produced by MicroMusic. Despite its potential to recreate the exact same presets due to using the same synthesiser as the original sounds, this potential is rarely utilised. Most of the created presets fall into a category of unoriginal, basic sounds within the electronic music genre, deviating significantly from the original designs. This is even noticeable visually when comparing the waveforms (Fig. 17).

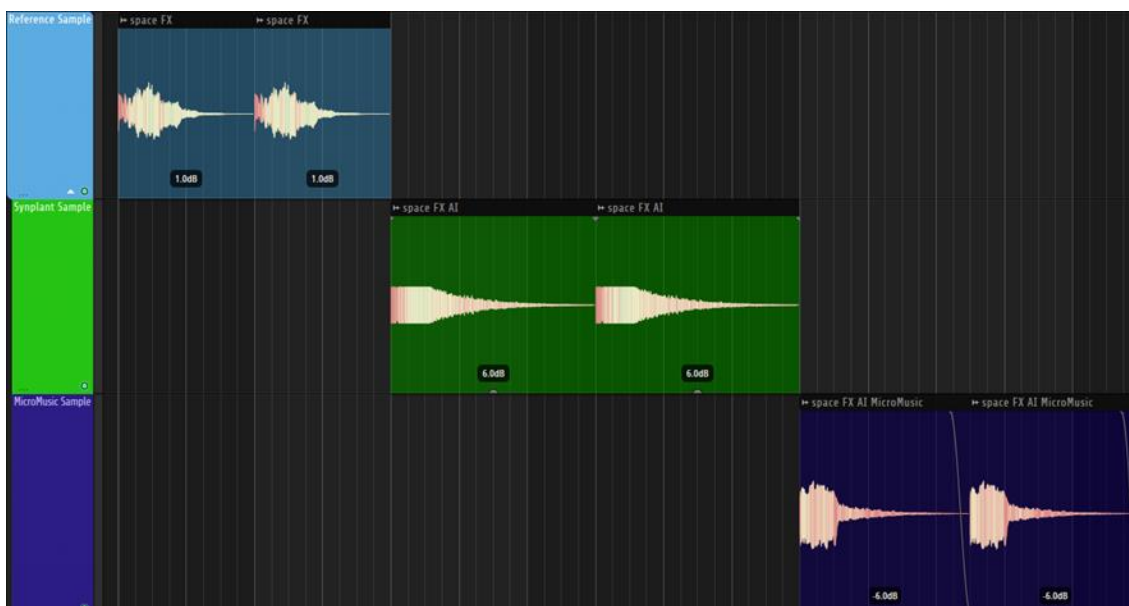


Figure 17: Visual Comparison between Space FX Samples

Although these presets can be adjusted to resemble the original sounds more closely, it raises the question of the utility of such a tool. Given the wide availability of Vital, a plethora of presets created by artists can be found online, often free and of higher quality. Additionally, the option to purchase preset packs exists. Long-term, learning the

fundamentals of sound design proves more beneficial to amateurs, allowing to create own unique sounds and express individuality over time.

To summarise, Synplant 2 provides interesting and often close-to-original replica sounds, albeit with missing depth, but offers unique ways to shape the sounds through its organic design features. MicroMusic, on the other hand, fails to replicate even very basic sounds and may produce results that are more lacklustre than simply using pre-made presets.

## 6.2 Replacing Samples with AI-Generated Samples

This experiment tried to explore the idea of recreating and replacing existing original sounds through AI in the context of a piece of music. The sounds were replicated using the genopatch tab within the Synplant 2 VST plug-in described in Synplant 2. For this purpose, all sounds of an existing song were resampled and replaced by AI-generated replica sounds. To show the grade of alteration by processing, this experiment was performed using four different approaches. Version 1 and 2 were processed through the same mixer channels and effects as the originals.

- **Version 1:** Unprocessed samples replicated, then processed
- **Version 2:** Unprocessed samples replicated, optimised, then processed
- **Version 3:** Processed samples replicated, then put into mixer
- **Version 4:** Processed samples replicated, optimised, then put into mixer

The goal was to replicate the sound design choices as accurately as possible, and to determine to what extent added effects on the mixer channel influence the AI outcome. The test sounds were designed and chosen by the author, to ensure the best possible evaluation. The same procedure was used for every sample. The replicated sounds were normalised and brought to the same loudness level, as to ensure that loudness differences do not change the perception of the sound (Boer 2024). The option to replicate in tune with the original sample was always enabled.

When replicating unprocessed samples, i.e. without mixer inserts, the replicated samples were routed to the same mixer channel afterwards to ensure that they were processed in the same manner as the original sounds. As Synplant 2 is not capable of replicating voice samples, the only unaltered sample within all the examples was the voice sample, which is recorded spoken word through a distortion plug-in. Except for the drums, all sounds were created with Vital.

The chosen reference song is called “NO PAIN NO PAIN” and was created to be an intense, energetic, dark hard techno track with industrial influences. It was chosen for its lack of dynamic development and simple arrangement, as dynamics prove to be difficult to replicate due to Synplant 2’s limited duration for input samples of 2 seconds. The generated patches were curated, and the most similar was chosen.

Original Song Reference: [NO PAIN NO PAIN](#)

### 6.2.1 Evaluation

#### Version 1: Unprocessed samples replicated, processed through original effect channels

This Version employs samples and patches that were cloned from the original samples before further processing. The resulting AI-generated samples were then put into the same arrangement position as the originals and processed equally.

Reference: [AI Version 1](#)

Evaluation:

It is immediately noticeable that the kick and synth sound different from the original. The entire song lacks rhythmic aspects because the AI-generated kick exhibits significantly less complexity. Similarly, the created top loop and rhythmic groove are lacking. On a positive note, the percussion and synth stab were very accurately reproduced, which could also be due to the processing. Additionally, a hissing sound accompanies the main synth lead throughout the song and the reverb effect appears too metallic.

#### Version 2: Unprocessed samples replicated, optimised, processed through original effect channels

This Version employed samples and patches that were cloned from the original samples before further processing. The resulting AI-generated samples were then adjusted within Synplant 2, and then put into the same arrangement position as the originals and processed equally.

Reference: [AI Version 2](#)

Evaluation:

It is immediately noticeable that the synth sounds more akin to the original synth lead, but is lacking in complexity and depth, nonetheless. While the kick improved in punchiness, it did not gain complexity. Again, the created top loop and rhythmic groove are lacking. Furthermore, the hissing sound accompanies the main synth lead again and the mix sounds empty.

#### Version 3: Processed samples replicated, then put into empty mixer channels

This Version employed samples and patches that were cloned from the original samples after processing. The resulting AI-generated samples were then put into the same arrangement position as the originals and remained unprocessed.

Reference: [AI Version 3](#)



**Evaluation:**

It is immediately noticeable that the synth sound changed again, but is lacking in complexity and depth, nonetheless. It appears to be more detuned and less transient. While the kick improved in punchiness, it did not gain complexity. Again, the created top loop and rhythmic groove are lacking. Furthermore, the synth stab now demonstrates more percussive and transient characteristics. On a positive note, the hissing sound is absent in this version. Overall, it shows a reduction in quality compared to Version 1 and 2.

**Version 4: Processed samples replicated, optimised, then put into empty mixer channels**

This Version employed samples and patches that were cloned from the original samples after processing. The resulting AI-generated samples were then adjusted within Synplant 2 and put into the same arrangement position as the originals and remained unprocessed.

Reference: [AI Version 4](#)

**Evaluation:**

It is immediately noticeable that the synth sound is more transient but lacking in complexity and depth, nonetheless. It represents the closest to the original, however. The kick improved in punchiness and slightly in complexity due to an added off beat hit. Again, the created top loop and rhythmic groove are lacking. Furthermore, the synth stab still demonstrates more percussive and transient characteristics than the original. Additionally, the hissing sound is absent in this version. Overall, a miniscule improvement over Version 3.

**6.2.2 Conclusion**

Replicating existing sound design through the aid of AI seems tempting and holds the possibility to generate impressive results with singular, uncomplex audio material. However, as evidenced by the lack of depth and complexity in the AI-versions of “NO PAIN NO PAIN”, complex and layered sounds such as hard techno rumble kicks or loops prove difficult to replicate (Fig. 18).

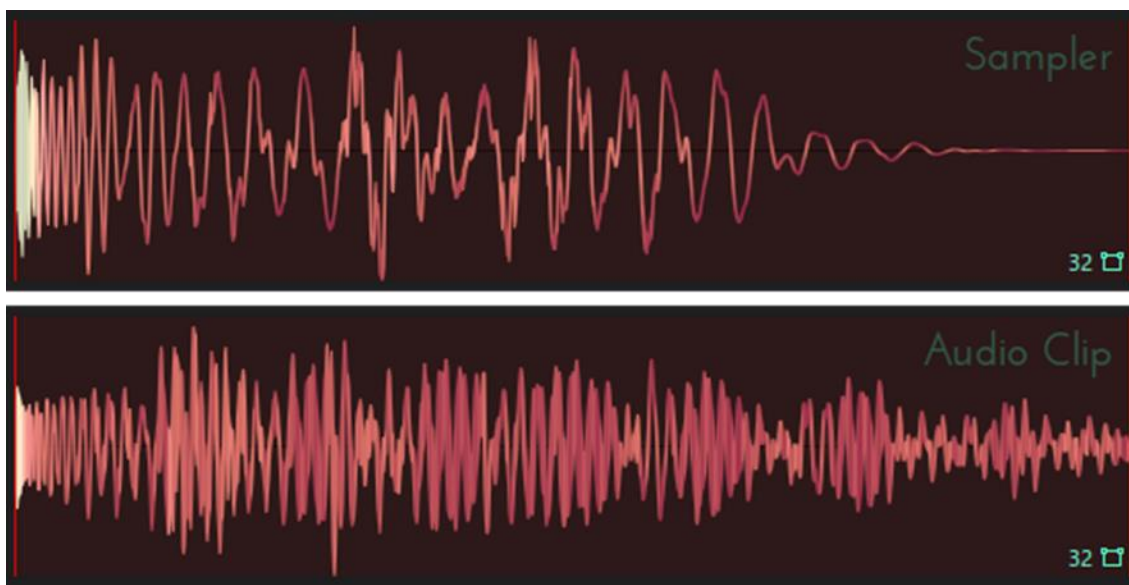


Figure 18: Comparison Kick Samples; above Original, below Synplant Replica

How much of this was caused by the sample length limit of 2 seconds remains unclear. Generally, audio samples with higher percentage of noise were replicated more efficiently. Additionally, rhythmic features of songs are lost in AI replication as well. However, it should be noted that the generated patches came close to the original with some sounds, which brings the question forward who constitutes the target audience for such features within a synthesiser. Finally, the framework conditions of this experiment are somewhat removed from real-world applications, as obtaining individual samples for songs and, even more so, individual mixer channels, presents significant challenges.

### 6.3 Producing with AI-Generated Samples

This experiment tried to explore the utilisation of AI-generated samples in a practical application within a music production environment.

Production 1: [Looped Memories](#)

Production 2: [WELCOME /// FUTURE](#)

The idea was to produce a complete song in the genre of the prompts – hard acid techno - using AI-generated samples. The samples were generated by ElevenLabs during the testing phase, as detailed in Chapter Testing. Firstly, the AI-generated samples were compiled into a sample pack and categorised by sound type. For artistic integrity, only select sounds from the pack were used to ensure the resulting compositions were genuine artistic creations rather than a technical demonstration trying to show every generated sound.

#### 6.3.1 Production 1

This production exclusively employed the use of loops generated by ElevenLabs, to explore potential results achievable with minimally altered AI-content. The loops were

arranged and mixed in a manner consistent with processing regular samples. At least one loop was utilised from each sound category. The goal was to produce hard techno with acid influences, however the desired genre characteristics proved difficult to achieve without such sounds. Consequently, the resulting composition deviates from the intended genre and presents a less intense composition.

Mixing and mastering were performed manually, with the entire production process (excluding sample generation) taking about 3 hours. This duration is exceptionally short and reflects the simplicity of the arrangement, as well as the experiment-given restrictions.

### 6.3.2 Production 2

For this production, a more liberal and creative approach was taken, and only two loops were used. The samples underwent various transformations, including pitch adjustments, cutting, warping, and tempo alignment. These modifications were essential to integrate the samples into the desired genre. Given that Hard techno often features harsher sounds, the quality limitations of the samples at 128 kbps were less detrimental compared to genres where high-fidelity audio is critical.

Due to the static nature of samples, the song's arrangement emphasised repetitive, percussive, and industrial elements, avoiding reliance on evolving controllable parameters. A manipulated screech sound, copied several times and fused back together, served as the lead melody, after manipulating its pitch and time settings. Multiple hi-hat and percussion samples were combined to create a top loop from scratch, although two AI-generated loops, a hi-hat loop and a percussion loop for the break section starting at 01:42, were implemented.

A high-pass-filtered acid bass line was designed to complement the lead melody. This bass sound, audible from 02:50 onward, was created by manipulating an acid stab loop from the sample pack. Adjustments to the ADSR envelope settings and sixteenth-note arrangements, along with heavy distortion and compression, achieved the desired acid sound. Despite these efforts, replicating the dynamic squelching effect proved challenging with static samples, and thus, it was omitted from the final composition.

Tension and release were established by strategically cutting frequencies in specific sections and employing build-up and drop sections. Automation of stereo width and common effects such as delay, and reverb further enhanced the track. The phrase "welcome to the future" was deliberately used to set the theme by using one of the AI-generated voice lines and sending them through a vocoder and saturation.

Mixing and mastering were performed manually, with the entire production process (excluding sample generation) taking about 8 hours. This duration is relatively short, reflecting the song's simplicity, which is not atypical for shorter compositions.

#### 6.3.2.1 Limitations

While working with the samples, the lack of quality was noticeable throughout the production. Admittedly, as techno can be generated entirely digitally, no strict limitations

regarding quality apply. Thus, a qualitatively sub-par sample can achieve a desired sound profile for the genre. The most noticeable issues were the lack of precision in the overhead section, such as hi-hats, and the sub-par effect samples, which did not demonstrate significant usefulness compared to non-AI samples. For voice samples, there was almost no discernible difference from real human voices, as they were processed to de-humanise them further. The significant change in timbre resulting from this is desirable, effectively masking any deficiencies in the original samples. Finally, the lack of authentic acid sounds resulted in the creation of different genres. Overall, while the quality loss remains apparent, these samples can be useful when processed appropriately according to genre-specific requirements.

## 7 Conclusion and Recommendation

This study analysed the current state of AI-generated content and its sound design within the framework of the electronic music genre of techno, providing a detailed assessment of its quality. The objective was to determine how it compares to human-made design and quality, and to what extent AI-generated audio content can be integrated into music production. Expert interviews were conducted and systematically evaluated to excerpt valuable knowledge from them. The study also examined the extent to which techno sound design can already be replicated by AI and utilised in a music production environment.

The results show that AI-generated content varies significantly in quality between different models and even generations. While some AI applications, such as Suno and Udio, are already capable of generating and expanding entirely arranged songs, others struggle to produce anything more than generic drum loops with low sample quality and bad mixing. Initially, the ability of these models to generate entire songs is impressive. However, upon closer examination, it becomes evident that the prompt adherence desired by producers is not yet achieved and the sound characteristics often generic sounding. This is particularly concerning given that AI models primarily base their outputs on datasets of successful musical works they were trained on, casting doubt on their ability to generate truly original content. It is also noted that, with rare exceptions, these models are unable to accurately implement defining aspects from the given prompts, even when provided with detailed descriptions.

Furthermore, the models lack flexibility regarding the influence on individual tracks or instruments, which hinders musically inexperienced users from fully realising their creative ideas, despite the effect of democratisation. As a result, without further creative processing, no original pieces can be created from purely generative AI models. The interviewees also confirmed that AI lacks its own identity.

Additionally, there are a few applications that can be specifically useful to producers. For example, ElevenLabs' sound effect generation feature produces surprisingly high-quality samples for certain sound types that adhere to the given prompt in terms of tempo, key, and sound characteristics. Particularly, the generation of percussive loops demonstrated remarkable quality and prompt adherence, despite the limited information density given in the prompts. This ability to generate tailored samples, including mostly convincing vocal samples, can lead to increased efficiency and creative inspiration in music production, analogue to human-produced samples. However, the generated content still exhibits audible acoustic artifacts, limiting its potential use to more aggressive or lo-fi genres, as they would otherwise stand out negatively in the production. In practical application, it was found that the audio content can be integrated into projects with existing production knowledge without much effort, though the quality of the result has not yet reached the level of comparable professionally produced works created without AI-generated content.

Attention should especially be given to innovative tools related to sound design that allow existing sounds to be replicated and transformed into manipulable presets for synthesizers. Plug-ins like Synplant 2 and MicroMusic aim to increase efficiency, simplify, and enhance the sound design process. This could assist inexperienced producers in efficiently achieving their desired sound. Particularly, Synplant 2 not only offers the most accurate replication of given sounds but also provides the capability to generate unique new sounds through the creation of various in-between-presets, making it a versatile tool for sound design in general.

It is important to emphasise that the art and the process of musical sound design cannot be replaced by these applications, nor can they serve as a substitute for learning fundamental techniques. AI-replicated sound design exhibits significant quality fluctuations and never achieves the depth and detail of human-created sound design. Currently, replicating dynamically rhythmic and complex sounds is not possible with these AI models. The experiments demonstrated that replicated sounds alone do not constitute coherent musical pieces. However, these applications can be excellently employed for creative inspiration or for unconventional processing of existing sounds, particularly Synplant 2, which significantly outperforms MicroMusic in terms of quality.

The use and distribution of purely AI-generated audio content remains a topic that must be critically examined from both ethical and legal perspectives. According to expert interviewees, the use of AI-generated samples and MIDI sequences is to be seen as unproblematic, both legally and in terms of authenticity. The comparison to traditional sampling was frequently mentioned, as was the role of the producer as a curator of sounds. It was particularly emphasised that humans remain essential for the creation of musical works and represent the most important factor. Therefore, purely generative songs can never exhibit true originality and should not be considered creative or authentic works.

However, it is precisely these more advanced applications whose generated content and datasets need to be scrutinised more closely from a moral perspective. The significant advancements in generative AI music raise numerous questions regarding ethics and regulation. Protecting individual human artists is of fundamental importance. This has recently been highlighted by various copyright lawsuits filed by major music labels against the developers of the AI models Suno and Udio, as well as an open letter signed by over two-hundred artists.

The interviewed Experts see the future of AI-generated music less in the realm of entertainment music, although its quality has significantly improved in recent months. Instead, application areas are frequently cited in the context of video platforms and other multimedia content provided by companies, which could use AI to create cost-effective and customised background music. Nevertheless, the individual person creating the music will remain the most significant factor, and, therefore, the influence of live music is expected to increase.

Given the rapid development in recent years, it must be assumed that the currently acceptable quality of AI-generated content will continue to improve. This will also lead to

higher prompt adherence, which is still noticeably lacking at present. Furthermore, the interviewed experts agree that the quantity of music will increase, but the quality thereof is likely to remain average. However, experienced producers could further enhance the quality of their productions by incorporating AI-generated audio content, provided its quality improves further. So far, AI applications have primarily been used for analysis and MIDI note generation, making the field of sound design a relatively unexplored area of music production. Whether AI will be able to produce music entirely independently of any human influence remains uncertain—this is a topic of general AI. In the future, quality, full integration into DAWs, and the agentification of AI models will play a significant role, while further democratisation of music production is to be expected.

This work has demonstrated that the creation of music by artificial intelligence with acceptable quality is no longer a distant future scenario. Rather, AI's relevance to music production is not limited to experienced producers, but also made its way into the broader public domain, catering to non-musicians.

Technology is constantly evolving and has historically led to the democratisation and creation of new music cultures and genres, as seen with the Roland TB-303 bass synthesiser and the emerge of techno in general. This progress has now advanced to algorithms and deep learning, offering the potential to make music creation and individual expression even more accessible to non-musicians. Therefore, the development and use of AI in music production should be approached with an open mindset. Producers should engage with this technology, as it holds significant potential for the creation of new art. However, legislation on the training and use of generative AI models are equally as important. Nonetheless, it is certain that the human element will remain the most important factor in the creation thereof.

## References

- A Brief History of Computer Music (1994) – Eric Kuehnl (2024). Available online at <https://erickuehnl.com/1994/12/30/a-brief-history-of-computer-music/>, updated on 8/4/2024, checked on 8/4/2024.
- Artist Rights Alliance (2024): 200+ Artists Urge Tech Platforms: Stop Devaluing Music. In *Medium*, 1/4/2024. Available online at <https://artistrightsnow.medium.com/200-artists-urge-tech-platforms-stop-devaluing-music-559fb109bbac>, checked on 8/13/2024.
- Audiobox | Meta FAIR (2024). Available online at [https://audiobox.metademolab.com/capabilities/text\\_to\\_audio](https://audiobox.metademolab.com/capabilities/text_to_audio), updated on 7/29/2024, checked on 8/6/2024.
- Avid (2024): Guide to Using Samples in Music & Digital Audio Production. Available online at <https://www.avid.com/resource-center/music-sampling-guide>, updated on 11/8/2024, checked on 11/8/2024.
- Baetz, Klaus (2023): Sounddesign: Wir bauen eine 303-artige Bassline | SOUND & RECORDING. Available online at <https://www.soundandrecording.de/tutorials/sounddesign-wir-bauen-eine-303-artige-bassline/>, updated on 8/12/2024, checked on 8/12/2024.
- Boer, Martin de (2024): FabFilter Learn - The science of sound - Perception of frequency and loudness. Available online at <https://www.fabfilter.com/learn/science-of-sound/perception-of-frequency-and-loudness>, updated on 8/16/2024, checked on 8/16/2024.
- Brown, Oliver (2021): Handbook of Artificial Intelligence for Music. Foundations, Advanced Approaches, and Developments for Creativity. With assistance of Eduardo Reck Miranda. Cham: Springer International Publishing AG. Available online at <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6676030>.
- Bundesverband Musikindustrie e. V. (2022): Musikindustrie\_in\_Zahlen\_2021\_E-Paper, Article 978-3-947253-03-6. Available online at [https://www.musikindustrie.de/fileadmin/bvmi/upload/06\\_Publikationen/MiZ\\_Jahrbuch/2021/Musikindustrie\\_in\\_Zahlen\\_2021\\_E-Paper.pdf](https://www.musikindustrie.de/fileadmin/bvmi/upload/06_Publikationen/MiZ_Jahrbuch/2021/Musikindustrie_in_Zahlen_2021_E-Paper.pdf).
- Coldewey, Devin (2022): Try ‘Riffusion,’ an AI model that composes music by visualizing it. In *TechCrunch*, 12/16/2022. Available online at <https://techcrunch.com/2022/12/15/try-riffusion-an-ai-model-that-composes-music-by-visualizing-it/>, checked on 7/21/2024.
- ElevenLabs (2024): We're building AI to shape the future of digital communication, ensuring content is understood by everyone, everywhere. | ElevenLabs, updated on 7/23/2024, checked on 7/23/2024.
- Exclusive Royalty Free Stock Music | Shutterstock (2024), updated on 8/16/2024, checked on 8/16/2024.



Fabian J. G. Westerheide (2017): Antwort auf den Fragenkatalog für das Fachgespräch zum Thema „Künstliche Intelligenz“ des Ausschusses Digitale Agenda. Available online at <https://www.bundestag.de/resource/blob/498712/7a03d0356e35cf64b26bbedc12e8a56d/a-drs-18-24-130-data.pdf>, checked on 8/6/2024.

Festival Fire (2022): FF-GOING-OUT.SIDE-Studie-2022. Available online at <https://www.festivalfire.de/15-000-festival-insights-in-der-going-out-side-studie-2022/>.

Gensch et al. (2009): Musikrezeption, Musikdistribution und Musikproduktion. Der Wandel des Wertschöpfungsnetzwerks in der Musikwirtschaft. Wiesbaden: Gabler Verlag / GWV Fachverlage GmbH Wiesbaden (SpringerLink Bücher).

Google Cloud (2024): What Is Artificial Intelligence (AI)? | Google Cloud. Available online at <https://cloud.google.com/learn/what-is-artificial-intelligence>, updated on 8/5/2024, checked on 8/5/2024.

Hight, Jewly (2024): AI music isn't going away. Here are 4 big questions about what's next. In *WPLN News | Nashville Public Radio*, 4/25/2024. Available online at <https://wpln.org/post/ai-music-isnt-going-away-here-are-4-big-questions-about-whats-next/>, checked on 1/6/2024.

Horea, Tiki (2023): Fixing phase issues. In *Sonarworks Blog*, 3/17/2023. Available online at <https://www.sonarworks.com/blog/learn/fixing-phase-issues>, checked on 12/8/2024.

Hua, Jane (2018): The Impact of Technology on the Musical Experience. In *www.music.org*, 7/31/2018. Available online at [https://www.music.org/index.php?option=com\\_content&view=article&id=2675:the-impact-of-technology-on-the-musical-experience&catid=220&Itemid=3665](https://www.music.org/index.php?option=com_content&view=article&id=2675:the-impact-of-technology-on-the-musical-experience&catid=220&Itemid=3665), checked on 8/4/2024.

IBM. IBM's computer checkmated a human chess champion in a computing tour de force (2024). Available online at <https://www.ibm.com/history/deep-blue>, updated on 8/8/2024, checked on 8/8/2024.

javatpoint (2024): Expert Systems in Artificial Intelligence - Javatpoint. Available online at <https://www.javatpoint.com/expert-systems-in-artificial-intelligence>, updated on 8/8/2024, checked on 8/8/2024.

Kaiser, Robert (2021): Qualitative Experteninterviews. Konzeptionelle Grundlagen und praktische Durchführung. 2., aktualisierte Auflage. Wiesbaden: Springer VS (Elemente der Politik).

Keeley, Ethan (2018): Masteringbox. Mixing in The Box: Tips for Great Sounding Mixes Without the Hardware - MasteringBOX. Available online at <https://www.mastering-box.com/learn/mixing-in-the-box>, updated on 11/8/2024, checked on 11/8/2024.

- Klawonn, Thilo (2023): Künstliche Intelligenz, Musik und das Urheberrecht. Edited by Burkhard Nonnenmacher. Erscheinungsort nicht ermittelbar: Mohr Siebeck. Available online at <https://directory.doabooks.org/handle/20.500.12854/133165>.
- Lauder, Ed (2017): Aiva is the first AI to Officially be Recognised as a Composer. In *AI Business*, 10/3/2017. Available online at <https://aibusiness.com/verticals/aiva-is-the-first-ai-to-officially-be-recognised-as-a-composer>, checked on 8/6/2024.
- LimeWire (2024): LimeWire. Available online at <https://limewire.com/>, updated on 7/31/2024, checked on 7/31/2024.
- Magnus; Lidström, Fredrik (2024): Synplant 2. Available online at <https://son-iccharge.com/synplant>, updated on 1/8/2024, checked on 1/8/2024.
- Mayring, Philipp (2014): Qualitative content analysis: theoretical foundation, basic procedures and software solution. Available online at [https://www.ssoar.info/ssoar/bitstream/handle/document/39517/ssoar-2014-mayring-qualitative\\_content\\_analysis\\_theoretical\\_foundation.pdf](https://www.ssoar.info/ssoar/bitstream/handle/document/39517/ssoar-2014-mayring-qualitative_content_analysis_theoretical_foundation.pdf).
- Mayring, Philipp (2023): Einführung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken. 7., überarbeitete Auflage. Weinheim, Basel: Beltz. Available online at [https://www.content-select.com/index.php?id=bib\\_view&ean=9783407296023](https://www.content-select.com/index.php?id=bib_view&ean=9783407296023).
- McKee, Ruth; Jamie Grierson (2017): Roland founder and music pioneer Ikutaro Kakehashi dies aged 87. In *The Guardian*, 2/4/2017. Available online at <https://www.theguardian.com/music/2017/apr/02/roland-founder-and-music-pioneer-ikutaro-kakehashi-dies-aged-87>, checked on 8/8/2024.
- McMahon, Liv (2024): Billie Eilish and Nicki Minaj want stop to 'predatory' music AI. In *BBC News*, 2/4/2024. Available online at <https://www.bbc.com/news/technology-68717863>, checked on 8/13/2024.
- mebis Magazin (2024): KI | Die Geschichte der Künstlichen Intelligenz. Available online at <https://mebis.bycs.de/beitrag/ki-geschichte-der-ki#sec7>, updated on 8/6/2024, checked on 8/8/2024.
- Mehta, Ivan (2024): ElevenLabs debuts AI-powered tool to generate sound effects. In *TechCrunch*, 5/31/2024, checked on 7/23/2024.
- MicroMusic Inc. (2024): MicroMusic. Available online at <https://micromusic.tech/about>, updated on 7/25/2024, checked on 1/8/2024.
- Mockenhaupt, Andreas; Schlagenhaut, Tobias (2024): Digitalisierung und Künstliche Intelligenz in der Produktion. Grundlagen und Anwendung. 2. Aufl. Wiesbaden: Springer Vieweg.
- Mubert (2024). Available online at <https://mubert.com/render/faq>, updated on 8/6/2024, checked on 8/6/2024.

Mucci, Tim (2024): Getting ready for artificial general intelligence with examples - IBM Blog. Available online at <https://www.ibm.com/blog/artificial-general-intelligence-examples/>, updated on 4/25/2024, checked on 8/6/2024.

MusicFX (2024). Available online at <https://aitestkitchen.withgoogle.com/tools/music-fx>, updated on 7/20/2024, checked on 7/20/2024.

MusicGen - Advanced AI Music Generation (2024). Available online at <https://musicgen.com/>, updated on 11/4/2024, checked on 7/31/2024.

Open Music Academy (2024). Available online at <https://openmusic.academy/docs/WAsrV8ETFnJavaS1Lbgs6m/techno>, updated on 11/8/2024, checked on 11/8/2024.

Phleps, Thomas (2015): Pop Sounds. Klangtexturen in der Pop- und Rockmusik. Basics - Stories - Tracks. With assistance of Ralf von Appen. 1st ed. Bielefeld: transcript (texte zur populären musik). Available online at <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6955626>.

Rightsify (2018): blog - Rightsify. In *Rightsify*, 1/11/2018. Available online at <https://rightsify.com/blog/>, checked on 8/6/2024.

Schumann, Nicolai (2024): Gamechanger Künstliche Intelligenz. Wie künstliche Intelligenz inspiriert und kreatives Potenzial entfesselt. 1. Auflage. München: Haufe; Imprint Haufe. Available online at <https://link.springer.com/book/10.34157/978-3-648-17563-7>.

SIX AM (2024a): A Beginner's Techno Music Guide: Brief History, Artists & Clubs. Available online at <https://6amgroup.com/articles/guides-all/a-beginners-techno-music-guide-brief-history-artists-clubs>, updated on 8/8/2024, checked on 8/8/2024.

SIX AM (2024b): Acid Techno Guide: History, Artists & Classics. Available online at <https://6amgroup.com/articles/guides-all/acid-techno-guide-history-artists-classics>, updated on 7/19/2024, checked on 7/19/2024.

Snowpixel (2024). Available online at <https://make.snowpixel.app/>, updated on 8/6/2024, checked on 8/6/2024.

Soundtrap (2024): What Are One Shots in Music & Best Ways To Use Them. Available online at <https://www.soundtrap.com/content/blog/what-are-one-shot-samples>, updated on 11/8/2024, checked on 11/8/2024.

Soundverse (2024): Introduction - Soundverse. Available online at <https://help.soundverse.ai/introduction>, updated on 7/22/2024, checked on 7/22/2024.

Stable Audio - FAQs (2024). Available online at <https://stableaudio.com/about>, updated on 7/23/2024, checked on 7/23/2024.

Statista (2024): Preferred audio content by genre in Germany 2024 | Statista. Available online at <https://www.statista.com/forecasts/1466148/preferred-audio-content-by-genre-in-germany>, updated on 8/4/2024, checked on 8/4/2024.

Sun, Michael (2023): Paul McCartney says there's nothing artificial in new Beatles song made using AI. In *The Guardian*, 6/23/2023. Available online at <https://www.theguardian.com/music/2023/jun/23/paul-mccartney-says-theres-nothing-artificial-in-new-beatles-song-made-using-ai>, checked on 8/13/2024.

Suno (2024). Available online at <https://suno.com/>, updated on 4/6/2024, checked on 4/6/2024.

Teneo.ai (2024): John McCarthy: homage to the father of Artificial Intelligence (AI). In *Teneo.ai*, 11/6/2024. Available online at <https://www.teneo.ai/blog/homage-to-john-mccarthy-the-father-of-artificial-intelligence-ai>, checked on 8/8/2024.

The Guardian (2024): Music labels sue AI song generators Suno and Udio for copyright infringement. In *The Guardian*, 6/25/2024. Available online at <https://www.theguardian.com/music/article/2024/jun/25/record-labels-sue-ai-song-generator-apps-copyright-infringement-lawsuit>, checked on 6/26/2024.

Udio (2024): Udio | AI Music Generator - Official Website. Available online at <https://www.udio.com/guide>, updated on 7/18/2024, checked on 7/18/2024.

Venutti, Isabella (2024): A brief history of the DAW, or Digital Audio Workstation. Available online at <https://mixdownmag.com.au/features/a-brief-history-of-the-daw-or-digital-audio-workstation/>, updated on 4/30/2024, checked on 8/4/2024.

Vetter, Jens (2018): Techno-History 1: Die Fundamente des Techno - AMAZONA.de. In *Amazona.de*, 1/18/2018. Available online at <https://www.amazona.de/techno-history-1-die-fundamente-des-techno/>, checked on 8/8/2024.

Voigts, Eckart; Auer, Robin Markus; Elflein, Dietmar; Kunas, Sebastian; Röhnert, Jan; Seelinger, Christoph (2024): *Artificial Intelligence - Intelligent Art?* Bielefeld, Germany: transcript Verlag (64).

Ward, Abby (2023): How to Use Microsoft Copilot's New Suno AI Music Creation Tool. In *Tech.co*, 12/21/2023. Available online at <https://tech.co/news/how-to-use-microsoft-copilot-suno-ai-music>, checked on 7/18/2024.

Wikipedia (Ed.) (2024): Techno. Available online at <https://en.wikipedia.org/w/index.php?title=Techno&oldid=1237439325>, updated on 7/29/2024, checked on 6/8/2024.

Zantal-wiener, Amanda (2017): From the Phonograph to Spotify: The History of Streaming Music. In *HubSpot*, 8/3/2017. Available online at <https://blog.hubspot.com/marketing/history-of-internet-radio>, checked on 8/4/2024.

## Appendix

### Digital Appendix

All sound examples, references, presets, and productions can be downloaded [here](#). The link expires 31 March 2025.

## Transcription Expert Interviews

### Interview 1

#### Dr. Tristan Behrens

**Interviewer:** Gut, dann würde ich dich jetzt erst mal fragen oder willkommen heißen: Hallo! Hallo Tristan. Ich würde dich gerne fragen, wie jetzt deine Berufsbezeichnung ist. Erst mal sowas allgemeines zu dir ein bisschen.

**Tristan:** Es hat lange gedauert, bis ich das gefunden habe, aber seit ein paar Jahren nenne ich das, was ich mache AI Hands-On-Advisory. Das ist relativ nah am Consulting, also IT-Consulting, wobei ich auch strategisch berate. Und ganz wichtig dabei ist, ist das Hands-On. Das heißt, wenn ich beim Kunden bin, dann ist es meistens so, dass ich dann auch für den Kunden und mit dem Kunden programmiere und an Lösungen arbeite.

**Interviewer:** Okay, okay. Und wie bist du dazu gekommen? Also wie bist du dahin gekommen?

**Tristan:** Das war ein etwas längerer Weg. Ja. Also mein Weg, wie ich zu dieser Tätigkeit gekommen bin, ich lasse mal viele Sachen aus. Ich fange an bei - ich habe in den 90er Jahren, also 1990er Jahren, gelernt, zu programmieren. Das habe ich mir selbst beigebracht. Ich habe ein paar Sachen dann auch danach an der Schule gelernt. Und mein Ziel war es immer, Spieleprogrammierer zu werden. Spiele waren für mich so wunderbar holistisch. Da kann man ganz viel machen. Storytelling, da hat man Grafik, da hat man Klang und man hat noch Musik dazu und Physik, die man programmieren muss.

Und dann habe man mir gesagt, um in Deutschland als Programmierer Geld zu verdienen, baue ich mal ein Diplom in Informatik. Ja, dann habe ich mich eingeschrieben. Nach dem Wehrdienst war ich an der TU Clausthal, habe da 4,5 Jahre studiert, habe 2006 abgeschlossen im Diplom. Und gerade, als ich überlegt habe, in die Industrie zu gehen, kam mein Doktorvater auf mich zu, der Professor Dix. Und hat gefragt, ob ich nicht sein, also einer seiner wissenschaftlichen Assistenten werden möchte. Und ich könnte dann auch in dieser Zeit promovieren im Thema künstliche Intelligenz. Dann habe ich das gemacht. Die ersten paar Jahre den Uni-Betrieb gemacht und dann langsam reingegangen in die Forschung und Thesis schreiben. 2012 habe ich dann meine Dissertation verteidigt über Multiagentensysteme.

Das war ganz, ganz schön. 2012 ist auch für die KI eine besondere Zeit, weil man da gesehen hat: ImageNet, dass diese tiefen neuronalen Netze, die auf den GPUs trainiert wurden, wirklich toll sind. Nun konnte man in der Zeit in Deutschland eigentlich kein Geld mehr KI verdienen. Deshalb habe ich da IT-Beratung gemacht. War erst bei der Volkswagen AG im Innovation Lab und bei den Entwicklern, habe da ganz, ganz viel Digitaltransformation gemacht. Ich bin dann gewechselt zur MHP, der Porsche-Tochter, wobei das ist jetzt keine Tochter mehr, das ist glaube ich ein hundertprozentiges Unternehmen. Und bin dann am Ende 2017, 2018 war ich dann im Porsche Digital Labor.

In Berlin habe ich dann Intensiv KI und Deep Learning gemacht und bin Anfang 2018 in die Selbstständigkeit. Und seitdem treibe ich mich wunderbar um und berate Firmen. Und sozusagen, ich war auch damals schon gesegnet mit dem ersten Projekt, was direkt in Audio war. Da habe ich dann Klänge analysiert für jemanden aus der Musikindustrie. Ich war auch mal in einem Start-up tätig, da haben wir Beats generiert mit Künstlicher Intelligenz. Und 2021 habe ich mich dann sehr öffentlich aufgestellt mit meiner Heavy Metal KI, die damals dann angefangen hat. Um es zum Abschluss zu bringen, im letzten Jahr bis, ich glaube, März dieses Jahres war ich AI Artist in residence im KI-Salon Heilbronn. Da habe ich dann Ausstellungsstücke gebaut, die mit generativer KI in der Musik zu tun haben.

**Interviewer:** Ja, super, perfekt. Damit hast du auch meine Frage beantwortet noch. Seit wann dein erstes Musik-KI-Projekt begonnen hat? Oder wie dein erster Kontakt zu Audio war?

**Tristan:** Wobei, da muss ich antworten, ich glaube die ersten Berührungspunkte mit Musik-KI hatte ich, also mit Musik-KI, also generativer Musik mit Computern, das war schon vorher. Musik-KI, das muss 2017 gewesen sein. Mit den ersten handfesten Ergebnissen, weil ich dann ein bisschen Zeit hatte, in 2019 und seit 2021 ist es einfach nur weniger.

**Interviewer:** Ah, okay, ja, alles klar. Gut, dann würde ich mal zu meinen Fragen übergehen, die ich so ein bisschen geordnet habe, schauen wir mal, wie das funktioniert. Da würde ich jetzt gerade so ein bisschen um die Entwicklung der letzten Jahre gehen und vielleicht den aktuellen Stand. Was denkst du, wie hat sich die KI im Bereich Musik in den letzten Jahren so entwickelt oder vielleicht sogar in den letzten zwei Jahren?

**Tristan:** Da gab es, also die KI hat sich in den letzten Jahren, und es tut mir ein bisschen, ich tue mich ein bisschen schwer zu sagen, wann das wirklich akut gewesen ist -also so KI in der Musik oder im Computer, das ist ja schon ein relativ altes Thema, aber das ist jetzt heute so, dass es relativ gut funktioniert. Es gibt da verschiedenste Tools, die jetzt gut funktionieren, und das kann man so ein bisschen eingrenzen, weil es sehr, sehr viel gibt. Ich meine, in der Musikproduktion gibt es sehr viele Anwendungsfälle für künstliche Intelligenz. Das Mastering ist was, das Generieren von Noten, das Generieren von Klängen, das sind verschiedene Themen, die man da abdecken kann. Und wenn es jetzt darum geht, was gerade eine sehr, sehr große Prominenz hat, das ist das Text to Music.

Da gibt es verschiedene Tools, wo man dann über einen Prompt - das kennt man ja mittlerweile, das Prompt Engineering - dann Musik generiert, die jetzt schon eine relativ hohe Qualität hat. Relativ hoch heißt, was dabei rauskommt, das überzeugt schon so ein bisschen. Nicht alles, was rauskommt, überzeugt, sondern das sind so gewisse Prozentsätze, die dann noch nicht so gut funktionieren. Und wenn man hinhört, dann hört man immer noch Artefakte raus, die halt verraten, dass es KI generiert ist, bin aber selbst der Vermutung erlegen, es sind relativ wenige Leute, die das dann wirklich hören.

**Interviewer:** Kommt natürlich auch auf das Medium drauf an, wo man es hört.

**Tristan:** Genau, genau.

**Interviewer:** [...] ist das dann auch der Bereich, den du jetzt gerade am vielversprechendsten findest?

**Tristan:** Der ist am vielversprechendsten für Leute, die die Musikproduktion nicht gelernt haben und die auch keine Zeit und Energie haben, das zu lernen. Also das, was man unter Musikproduktion versteht, wenn man dann sich verschiedene Samples, Events in der Digital Audio Workstation und herschiebt, sich um die Automatisierung kümmern, das Mixing macht, das Mastering macht. Diese ganze Prozesskette, die Musikproduktion da ist, das können Leute jetzt halt umschiffen.

Bedeutet, diese Modelle sind momentan sehr, sehr charmant für Leute, die da relativ leicht sogar, die eigentlich gar nicht ausgebildet sind, was sehr schön ist für die Demokratisierung. Das heißt, dass viel mehr Leute in der Lage sind, sich musikalisch auszudrücken. Die müssen ja nicht mehr Noten lernen, die müssen ja nicht mehr lernen, mit dem Tool umzugehen, weil sie mit der natürlichen Sprache arbeiten. Und das ist für mich eine sehr, sehr große Hoffnung, weil ich das halt mag, wenn viel mehr Leute, viel mehr Musik machen, mit angenehmer Qualität.

**Interviewer:** Ja, das ist die Sache, denke ich auch. Dann du sagst also, dass die Qualität noch nicht hoch genug ist. Das heißt, das fehlt dir, sagen wir mal, als Meilenstein noch bis zu dem Punkt, dass man sagen kann, dass die Musik jetzt auf der gleichen qualitativen Ebene ist, wie menschlich erzeugte Musik. Oder gibt es da noch andere Dinge, die dir fehlen?

**Tristan:** Also bei diesen Ansätzen, Text to Music, müssen wir auf der einen Seite davon ausgehen, analog zur Bildgenerierung, Text to X gibt es ja schon eine Weile. Text to Image, mit Anbietern wie Dall-E oder Midjourney, da haben wir gesehen, dass es sehr schnell geht, dass die Qualität der generierten Artefakte halt überragend ist und dass man das auch gut steuern kann. Die Musik hängt immer ein paar Jahre dem Bild hinterher.

Das bedeutet, ich gehe davon aus, in diesem Jahr, wenn nicht sogar nächstes Jahr, vielleicht in dem darauffolgenden Jahr, sind diese Ansätze so gut, dass man ganze Lieder mit sehr guter Qualität generieren kann. Das ist nur eine Frage der Zeit. Schrägstrich des Trainings dieser Modelle.

**Interviewer:** Ja, okay. Stimme mich zu so weit. Gut, dann würde ich ganz gerne mal ein bisschen zu den künstlerischen Aspekten mal übergehen, weil die tragen ja auch zur Qualität mit bei. Könnte man behaupten? Ja, könnte man sich darüber streiten. Wie siehst du denn, inwieweit beeinflusst denn die KI die Kreativität und die Intuition, die man jetzt als kunstschaftende Person auch hat in der Musik?

**Tristan:** Ich meinte ja gerade, dass so eine KI als Enabler funktioniert, dass man Leute damit arbeiten kann. Eine andere Säule von generativer KI ist natürlich Leute, die schon in dem Thema drin sind, die Musik produzieren können, die da schon gewisse Fortschritte gemacht haben. Da geht es dann in so Richtung Inspiration. Es geht nichts über, man setzt sich kurz hin, lässt sich von der KI mit vielleicht einem Prompt was generieren und dann bekommt man Ideen, die man dann später nochmal selbst umsetzt. Man kann



auf dem Weg oder man wird wahrscheinlich auf dem Weg das KI generierte Material verwerfen und darauf auf der Idee aufbauen, dann was erschaffen. Auf der anderen Seite, wenn wir uns ein bisschen entfernen von Text to Music, wo dann wirklich nur das Zeitsignal dabei rauskommt, wer jetzt eine KI hat, die Noten generiert, also das Notenblatt ausspuckt, dann ist das nochmal eine ganz andere Qualität, weil man da auch innerhalb kürzester Zeit sehr viel Notenmaterial bekommt mit einer angenehmen hohen Qualität. Und dann, und das in meiner Erfahrung, ist man eher auf so einer Ebene, wo man an der Grenze zu Überinspiration ist. Es sind immer Ideen da, es sind auch sehr viele gute Ideen dabei und dann wird es halt schwierig, die zu kuratieren. Die KIs sind dann teilweise recht spielerisch und geben dann - auch ein bisschen gesteuert, so viele Ideen, dass man eigentlich gar nicht dahinterher kommt, da jeweils ein Lied daraus zu machen.

**Interviewer:** Das heißt also, dass man als Producer die Rolle des Kurators ein bisschen mehr übernimmt?

**Tristan:** Ja, genau. Also, man liefert der KI ja eine Vorlage, wenn man jetzt ein Drum-track hat und eine Basslinie, dass man dazu dann noch ein Bassinstrumental dazu kommt. Also, man kann halt steuern nicht über die Sprache, sondern über das Material, was man da z.B. Notenformen reingibt. Das ist so die eine Sache.

Und das dürfen wir nicht vergessen, wenn man jetzt an diese Noten generierenden KIs denkt, die sind trainiert auf Notenblätter, zum Beispiel auf MIDI-Dateien, das ist ein digitales Notenblatt. Und da gibt es Datasets, wie es so schön heißt, in unglaublichen Größen. Also, es gibt ein Dataset, da sind 400.000 Lieder dran. Da kann man dann überschlagen, mit wie vielen Musikern man da eine Jam Session macht. Wenn man überlegt, wie viele Leute- wie viele Musiker - dazu beigetragen haben zu so einem Dataset indirekt, dann wird einem das schon so ein bisschen schwindelig. Und das ist halt genau das. Wenn man jetzt überlegt, früher saß man in einem Proberaum mit seiner Band, das waren so vier Leute, man hat Bier miteinander getrunken und dann hat man Ideen gesponnen. Und jetzt ist es so, auf Knopfdruck bekommt man von weitaus mehr Künstlern, besonders auch Künstlern, die sich durchgesetzt haben, Ideen, mit denen man dann weiterarbeitet.

**Interviewer:** Ja, stimmt. Stimmt mich zu. Das heißt, es kommt viel auf die Datasets an und auf welche Musik, die KI trainiert wird. Genau. Wie stehst du dann, daher gehen zu den, das ist ja oft in Genres unterteilt oder in bestimmte Stile, die dann auch vertextlicht werden müssen. Wie stehst du dazu, dass KI potenziell diese Genres ausbauen könnte? Glaubst du, die können das erweitern oder vermischen, neue Genres erstellen oder fehlt da der gewisse menschliche Input?

**Tristan:** Den KI fehlt in der Tat sehr, sehr, sehr viel. Also auf der einen Seite die Datenverarbeitung kriegen die ja wunderbar hin. Was Du gerade gesagt hast, das ist dieses Interpolieren von Musikstücken, von Musikgenres, da ist die KI sehr, sehr gut drin. Es gibt ein paar Wissenschaftler, die streiten sich, ob da nicht vielleicht auch so ein gewisser Grad an Extrapolation drin ist. Das heißt, man findet ja Sachen, die zwischen den

Sachen sind, da sind eventuell auch neue Sachen dabei. Und auch die KIs machen Fehler, wo dann auch wiederum neue Sachen draus entstehen können. Also da kann man sagen, ja, das sieht an der Stelle schon mal ziemlich gut aus, was diese Kreativität betrifft. Aber auf der anderen Seite - stand heute - es sind die KIs, es sind zwei Punkte. Der erste Punkt ist, die KIs sind in ihrer eigenen kleinen Welt gefangen, die sind ja im Computer drin. Wir haben einen gewissen Nachteil, wir haben denen was voraus, weil wir sind ja in dieser „hier-Erfahrungswelt“ und bekommen weitaus mehr Input als nur das reine Notenmaterial. Wenn ich gerade verliebt bin, schreibe ich vielleicht andere Lieder, als wenn ich gerade eine Magenverstimmung habe oder ich habe gerade ein Lied im Radio gehört und denke, ja, das ist eine gute Idee, die muss man irgendwie erweitern. Oder ich bin gerade wandern und erfreue mich über die Landschaft irgendwie in Slowenien, so was. Das sind diese Inspirationen, die die KI in der Form nicht hat. Und was die KI auch nicht hat, das ist dieses Zielgerichtete, die ist ja rein reaktiv. Ich habe bislang noch keine kreativen KI oder kreativen Agenten gesehen, der das alleine jetzt selbst gerichtet, die Musik generiert. Es ist immer ein Mensch dabei, der dieses Werkzeug ansteuert und dann die Ergebnisse daraus kommen kuratiert und eventuell noch mal ediert. Aber dass die KI selbst gerichtet ist, das habe ich noch nicht gesehen.

**Interviewer:** Okay, ja, spannend. Dann haben wir da auch ganz gut schon den Weg gefunden zu meinem nächsten Fragenblock; der Interaktion zwischen dem Mensch und der Maschine oder in deinem Fall der „Hexagon-Maschine“ vielleicht. Genau. Wie sieht denn da die Zusammenarbeit zwischen dir und der KI? Du hast die ja trainiert auch, oder?

**Tristan:** Genau, genau. Die Zusammenarbeit von mir und der KI, das sind verschiedenste Phasen gewesen. Die erste Phase war überhaupt erst einmal in das Thema reinkommen. Das ist das, was du gesagt hast, das Trainieren der KI. Das ist auf der einen Seite, die Implementierung des neuronalen Netzes und des Trainings von demselben, was für das Projekt Hexagon Machine relativ schnell ging. Das waren so ein, zwei Tage maximal, an einem Osterwochenende und dann war es in Ordnung. Danach mussten Daten besorgt werden, die Daten mussten kuratiert werden, die Daten müssen vorverarbeitet werden. Das ging dann auch relativ schnell innerhalb von wenigen Tagen. Und dann muss ein neuronales Netz trainiert werden, nochmal zwei Tage.

Also nach guter Woche ist das fertig. Schwierigkeit dabei ist, ja, diese KI spuckt munter Noten raus mit mehreren Instrumenten gleichzeitig, aber damit kann man ganz, ganz schlecht interagieren. Das ist wie beim ChatGPT, dass man irgendwie eine Note reinschreiben muss und dann werden die anderen Noten generiert. Aber dass man jetzt nochmal was wiederholt, dass man von mehreren Tracks die Gitarren nochmal neu macht, dass man vielleicht noch einen Tag dranhängt und andere kleine Operationen macht, dafür musste ich dann erst mal ein Tool programmieren. Es hat dann auch wieder ein paar Tage gedauert, dass ich dann per Klick mit dieser eigentlich musiktextbasierten KI interagieren konnte. Und dann hat sich ein Prozess etabliert, dass ich mit so einem Tool fünf bis fünfzehn Minuten gearbeitet habe, um reines Notenmaterial zu generieren, also 30, 40 Takte, die so eine gewisse Geschichte andeuten und dann wurde dieses

Material übergerettet in die Digital Audio Workstation, wurde arrangiert im Sinne von, dass dann die Songstruktur festgelegt wurde, also was ist ein Chorus, was ist ein Verse, Intro, Outro und so weiter. Und dass ich mich dann auch noch um das Timbre kümmern muss.

Das heißt, welche Melodie wird es mit welchem Instrument in welcher Tonhöhe ausgedrückt, welche anderen Harmonien oder anderen Timbres kommen noch dazu, um das Gesamtbild zu vervollständigen. Also da, die Klanggestaltung, die lag da noch komplett bei mir, das hat die KI ja nicht gemacht, weil die KI ja nur die Noten symbolisch generiert hat. Und dann, die Automatisierung habe ich dann noch selbst gemacht. Und ich meine, dass ich für das erste Album dann auch ein Tool benutzt habe, E-Mastered, müsste es gewesen sein, das auch direkt online Mastering in angenehmer Qualität bietet, und dann konnte ich das Mastering dann auch abhaken. Und die Bottom Line davon ist, es geht ja auch um die Steigerung der Kreativität, das hatten wir gerade schon, ich hatte sehr, sehr viele neue Ideen, die ich allein nicht hätte generieren können.

Und die Produktivität, ich habe das damals so ein bisschen überschlagen, ich habe Pro-Lied ungefähr vier Stunden gebraucht, mit allem Drum und Dran, wobei ich vorher, wenn ich vorher instrumentale Musik gemacht habe, dann waren das acht Stunden pro Stück. Also nicht unbedingt direkt in einer Session, sondern so ein bisschen verteilt, aber ich habe gemerkt, ich habe dann eine doppelte Reichweite in der gleichen Zeit.

**Interviewer:** Das heißt, die KI unterstützt dich überwiegend oder hat dich zumindest bis dahin überwiegend in der Komposition unterstützt, also wenn du das Arrangieren ja auch noch selber übernommen hast?

**Tristan:** Genau, Komposition, meistens so sechs Spuren gleichzeitig, also ich habe immer gearbeitet, mit natürlich Drums und ein Bass hat man drin. Dann hatte ich eine Gitarren- oder Pseudogitarrenspur, Hexagon Machine ist ein bisschen an Heavy Metal angelehnt, aber es sind weniger Powerchords drin, als man erwarten würde. Und dann meistens noch so ein, zwei Spuren als Dekoration drum herum, die man dann an und ausgemacht hat.

**Interviewer:** Und das sieht jetzt immer noch so aus, weil sich ja die KI-Tools für die Produktion auch deutlich weiterentwickelt haben. Verwendest du die jetzt auch?

**Tristan:** Ja, es hat sich ein bisschen angepasst. Also ich bin jetzt auch gerade zugeben-ermaßen in der Neufindungsphase, wo ich das Projekt nochmal überarbeite. Ich habe noch bessere Musik-GPTs, das ist ganz wichtig, das ist dann halt GPTs, diese Musikgeneratoren, mit einem viel größeren Dataset trainiert. Da kommt jetzt noch hinzu, dass ich gerade diese Integration mit Ableton Live mache, über Max for Live, was das auch ein bisschen beschleunigen sollte. Weiterhin, ich habe mir auch angewöhnt, von klanggenerierenden KIs so Flächen einzubauen. Wenn man jetzt ein Sample generiert, und das jagt man dann durch einen Granular Synthesizer, kann man damit dann schöne Klangumgebungen in die Lieder miteinfügen. Und was auch noch ein ganz großer Schritt gewesen ist, ist, ich habe einen Gitarren-Synthesizer, den ich zusammen mit einer KI gebaut habe. Und ich nutze auch von [...] einen Vocal Synthesizer.

Das ist eine Gesangsstimme, wird gemacht über ein neuronales Netz. Und man gibt im Wesentlichen, wie bei einem Synthesizer, die Melodie vor. Und an jede Note schreibt man dann noch die Silbe dran. Und dann singt die KI das in einer Qualität, die ziemlich überragend ist.

Also ich kenne Produzenten, die sagen, man kann es gar nicht mehr raushören, dass es genau diese spezifische Gesangs-KI ist. Und wenn man dann noch ein gewisses Token davor schreibt, dann atmet die sogar ein. Und das macht es dann noch natürlich. Also das ist eine wunderbare Qualität.

**Interviewer:** Das heißt, für dein letztes Album war das dann auch schon die Vocal-KI oder warst das noch Du?

**Tristan:** Das war es noch nicht, das war ich durch den Verzerrer gejagt.

**Interviewer:** Ja, okay. Dann würde ich gerne sehen. Also, inwieweit siehst du die KI als Partner oder Partnerin in deiner Musikschaftung dann? Weil das ist ja immer noch getrennt jetzt für euch. Oder ist nicht die gesamte Arbeit, die von der KI übernommen wird, nicht auch die gesamte Arbeit von dir?

**Tristan:** Ja, das ist eine sehr, sehr spannende Frage. Und ich habe das schon öfter mal darüber nachgedacht. Also es ist eher so ein Gefühl. In der Musikproduktion sind das ja mittlerweile mehrere. Mit den KIs fühle ich mich eher so in einer Gruppe von Entitäten, vielleicht sogar so kreative Orakel, die ich da teilweise habe, mit unterschiedlichen Tätigkeiten verteilt auf die unterschiedlichen Agenten, die wir dann haben. Ich fühle mich immer noch so ein bisschen in der Mitte sitzen.

Ich weiß aber nicht, ob ich sagen kann, dass das eine Hierarchie ist, wo ich oben bin. Ich fühle mich da eher so mittig, sozusagen als Mediator zwischen den verschiedenen Ideengebern. Wir haben das ja schon festgestellt, dass ich auch sehr viele Ideen dann ablehne, dass ich Ideen auch mal in eine andere Richtung schubse, dass ich sage, hier vielleicht weniger Noten, hier mehr Noten, macht doch mal das Schlagzeug hier neu. Also quasi so ein, auch so ein Ideengeber-Speicherung in die KI hinein.

Bin aber immer sehr überrascht, wie schnell das eigentlich geht, dass ich Musikmaterial bekomme, was zu der aktuellen Richtung der aktuellen Stimmung passt. Das geht immer sehr, sehr schnell. Und das Schöne für mich bei diesem Prozess ist, ich mache ja in meinen Arbeiten immer die Unterscheidungen zwischen „Generated by AI“, das ist, wenn die KI auch Knopfdruck was generiert und „Generated with AI“. Was bedeutet, dass die KI in mehreren Schritten ab und zu mal herangezogen wird, um zu helfen, um Material zu liefern. Und das ist das, worauf ich mich fokussiere.

Obwohl ich Knöpfchen drücke oder Text eingeben, Enter und es kommt dann was heraus, finde ich charmant, aber das ist nicht meine Arbeit. Ich bin liebend gerne ein bisschen kleinteilig unterwegs. Und das ist auch das Problem von vielen Text-to-Music-Anbietern, dass das Editieren des generierten Materials einfach nicht geht. Dass man hier nochmal eine Note anpasst und hier die Akkordfolge leicht ändert. Oder hier in dem Text eine Silbe ändert. Das geht mit diesem KI-Stand heute noch nicht.

**Interviewer:** Ja, ich habe das, das wertschätzt du dann auch und kennzeichnest du dann ja auch so, oder? Also ich hatte gesehen auf Spotify zum Beispiel, stand dann **Tristan Machine** als Composer drin oder Heavy Metal GPT. Deswegen wollte ich das auch fragen, finde ich cool.

**Tristan:** Genau, genau. Das war auch meine, das müsste die zweite Metal-KI gewesen sein, die ich dann hatte.

**Interviewer:** Ja, und ganz aktuell, ich weiß nicht, ob du das mitbekommen hast mit, also kennst du udio bestimmt, oder? Udio AI. Ja, die hat jetzt - am Wochenende kam das raus, dass jetzt auch In-Painting möglich ist. Hast du das probiert schon mal?

**Tristan:** Also ich kenne In-Painting. Es ist ja ein Diffusion-Model. In-Painting habe ich bei den visuellen Künstlern immer benutzt. Ich weiß also direkt, was hier bei Audio Phase ist. Die Frage ist, ist das In-Painting dann hier auf einem Teil von dem gesamten Song? Oder kann das auch auf einzelne Spuren einwirken?

**Interviewer:** Auf den gesamten, ja.

**Tristan:** Und das ist nämlich die Schwierigkeit. So, wo ich das verstanden habe, und das haben die Leute nicht so veröffentlicht, bei ein paar weiß man das, generieren die direkt einen Track. Keine Stems, sondern einen Track, da wird zwischendurch abgemischt, das ist wirklich end-to-end. Und es wäre ein klarer Vorteil, wenn man einfach Multitrack hätte.

Und das gegenseitige Konditionieren, das würde dann auch wunderbar funktionieren. Man sagt hier den Basslauf an der Stelle, den möchten wir nochmal neu generiert haben. Es muss technisch möglich sein. Schwierigkeit ist aber hier das Dataset. Es ist viel, viel einfacher, ganz YouTube zu [lacht] - wie man sagt - zu scrapen, um Stereo-Tonmaterial zu bekommen, als dass man an wirkliche Produktionssysteme, also Multitracks rankommt, auf die man dann trainieren kann, die sind eher selten.

**Interviewer:** Ansonsten müsste man ja die Gesamt-Songs nochmal in die Stems separieren und diese analysieren und diese sind ja auch noch nicht ganz so genau...

**Tristan:** Genau, genau. Zu viele Artefakte auf dem Weg.

**Interviewer:** Ich hätte dann natürlich noch die obligatorischen ethischen Überlegungen dazu. Was denkst du, was für ethische Fragen werfen denn jetzt KI generierte Musikstücke auf oder können die aufwerfen? Was muss man da beachten?

**Tristan:** In der Kreativität, wie eigentlich überall, gibt es so ein gewisses Damoklesschwert. Und das Damoklesschwert heißt Pareto-Verteilung. Heißt im Wesentlichen, dass in der Musikindustrie, wie auch in anderen Kreativen und darüber hinaus Industrien, dass sehr, sehr wenige Leute eigentlich das ganze Geld verdienen. Bedeutet im Umkehrschluss, dass es heutzutage, und es war eigentlich schon immer so, dass es relativ unwahrscheinlich ist, dass man seinen Lebensunterhalt mit Kreativität bestreiten kann. Natürlich, es gibt immer welche.

Und die Möglichkeit da auch hinzukommen ist, immer kreativ zu sein, immer nach außen damit zu gehen, immer dranzubleiben, damit erhöht man die Wahrscheinlichkeit, dass man in der Pareto-Verteilung eher so bei den wenigen landet, die damit Geld verdienen. Und es ist immer ein Konfliktpotenzial gewesen, bis heute. Und jetzt stellt sich so die Frage, bekommt man durch diese KIs Vorteile?

Auf der einen Seite steht ja die Möglichkeit, dass weitaus mehr Leute an diesem kreativen Spiel teilnehmen. Das hatten wir aber vorher schon. Wenn plötzlich Leute mit FL-Studio, damals Fruity Loops, arbeiten konnten, das ist schon mal ein großer Unterschied gewesen zu „ich habe nur eine Vierbandmaschine und muss Instrumente lernen und spielen“.

Jetzt kann man das klicken. Oder wenn man sich jetzt so Chord Progressions aus dem Internet runterlädt und die dann einsetzt, dann hat man auch schon etwas größere Reichweite. Instrumente mit genialen Presets, wo man eigentlich gar nichts mehr einstellen muss, so was. Und wir haben [...], wir stehen jetzt gerade und gucken zurück in die Vergangenheiten und da ist so eine riesige Kette an technologischen Entwicklungen gewesen, die gesagt hat, ja, okay, plötzlich können mehr Leute einfacher Musik generieren. Beispiel in der Kirche, wenn man da an die Orgel denkt. Das ist ja jetzt der göttliche Atem.

Da hat man eine ganz andere Ausdruckskraft, als man jetzt mit dem Gesang hatte, der vorher eingesetzt wurde. Dann plötzlich kann man noch mehr Stimmen gleichzeitig haben, mit einem weitaus größeren Tonumfang. Und es ist eine Person, die das macht. Plötzlich gibt es eine Person, die dann mehr Stimmen singen kann. Das ist auch ein großer Schritt da gewesen. Und so kann man dann die ganze Reihe zurückgehen, bis hin zu wahrscheinlich zu irgendwelchen Flöten aus Knochen gemacht, in irgendwelchen Höhlen, irgendwelchen Höhlenmenschen.

**Interviewer:** So, denkst du, also es ist der normale Verlauf der Technologie?

**Tristan:** Es ist der normale Verlauf der Technologie. Es ist natürlich ein Werkzeug. Die Werkzeuge sind immer mächtiger geworden, exponentielles Wachstum. Das ist gerade ein Werkzeug, das ist sehr, sehr mächtig. Ich sehe, dass der Vorteil gerade überwiegt, dass Demokratisierung mehr Leute diese KIs einsetzen können. Ich weiß nicht, wie sehr jetzt die Musikindustrie dadurch eine Disruption erfahren wird. Wie weit sich die ganzen Einkommenströme jetzt umsortieren werden.

Stand heute können wir sagen, dass es vielleicht gar nicht mehr so einfach ist, seinen Lebensunterhalt mit der Erstellung von Jingles, Fahrstuhlmusik, Hintergrundmusik für Werbung usw. zu verdienen. Weil das können die Kreativagenturen jetzt per Text einfach machen. Für ein Werbespot Musik zu bekommen, das geht heute relativ einfach.

Und auch viel schneller, da braucht man keine Vertragsverhandlungen, da braucht man nur ein Abonnement bei einem entsprechenden Dienstleister und dann geht das schon. Moralisch muss man auch überlegen, wie es aussieht mit den Daten. Das ist bei uns in Deutschland, da bin ich immer ein bisschen schwer, über das Copyright zu reden, weil

das Copyright ja nicht, was nicht Deutsches ist, sondern das Deutsche ist ja das Urheberrecht.

Dass solche Sachen dann auch wirklich schon geklärt hat. In Deutschland haben wir ein Alleinstellungsmerkmal, das ist im Urheberrechtsgesetz, das muss § 44 b sein, Data Mining. Data Mining sagt, dass man Daten, die frei verfügbar sind, zum Extrahieren von Mustern, das ist ja das, was neuronale Netze machen, heranziehen darf. Man muss aber hinterher die Daten von der Festplatte löschen. Und man darf Daten nicht berücksichtigen, wo es ein Opt-Out gibt, wo jemand gesagt hat, diese Daten zum Training nicht benutzen.

Eigentlich wunderbar, was bedeutet, wenn keiner Nein sagt, alles an Daten benutzen, worauf man trainieren könnte und danach die Daten wieder von der Festplatte löschen, dann ist das schon mal geklärt. Das ist dann wahrscheinlich eher für kommerzielle Applikationen, für Forschung ist das egal, die Forschung ist ja frei in Deutschland, da darf man ja alles benutzen. Also rein rechtlich kann man sagen, in Deutschland ist da schon ein Haken hinter, das wird nochmal angepasst und dann stellt sich aber noch eine andere Frage. Und das ist die Frage, ob man ein Urheberrecht anmelden kann, auf etwas, was von der KI generiert ist. Da ist es dann ein bisschen schwieriger, weil wir für ein Urheberrecht eine Schöpfungshöhe brauchen, die wir mit einem Text, und wir haben daraus ein bisschen Musik generiert, einfach nicht erreichen, ist nicht da.

Wenn man, davon muss ich ausgehen, sich von der KI ein Stück Musik generiert, und das dann samplet und das neu zusammenschneidet, einen Remix davon macht, muss ich davon ausgehen: Schöpfungshöhe erreicht und dann kann man an der Stelle wieder ein Urheberrecht anmelden. Wenn man mit den KIs was baut, kann jemand anderes ein Urheberrecht gelten machen, das ist eigentlich auch schon geklärt. Das machen wir schon lange, Urheberrechtsfeststellungen, da gibt es teilweise noch Verfahren, die seit 20 Jahren laufen. Ich glaube, das müssen jetzt schon 20 Jahre sein wegen 1, irgendwas Sekunden Samples, die er damals genommen hat. Das müssen wir dann mit einem Richter entscheiden. Also ja, auf der einen Seite die Schöpfungshöhe ist eine Sache, die festgestellt werden muss, um ein Urheberrecht gelten zu machen. Es muss hoch genug sein, auch das ist sozusagen eher so eine Spektrumsfrage, das ist nicht schwarz-weiß, das ist irgendwo dazwischen. Und dann auch, wenn ein Urheberrecht von jemandem verletzt wurde, das geht dann auch durch die entsprechenden Instanzen.

**Interviewer:** Okay, ja. Spannend, du hast auch das schon angeschnitten, [...] mit der Zukunft oder mit der Hintergrundmusik vor allem. Wie würdest du sagen, wo siehst du gerade am 1. die Zukunft der KI-generierte Musik, also wo wird die sich wiederfinden, sagen wir mal im Bereich Konsum? Oder wo wird die am ehesten Verwendung finden?

**Tristan:** Ich habe eine Vermutung, die kann ich aber nicht belegen, weil ich da nicht dabei bin. Aber bei TikTok ist das ein Thema. Da würde ich es vermuten, bei YouTube würde ich sagen, sehr wahrscheinlich, Hintergrundmusik für die eigenen Videos. Und dann natürlich, wenn jemand ein Intro haben will für seinen eigenen Podcast, da auf jeden Fall. Und dann auch überall im öffentlichen Raum, wo man sich bewegt. Und wenn

man da eine KI hat, die Musik generiert, dann stellt sich dann natürlich auch noch die Frage, wie es mit der GEMA aussieht. Und GEMA da was anmelden könnte, was dann nochmal komplett anderes Fass ist, das man dann aufmacht. Also dieses etwas Seichtere.

Und da ist nichts irgendwie berücksichtigt, was mit Musik zu tun hat, die wirklich dann auf irgendeine Art und Weise nochmal performt wird. Oder die im Zweifel sogar einen Platz in den Charts landet. Also es ist wiederum was anderes.

Also diese eher so, diese Content-Musik, die man in dem Content dann verarbeitet. Das ist die, wo jetzt die Wahrscheinlichkeit sehr hoch ist, dass man da mit der KI was einsetzen kann. Ich vermute, dass die Zahlen dann einfach stetig hochgehen. Das Nächste, was dann kommt, das ist nicht, das ist dann auch wieder so eine Hintergrundmusik. Leider ist es so, dass Leute, die heute Musik konsumieren eigentlich keine aktive [...] die konsumieren es nicht mehr aktiv. Man hört kein Album mehr von Anfang bis zum Ende. Das hat Sven Väth vor ein paar Jahren auch schon gesagt, dass die Leute nicht mehr in der Lage sind, ein Album komplett zu hören, sondern die brauchen Musik eigentlich nur als Hintergrund, um irgendwie die Sachen auszublenden.

Und wenn man da in einem gewissen eigenen Stil über die KI-Content generieren kann, ist das vielleicht auch für viele, viele Leute charmant. Was ich noch nicht ganz so sehe, aber es kommt wahrscheinlich, ist wirklich, wenn es um Musik geht, die in den Charts landet. Und da halte ich mich so ein bisschen fest an Garry Kasparov, der war ja auch sehr kreativ im Schachspiel. Der wurde Ende der 90er geschlagen von Deep Blue.

Und der hat danach auch eine Zeit lang zusammen mit KIs gegen Menschen mit KIs gespielt. Centaur Chess hieß das. Es war Augmentation, das heißt die eigene menschliche Kreativität, die eigene menschliche Intelligenz durch den Computer erweitert und hat damit natürlich auch eine gewisse Reichweite abgedeckt. Aber ich glaube, sowas macht man eigentlich gar nicht mehr.

Das ist das, was ich gerade sehr spannend finde. Leute, die in der Musikproduktion sind, wenn die dann mit, wenn die halt sehen, was die KIs produzieren, versuchen die vielleicht auch im Positiven auszuweichen und schaffen darüber, über diesen Druck wiederum neue Sachen. Ich bin sehr, sehr neugierig auf die neue Populäre und die neue Untergrundmusik vor allen Dingen, die durch diesen, es ist ein evolutionärer Druck geschaffen wird, den man der KI ausweichen muss, weil man weiß, was die KI generiert.

**Interviewer:** Man muss sich ja abheben davon, dann in Zukunft.

**Tristan:** Genau, genau, genau. Und da wird sehr viel passieren.

**Interviewer:** Dann hätte ich dazu aber auch noch die Frage, weil du ja, also weil wir gesagt haben, es geht [...] um Content-Musik, bedeutet das im Umkehrschluss für dich, denn dass die meiste komplett KI generierte Musik noch nicht den künstlerischen Wert hat, wie menschlich produzierte Musik, vielleicht auch im Zusammenhang mit der Person, die dahinter steht.



**Tristan:** Schwierig an Musik einen künstlerischen Wert dran zu packen. Ja. Ganz, ganz, ganz, ganz, ganz, ganz, ganz, ganz, ganz schwierig. Also es gibt Musik, die sehr, sehr einfach gebaut ist, die trotzdem sehr viel Erfolg hatte.

[beide lachen]

**Interviewer:** Ja, definitiv!

**Tristan:** Gibt es einen objektiven Maßstab, den man da ansetzen kann? Gibt es den für KI? Muss man ganz, ganz schwer überlegen. In der KI nennt man das ja dann den Benchmark. Und es gibt ja relativ wenige Benchmarks für Musik, und die Benchmarks, die sehr, sehr deutlich sind, sind, wie viel Geld man zum Beispiel mit einem Musikstück oder einem Album verdient hat, wie bekannt das geworden ist, ob man dafür Preise bekommen hat, wie lange man in den Charts gewesen ist. Das sind diese, diese KPIs, die Key Performance Indicators, die man da hat. Das sagt aber relativ wenig über die musikalische Güte aus.

Wenn man jetzt vergleicht, dass was in 80er Jahren in Charts gewesen ist, verglichen mit was im Barock und danach dann komponiert wurde, was dann komplett andere Qualität hatte, merkt man, dass es da dann auch schon so ein bisschen klafft.

**Interviewer:** Ja, ja. Stimme ich zu. Du [...] bist da positiv gestimmt, auch was die Entwicklung der gesamten Musikbranche angeht, in Bezug auf KI, dass man sich da, wie bisher immer, auch schon anpassen musste?

**Tristan:** Ja. Ich weiß auch gar nicht, ob das als Anpassung zu sehen ist. Es gibt ja heute immer noch Leute, die benutzen keinen Synthesizer, was ja okay ist. Es gibt Leute, die benutzen Synthesizer, aber keine Digital Audio Workstations. Es gibt Leute, die machen dann immer noch das Mastering selbst. Andere Leute benutzen dann Izotope und machen dann das darüber oder E-Mastered und so weiter und so fort. Ich glaube, das Spektrum bleibt uns immer noch erhalten, nur dass jetzt halt eine Tür aufgeht mit einem nächsten Tool, was von unterschiedlichen Leuten unterschiedlich angenommen und umgesetzt wird und was im Wesentlichen dazu beiträgt, dass wir dann weitaus mehr musikalisches Material haben und, wie gesagt, die Hoffnung, dass daraus dann auch neue Sachen erschaffen werden.

**Interviewer:** Schön, dann glaube ich habe ich so ungefähr alles ziemlich durch. Ich hatte vielleicht noch die letzte Frage. Was denkst du, worauf sollte man ein Auge werfen, vielleicht als selbst-Musikschaffende in Bezug mit KI oder auch an der anderen Stelle der Konsumierenden? Was wird jetzt die nächsten Jahre noch kommen, was wird Großes passieren?

**Tristan:** Was sehr interessant ist, das ist die sogenannte Agentisierung, dass die KIs auf der einen Seite autonom gemacht werden und dass die auf der anderen Seite auch zielgerichtet agieren. Das heißt, meine aktuelle Frage ist - und da achte ich gerne drauf - wann sind wir so weit, dass diese KIs ohne uns funktionieren und dann neue Sachen schaffen, vorzugsweise mit einem Feedback-System aus der realen Welt?

Also die Frage ist, was kann vielleicht eine Gruppe an KIs oder eine große KI aus sich heraus schaffen, ohne dass ein großer menschlicher Einfluss auf diese KI ist? Kriegt man das autonom hin? Und an der Stelle muss ich sagen, dass diese Agentisierung sich gerade langsam durchsetzt.

In ChatGPT gab es diese GPTs, die machen teilweise, die erarbeiten sich Pläne und arbeiten die Pläne ab und gehen einmal ins Internet und suchen da Sachen. In der Softwareentwicklung ja sowieso. Agenten, die miteinander arbeiten, der eine schreibt einen Quelltext, der andere testet den, der andere schreibt Spezifikationen und ein anderer prüft das alles, das auf jeden Fall hier. Und dann ist halt die Frage, wie sieht es dann aus mit der Musik? Wann kommen diese Agenten und auf die sollte man achten.

**Interviewer:** Okay, dann halten wir ein Auge auf die Agenten, wann die kommen, wann der Mensch quasi nicht mehr den Faktor bestimmt dahinter; wann der Mensch nicht mehr da ist.

**Tristan:** Auch da; Es wird wahrscheinlich so sein, dass wir uns dann mit den KIs austauschen und von denen lernen wechselseitig. Ich glaube nicht, dass wir dann auch wieder den ganzen kompletten Markt abdecken und keiner mehr was macht, weil die Menschen machen wir immer noch was.

**Interviewer:** Der Mensch ist bisher immer geblieben, ne? Muss man so sehen. Okay. Ja, das wär's dann. Da würde ich mal Pause drücken, ne? Vielen Dank dir!

**Tristan:** Danke Roland. Hat sehr viel Spaß gemacht.

## Interview 2

### Vincent Bittner

**Interviewer:** Okay, dann läuft das jetzt. Ja, herzlich willkommen!

**Vincent:** Hallo, schön. Ich freue mich!

**Interviewer:** Schön, dass du da bist. Dann würde ich jetzt mal beginnen mit ein paar einleitenden Fragen zu dir. Kurz kannst du kurzhalten, wie du möchtest. Wie heißt du?

**Vincent:** Vincent Bittner. Vincent Marius Gabriel Bittner von und zu Obermichelbach – ne (lacht).

**Interviewer:** Schön, dass du da bist. Dann würde ich dich gerne befragen, was deine aktuelle Position ist oder dein Beruf.

**Vincent:** Aktuell bin ich, also die offizielle Bezeichnung ist Video und Web Content Production Manager. Basically, Medienproducing. Hauptsächlich ist der Fokus auf Video, aber auch solche Sachen, wie zum Beispiel Podcast-Aufzeichnungen, eigentlich sämtliche Medienaufzeichnungen im audiovisuellen Bereich fallen darunter, inklusive Konzeption, Produktion, Postproduktion, von vorne bis hinten der ganze Prozess. Leider nicht so viel Musikalisches im beruflichen Feld, das ist dann eher im Privaten. Fällt es da auch schon mit rein? Oder ist es die nächste Frage?

**Interviewer:** Das fällt da auch schon mit rein.

**Vincent:** Ich produziere auch Musik unter verschiedensten Pseudonymen, vorrangig elektronische Musik, spezifischer Techno, noch spezifischer Raw, Deep, Hypnotic. Also sehr mit sehr viel Fokus auf ja, ich sag mal, perkussive Elemente, wenn man es so nennen möchte und weniger melodios. Auf dem Hauptprojekt unter dem Namen Leonhardt März habe ich über 50 Releases auf verschiedensten, auch internationalen Labels. Ja, genau.

**Interviewer:** Okay, sehr schön. Und kann ich dich dann fragen, wie du dahin gekommen bist, zu der Musikproduktion, zu Leonhardt März, kurz zusammengekommen?

**Vincent:** Ja, ich habe als Kind klassisch Musik Unterricht gehabt, so Föte, Klavier und sowas und wie es dann halt so ist, als Kind sagt man, oh das ist voll blöd, ich will viel lieber Videospiele spielen (lacht).

Und dann später kam das aber eben wieder so, auch durch die Familie geprägt, recht musikalisch, recht viel mit Musik unterwegs gewesen und dann festgestellt, dass ich es mag, Musik zu spüren, also physisch. Deswegen dann auch Bass gelernt, so als klassisches Instrument und durch härtere Rock- und Metal-Musik bin ich dann irgendwann zu Techno gekommen, ich weiß nicht, da gibt es immer so eine Brücke irgendwie. Festgestellt, das taugt mir sehr gut, weil es so ein, um es jetzt mal vielleicht ein bisschen esoterisch zu benennen - oder es ist so eine Urform, finde ich, als Mensch. Es ist so Getrommel und dazu tanzt man sich in Hypnose wie so Höhlenmenschen (lacht). Dann dachte ich mir, muss ich auch mal selber probieren zu machen, ewig viel rumprobiert,

am Anfang noch ganz viel so EDM-mäßig, bis ich dann gemerkt habe, okay, mir gefallen diese kleineren Nuancen mehr und wenn es wirklich hypnotischer bleibt. Aber es entwickelt sich immer weiter, man merkt immer mehr verschiedene Einflüsse, die reinfließen, deswegen so super exakt festlegen kann man sich dann am Ende doch nicht, was es ist.

**Interviewer:** Okay super, sehr schön. Ja du bist heute hier, da wir ein bisschen auch noch über KI reden, vor allem in Bezug zu der Musikproduktion und ich habe das folgende Interview ein bisschen in Kategorien eingeteilt und würde dementsprechend immer so ein bisschen von Kategorie zu Kategorie gehen, einleiten, wenn du möchtest und würde dann jetzt auch mal zum Hauptteil kommen. Zum aktuellen Stand würde ich dich ganz gern befragen, wie du dazu stehst, also beziehungsweise zur historischen Entwicklung, sagen wir mal der letzten Jahre vielleicht und im aktuellen Stand der KI in der Musik, nicht der KI allgemein, was würdest du denn sagen, wie hat sich die KI in der Musik in den letzten Jahren dann entwickelt bis zum Stand heute?

**Vincent:** Also es ist die Frage ab wo man im Allgemeingebrauch von KI spricht, weil es gibt ja schon länger jetzt auch solche Sachen, wenn ich jetzt an zum Beispiel Plug-Ins denke in der Musikproduktion, wenn man dann schon sagt, okay ich benutze jetzt zum Beispiel so ein Mastering Tool, Izotope Ozone mit einem Mastering Assistant, würde man das schon mit zu KI zählen oder ist es wirklich einfach nur derselbe Prozess, der da immer abläuft, ohne irgendwelche neuronalen Netzwerke dahinter? Ich würde jetzt eigentlich sagen schon, dass ich da so die Anfänge der aktuellen Entwicklung sehe, aber inzwischen jetzt gerade in den letzten Jahren, in den letzten Monaten, gerade auch mit der allgemeinen KI-Entwicklung, die großen Durchbrüche, die da auch bekannt wurden allgemein, jetzt wie so Chat GPT, haben dazu geführt, dass es jetzt bei Musikproduktion allgemein auch sehr, sehr stark vorangegangen ist, dass man jetzt nicht mehr nur noch so, dass man mehr hat als eben solche simplen, naja gut simpel ist auch untertrieben, dass man mehr hat als Tools, die eben nur eine Sache können, dass ich mehr Möglichkeiten habe, als jetzt nur irgendwelche Text-to-Voice Generatoren, sondern dass wir jetzt hier angekommen sind bei so zum Beispiel Tools, wo man verschieden prompten kann, was man möchte mit oder ohne Text. Das ist ja dann noch mal ein anderes Modell vermutlich, was da läuft, was den Text schreiben würde und daraus dann auch wirklich komplette Tracks schreiben kann, von vorne bis hinten, die fertig sind, bei denen man klar noch so einen Unterschied merkt zu, ich sage jetzt mal, nicht vollständig KI produzierter eigener Musik, hochwertig produzierter eigener Musik. Aber es ist auf jeden Fall eine exponentiell steigende Kurve, die da stattfindet in allen Aspekten der Produktion, also von Sound Design zu Arrangement zu Ideenfindungen, ist alles abgedeckt eigentlich. Manches mehr, manches weniger.

**Interviewer:** Okay, okay. Und kann ich dich fragen, was du dann da aktuell am spannendsten, am vielversprechendsten findest, was hat dich da am meisten beeindruckt, was passiert ist?

**Vincent:** Also was mich wirklich sehr, also was mich wirklich geflasht hat, weil ich das irgendwie relativ spät mitbekommen habe, war Suno AI, was wirklich, ich hatte da auch ein Gespräch dann mit einem Bekannten, man war ja wirklich nur promptet und dann bekommt man einen, na ja, kurzen, aber trotzdem eigentlich fully fletched out Song raus. Ich weiß noch, dass wir uns darüber, als wir das gefunden haben, unterhalten und er meinte noch so, es ist krass, dass sogar die Kick definiert ist und so auch für komplexere Musikrichtungen ein bisschen mit rumgespielt, wenn man dann in den, ich sag mal, weniger strukturierte Musik geht, die jetzt nicht so formell, formularisch aufgebaut ist, dass da die Bereiche gut abgedeckt sind. Das hat also immer, als Beispiel bei Suno AI, wenn man da irgendwie, mein Favorite ist dann immer 80s japanischer Fusion Jazz und hier bitte dann noch ein Solo rein und dann kommen da Sachen raus, wo man sich denkt, okay, das könnte ein echter Song sein, der auch so von kreativen Menschen geschrieben sein könnte. Ja, genau.

**Interviewer:** Okay, ja, das waren schon spannende Sachen, die da passiert sind mit Suno. Und würdest du auch sagen, dass das Stand heute schon Relevanz hat für die Musikindustrie oder ist das noch nicht so weit?

**Vincent:** Also in der Industrie, ich würde sagen, also es ist, ich sehe viele Parallelen zu anderer AI, ich muss vorneweg sagen, es nervt mich auf jeden Fall unfassbar, dass KI - viele Modelle, die gerade entwickelt werden, wie die eingesetzt werden - dass die kreative Aufgaben übernehmen sollen, weil das ja eigentlich die Sachen, die Menschen Spaß machen, so. Fände ich es geiler, wenn mehr Arbeit reingesteckt werden würde in KI, die Sachen übernimmt, auf die man eigentlich keine Lust hat. Und ich würde sagen, Industrie ist aktuell Stand, es ist, also wahrscheinlich noch teilweise verpönt, ist ja auch oft so bei neuen Technologien, wenn ich jetzt auch zum Beispiel ganz kurz die Brücke schlag zum Auflegen als DJ; als die Sync Funktion kam, die ist bis heute noch für Speed Matching „oh, wer das automatisch macht, der ist kein echter DJ.“ Da wird natürlich dann bei der KI in Musik auch viel diskutiert, genauso wie in anderen KI-Bereichen, wie jetzt auch bei Grafik und sowas, das eben fehlt, der, weiß ich nicht, ob das der offizielle Begriff ist, aber eben dieser Divine Spark of Creativity, der halt im Menschen verankert ist, so was kann die KI nicht erreichen und soll sie auch nie können oder wird sie auch nie können. Deswegen wird kritisch beäugt, aber es gibt natürlich hilfreiche Tools, die nützlich sein können, wenn man jetzt zum Beispiel sagt, okay, gerade als, wenn man jetzt nicht professionell unterwegs ist und sagt, okay, ich habe kein Geld und ich habe keine Ahnung von Mastering, dann kann man sich eben solche Sachen, solche Tools zu Nutze machen wie Dolby Mastering KI, die das Ganze, ja, dann für einen mastert, was dann, ich kenne das Tool jetzt nur vom eigenen benutzen, ich weiß nicht, was dahinter steckt, ob da irgendwie neurales Netz dahinter steckt oder irgendwelche Modelle, aber so was kann man sich dann zu Nutze machen, eben die eigenen Sachen zu erweitern. Aber jetzt gerade in Industry Professionals, würde ich mal sagen, in den technischen Abläufen ist es eher verpönt und wahrscheinlich auch in den meisten Fällen schlechter als was von den Menschen gemacht wird, also noch, aber eben je mehr man in diesen, in den amateurhaften, ins Hobbymäßige geht, desto mehr kann man da einsetzen.

**Interviewer:** Okay, das heißt also von der technischen Seite aus, glaubst du, es ist noch verpönt, wie sieht es denn mit der kreativen Seite aus, dann können wir da gleich die Brücke dazu schlagen. Was denkst du denn, welche Rolle spielt denn noch der, wenn man jetzt KI-Tools verwendet, der menschliche Input dahinter und deine Intuition als musikschaaffende Person mit der KI?

**Vincent:** Also einerseits ist es ja so, dass die meisten Modelle, die werden ja mit irgendwas trainiert, da könnte man jetzt sagen, okay, da kriege ich halt nur Sachen raus, die so ähnlich klingen, wie Dinge, die bereits existieren. Andererseits ist es aber auch so, dass man wirklich in einem Großteil der Fälle, wo Musik gemacht wird, lässt man sich ja selber auch von was anderem inspirieren, was dann so ähnlich klingt - jetzt nicht vielleicht unbedingt vom Text her oder von der Melodie oder von den Zahlen, aber auch einfach vom Gefühl her. Von daher würde ich sagen, ob es echte Kreativität dann überhaupt so gibt, kann man dann wieder anzweifeln, was mit den Ideen gemacht wird (zögert). Jetzt habe ich mich dann ein bisschen verrannt gedanklich gerade.

**Interviewer:** Alles gut! Ich würde jetzt sagen, ich würde dich jetzt fragen, wie es dann bei dir wäre mit dem Input. Du bist ja natürlich auch beeinflusst von anderen Sachen und die KI ist beeinflusst von anderen Sachen, aber wie wichtig ist dann deine Beeinflussung dann noch bei der Arbeit mit KI jetzt, wenn man mit KI arbeitet?

**Vincent:** Genau, die Inputs. Also natürlich ist es dann einerseits das Prompting ist sehr wichtig. Ja, man muss da nicht immer genau wissen, was man will, man kann ja auch einfach ein bisschen ausprobieren und gucken, ob was Cooles dabei rauskommt. Aber ich finde es zumindest wichtig, dass man, um da so eine gute, ich sag mal, Symbiose herzustellen zwischen KI und Produzierenden, dass man damit dann noch irgendwie weiterarbeitet, dass man damit dann noch mehr machen kann und das nicht einfach gegeben hinnimmt. Ich sehe eigentlich recht viele Parallelen zu Sampling, wo man ja blöd gesagt auch fertige Sachen nimmt, die man bekommt, woher auch immer, aus dem Schaffungsprozess von jemand anderem und die dann weiterverarbeitet. Und so würde ich das bei KI auch sehen, nur dass man eben durch den eigenen Input an dem Schaffungsprozess beteiligt war, dann kriegt man es von der KI und dann verarbeitet man das noch weiter.

**Interviewer:** Und das heißt, du fühlst dich da in ähnlichem Umfang auch kreativ beeinflusst, wie jetzt sagen wir mal durch Samples. Also könnte man das vergleichen, dass man die KI-Ergebnisse durchgeht mit einem Sample-Pack, das man sich besorgt hat und das durchgeht?

**Vincent:** Ja schon, also ich habe jetzt bei mir selber noch nicht den Fall gehabt, weil ich mein Sounddesign nicht so sehr mit KI mache. Aber wenn ich gerade auch so an die Suno-Vergleiche denke, da gehe ich dann auch 10, 20 mal dasselbe generiert mit leicht verschiedenen Inputs und dann gehe ich die durch und gucke, was ist da jetzt das Beste, wo würde ich denn da jetzt ansetzen? Also würde ich schon so handhaben, vielleicht auch nicht die mal ein Sample-Pack mit KI zu erzeugen und damit dann zu arbeiten. Das klingt nach einem Konzeptalbum.

**Interviewer:** Nur kein Geld dafür verlangen (lacht). Ja, wobei könntest du wahrscheinlich.

**Vincent:** Ich weiß nicht, wie die rechtliche Lage ist. Also ich meinte, dass es bei -aber da kann ich auch komplett falsch liegen, deswegen zitiere mich da nicht ohne zu gucken -dass es bei Grafik einen Gerichtsbeschluss gab, dass KI-generierte Werke keine Schöpfungshöhe haben.

**Interviewer:** Ja, doch, das war so, dass es bei Musik aktuell auch so, soweit ich weiß, deshalb. Schwieriges Gebiet (lacht).

**Vincent:** Schwieriges Thema. Einen Anwalt müsste man da jetzt haben (lacht).

**Interviewer:** Der nächste Interviewpartner dann (lacht). Na gut, aber wenn du jetzt, na ja, sagen wir mal, du arbeitest jetzt mit dieser generativen KI oder du, du verwendest die, glaubst du, dass die zur Entwicklung neuer Stile auch beitragen kann oder zur Erweiterung bereits bekannter Genres, so, sagen wir mal, also als konkretes Beispiel, du machst deine Musik und möchtest die vermischen mit einem anderen Genre, versuchst das aber zunächst erst mal in KI.

**Vincent:** Ja, auf jeden Fall, auf jeden Fall. Da kann man schnell Ideen sketchen und manchmal kommen dann gegebenenfalls auch Sachen raus, die man so nicht erwartet hätte, was wiederum dann Inspiration schaffen kann, in die Richtung zu gehen. Ich hätte jetzt auch, wenn du kein Beispiel gesagt hättest, dann hätte ich auch gesagt, Genres vermischen. Das ist das perfekte Beispiel dafür, wie man sich es zu Nutzen machen kann. So sagen, aber wie würde sich das denn anhören, wenn ich jetzt wirklich zwei Sachen miteinander kombiniere, dass es bisher noch nicht so gibt. Da ist ja dann auch der Gedanke dahinter, warum gibt es das noch nicht? Warum hat es noch keiner gemacht? Wo ist der entscheidende Punkt, dass es dann am Ende funktioniert und da kann man halt relativ schnell relativ viel ausprobieren. Das geht auf jeden Fall, ja.

**Interviewer:** Ja? Geht auf jeden Fall. Gut. Schön. Ja, dann würde ich dich gern auch dazu fragen, wie du denn interagierst mit der KI oder fangen wir mal an damit; wann hast du dich denn das erste Mal damit auseinandergesetzt oder sie verwendet?

**Vincent:** Ja, natürlich, was ich von schon meinte, diese Ursprungsformen von Ozone Master Assistant, Dolby KI zum Mastering... (denkt nach) 2017 oder so, habe ich die das erste Mal benutzt, schon ein bisschen her. Also tatsächlich KI in Musikproduktion, da muss ich jetzt nachgucken, das war letztes Jahr, da habe ich eine Stimme, eine Voice Line mit Lyrics generiert für eine Produktion mit einem Tool, das es inzwischen nicht mehr gibt, aufgrund rechtlicher Bedenken, die bekannten Stimmen imitiert haben oder geklont eben. Das war so der Wendepunkt, da kam es gerade so auf mit diesen Voice Cloning Tools, wo ich auch einige größere Beispiele kenne, die es ja ähnlich gemacht haben, die exakt das gemacht haben.

**Interviewer:** Ja, es gab ja den Fall bei dem Jay-Z-Song, bei dem Jay-Z nicht vertreten war, sondern das die KI war, mit der produziert wurde.

**Vincent:** Also das ist so der, bislang der größte Punkt, wie ich da KI benutzt habe, ist eben dieses Erstellen von Lyrics, Gesang, gut in dem Fall war es jetzt eher Rap oder Spoken Word, weil da auch, sag ich mal, bei mir dann persönlich die größte Schwäche ist, da was herzubekommen, wo ich dann sage, okay, da von KI unterstützt zu werden, nehme ich gerne mit.

**Interviewer:** Ja, und wie ist das jetzt aktuell noch, wenn du sagst, das ist ja schon eine Weile her, wie verwendest du die KI denn dann aktuell, also wo unterstützt du dich am meisten, im gesamten Prozess?

**Vincent:** Also im gesamten Prozess unterstützt mich die KI tatsächlich am meisten beim Release (lacht). Ja, mit nächtelangen Diskussionen, aber halt über eben Formalitäten. Ich habe neulich Works in Progress von mir hochgeladen und Mix and Master analysieren lassen und mir da dann Feedback geben lassen von der KI, was jetzt nicht so hilfreich war, muss ich mal sagen, aber das liegt eben auch wieder an der Nische, da gibt es einfach nicht so viele Daten dazu, zu der Musik, dass da KI ausgereift genug ist zu sagen /. Die sagt dann natürlich, oh, das ist viel zu wenig, hier, Mitten, aber das ist gewollt so, dass da nicht so viel in der Mitte passiert oder nicht durchgehend zumindest. Das war jetzt so der letzte Moment, wo ich tatsächlich AI eingesetzt habe, für die Produktionen, ansonsten ist es wirklich auch eher ein Spielzeug aktuell für mich, da ich sage, okay, ich habe hier einen Song fertig, wie könnte der sich denn anhören, wenn er in einem anderen Genre wäre oder wie würde es denn klingen, was in die Richtung zu machen? So wie, wie, wie denkt die KI? Was sagt die? Das sind die nennenswerten Punkte dieser Musik und daraus kann ich dann wieder Learnings ziehen, jetzt gerade so mit AI-Stimmen, also mit Lyrics, in dem Sinne arbeite ich nicht mehr. Einerseits, weil es eben auch dieses Super Tool, Über Duck hieß es, das gibt es nicht mehr in der Form und ansonsten war auch aktuell keine Notwendigkeit da, was zu erschaffen. Was halt jetzt dann interessant wäre für mich, aber so ganz bin ich da noch nicht drin, ist dann tatsächlich die Sound Creation, also das Sounddesign von Synthesizer Sound, sage ich mal.

**Interviewer:** Ja, das sind wir ja aber noch relativ im Anfangsstadium verhältnismäßig. Könntest du dann aber beurteilen, ob sich deine Qualität jetzt verbessert hat? Also deine Endproduktion durch die Verwendung von KI-Tools? Oder würdest du jetzt, wie du gesagt hast, sagen, du ignorierst das dann trotzdem, was die KI sagt, weil es nicht spezifisch genug auf deine auf dein Genre ausgerichtet ist?

**Vincent:** Also man nimmt, also Learnings nimmt man immer mit auf jeden Fall. Aber da muss ich halt auch sagen, die nimmt man auch mit, wenn man sich irgendwo anders Feedback einholt oder wenn man sich irgendwo anders Informationen anschaut, die eben auch Mensch gemacht sind, aber es ist halt trotzdem eine gute Möglichkeit, ich sage mal, nachts um 4 Uhr früh mal schnell eine zweite Meinung einzuholen. Ich denke in anderen Fällen kann das mehr helfen, wenn man jetzt mehr in Richtung wirklich Populärmusik geht, Pop Musik geht, aber ansonsten, ich muss auch sagen, jetzt zum Beispiel, der Song mit dem AI-Gesang, der wäre ohne Lyrics, wäre der auch scheiße (beide lachen). Also da hat es das auf jeden Fall verbessert (lacht)!



**Interviewer:** Ja, okay, wenn du das als Schaffender dann dazu sagst, das wäre dann aber jetzt auch eine Frage: Würdest du sagen, dass die KI dadurch, dass sie einen großen Beitrag geleistet hat zum persönlichen Wert für dich jetzt, dass die KI quasi als Partnerin fungiert hat, für den Song spezifisch jetzt zum Beispiel, oder sogar die Person selber, die du imitiert hast?

**Vincent:** Also, ich würde die KI nicht als Partner sehen, also nicht in dem Zustand, in dem sie gerade ist, sobald es, also das ist ja diese spezialisierten KI, die für eine Sache da sind, sobald wir so einen irgendwann einen AGI, also diese General KI haben, da wäre wahrscheinlich der Punkt, wo ich sage, hier sage ich, okay, ich arbeite zusammen mit KI.

Und dann wirft es ganz viele ethische Fragen auf und ganz viele gesellschaftliche Fragen, über die ich – für die jetzt nicht so qualifiziert bin. Aber aktuell ist die KI selber, die ich nutze, das ist wie ein Synthesizer, den ich hier stehen habe, den man nur anders bedient sozusagen. Und die Personen jetzt bei so Stimmenemulatoren, mit der da getraint wurde, das Modell, die hat dann wahrscheinlich mehr Beitrag geleistet, weil die Person hat ja eben diese Stimme und diese bestimmte Tonalität und Rhythmus und Sprechungsflow, worauf es eben alles basiert.

**Interviewer:** Okay, ja, wenn du schon die ethischen Überlegungen angesprochen hast; Hast du denn, fällt dir eine Problematik ein, eine ethische? Welche wäre für dich die größte, momentan mit KI-generierter Musik?

**Vincent:** Das größte ethische Problem bei KI, es ist vielleicht, wie drück ich das denn aus? So, im modernen gesellschaftlichen System, in dem wir uns befinden, was ja auch sehr, ich sag mal kapitalistisch geprägt ist, werden künstlerische kreative Sachen eh nicht so geschätzt oder so vergütet, wie sie es von der Schöpfungshöhe haben. Und ich habe das Gefühl, wenn KI auf diesem Level ist, dass man den Menschen nicht mal mehr dahinter braucht, um jetzt irgendwie Musik zu haben für was auch immer, dass dann Kreativität noch unwichtiger wird aus gesellschaftlicher Ebene und dass das ein bisschen so weniger gefördert wird, was ja schade wäre. Einerseits das und natürlich, da hängt dann alles andere mit dran, dass man jetzt auch sagt, man hat auch irgendwie Leute, die Stockmusic produzieren oder sonstige Sachen, die dann, die es dann nicht mehr braucht, die dann auch erst mal gucken müssen; oh wie verdiene ich denn jetzt überhaupt Geld, wie kann ich überhaupt jetzt überleben? Ist vielleicht ein bisschen mehr Kapitalismuskritik als KI-Kritik.

**Interviewer:** Naja, wenn man sich überlegt, wo die großen KIs herkommen, von welchen Unternehmen, ist das wahrscheinlich berechtigt. Du hattest erwähnt auch mit der Schöpfungshöhe, wie siehst du das dann mit KI und, sagen wir nicht nur der Schöpfungshöhe, vielleicht sogar auch der Authentizität der generierten Musik? Also ich würde jetzt sagen, wir müssten hier unterscheiden zwischen der komplett generativen Musik und sagen wir mal, wie du das verwendest hast, der KI als Tool. Ist das genauso authentisch wie, sagen wir mal, nur menschlich produzierte Musik?

**Vincent:** Es ist wieder schwer zu nuancieren. Also, ich meine, die Schnellantwort wäre nein, es ist nicht vergleichbar. Jetzt wo ich gerade gesagt habe, es ist wie ein Synthesizer, wenn ich überlege, gerade wenn ich Musik mit KI generiere, kommt auch viel auf den Input an. Also ich kann auch mit einem modularen System, wenn ich da bestimmte Kabel stecke und bestimmte Voreinstellungen treffe, dann kann es auch eine Stunde laufen und selbstständig eine Stunde lang Musik produzieren, die dann einen ähnlichen Vibe hat, aber immer durch irgendwelche Zufallsaktionen ein bisschen ja Nuancen reinbringt, ein bisschen verschiedene Dinge, die sich da dann in dem Leitmotiv bewegen oder was auch immer. Und so wäre das bei KI auch. Das heißt, ich würde tatsächlich eher vom User Input abhängig machen, als tatsächlich von dem Endergebnis, dass man sagt, die Schöpfungshöhe hier ist höher oder niedriger.

**Interviewer:** Okay, und würdest du dann auch nochmal unterscheiden, sagen wir zum Beispiel mit deinem Synthesizer jetzt, der von sich selbst quasi kreiert, du würdest davon jetzt einen Teil verwenden für eine Produktion. Aber das ist ja nur ein Teil der Produktion. Würdest du das, wenn jetzt ein Teil der Produktion KI wäre, der Rest aber nicht, nochmal unterscheiden zu rein KI generierter Musik?

**Vincent:** Ja, das würde ich. Da sind wir wieder bei dem Vergleich zum Sampling, wo ich sage, das wird dann weiterverarbeitet, da findet dann irgendwie dieser Transfer statt von du hast was generiert, du arbeitest damit weiter und du nutzt es als Grundlage oder als Ergänzung für zusätzliche Inspiration.

**Interviewer:** Ja, okay, und dann hätte ich noch die Frage dazu: Du bist ja auch in der, also Musik ist ja auch Kultur, muss man sagen, und du bist da natürlich tief drin in deinem Genre, in deiner Szene, sage ich jetzt auch mal. Glaubst du, dass KI, also die generative KI, dass sie diesen kulturellen Ausdruck und die Zugehörigkeit abbilden kann? Quasi die Genre-Authentizität?

**Vincent:** Also so von einem der Standpunkte, zumindest von Techno aus, kann das KI auf jeden Fall. Weil im Techno geht es ja auch um moderne Technologien (lacht). So, da sehe ich es tatsächlich am ehesten, klar wird es dann auch wieder Splitterfraktionen geben in der Kultur, die sagen, oh, das ist aber nicht mehr wie die Roots sind, aber es geht ja eben auch darum, dass man die Moderne betrachtet und erforscht und guckt, was kann ich für Technologien verknüpfen. Jetzt aber so von der menschlichen Kultur schwierig. Gerade auch weil viel elektronische Musik ist sehr inklusiv, anarchistisch, politisch auch, und klar kann man das ja einfließen lassen in Produktion oder in Sounddesign mit KI, je nachdem wie man promptet, aber man hat aber am Ende wieder große Korporationen hinter bekannten Modellen stehen, wo dann die Frage ist, ist das denn kulturell vertretbar in der Szene?

**Interviewer:** Also das heißt, es fehlt am Ende dann doch wieder der Mensch hinter der Musik?

**Vincent:** Genau, genau.

**Interviewer:** Dann würde ich dich gerne mal auf deine Prognosen, Einschätzungen für die Zukunft befragen, auch wenn wir schon in die Richtung gehen, wie das ja in Zukunft

vielleicht dann sein könnte. Glaubst du aber, dass die von KI generierte Musik auch eine eigene kulturelle Identität entwickeln könnte?

**Vincent:** Sehr, sehr weit in der Zukunft vielleicht, wenn es so cyberpunk-mäßig ist (lacht), mit so Androiden und so, ne. Wobei wenn ich das jetzt so ausspreche, es gibt schon so technologisch, ich sage mal extreme Menschen, die eben sehr viele neue Technologien sehr gut abfeiern. Und ich sehe da schon den Kreis der, ich sage jetzt mal Nerds, hocken, die sagen, hier ich habe die neue Platte von der neuen so-und-so-KI. Gut, vielleicht nicht Platte, aber das neue Album. Und dann kann es schon krass sein, was dabei rauskommt und irgendwelche Sounds, die man noch nie vorher gehört hat. Das heißt, es könnte sich schon ein kultureller Kreis drum bilden, oder wird sich ziemlich sicher, weil es wirklich alles gibt. Aber da ist dann eben auch wieder der Mensch ausschlaggebend, die Menschen, die das dann eben hören und dann eben diese Kultur prägen. Abgesehen davon sehe ich den, also in kürzerer Zukunft sehe ich den Einsatz von KI in Musik wirklich mehr bei dem Sounddesign, eben an diesem Sampling, ich habe kein besseres Wort dafür, die wir weiterverarbeiten und halt viel im kommerziellen Sektor. Also wenn ich einen Klamottenladen habe und da soll Musik laufen, die die Kunden zum Kaufen anregt, wieso soll ich irgendwelche Lizenzgebühren bezahlen, wenn ich mir auch einfach da Musik generieren lassen kann unter gewissen Aspekten? Oder wenn ich, da habe ich auch gelegentlich bei mir im beruflichen Leben, wenn ich sage, ich brauche Musik im Hintergrund bei Videos, bei Werbung und man möchte irgendwas, was nicht dieser Zahnarzt-Werbung-Corporate-Pop ist, der wahrscheinlich auch schon sehr formularisch und wahrscheinlich auch schon unter Hilfe von KI angefertigt wird inzwischen. Das so für die nähere Zukunft und dann für die weitere eben, genau wie gesagt, sich da schon, dass sich da kulturelle Kreise aus Menschen bilden, die das abfeiern, wahrscheinlich eher nischig bleiben und noch weiter in die Zukunft kann alles passieren.

**Interviewer:** Ja, allerdings. Da hast du schon auch eine Frage von mir gleich mit beantwortet, wie du das siehst und sagen wir speziell auf die Musikproduktion ausgerichtet aber; was siehst du da in der Zukunft? Was glaubst du wird, was werden die kommenden Tools beherrschen, was die Jetzigen vielleicht auch nicht können? Also mir fällt zum Beispiel ein, was jetzt die letzten zwei Jahre spannend war, was viel zugänglicher geworden ist, zum Beispiel die Stampseparation. Denkst du, was wäre denn, was wird da kommen? Was glaubst du?

**Vincent:** Also ich bin auch sehr gespannt, wir hatten ja jetzt auch ein paar Jahre lang, gab es diese große Plug-in-Flut, wo man, wenn man irgendwas gesucht hat, dann kamen zehn Plug-ins, wo es hieß, ja das Plug-in brauchst du dafür und das brauchst du dafür und da sind wir jetzt aktuell wieder bei dem Schritt, okay, nee, eigentlich braucht man es gar nicht so wirklich und ich denke, wir werden wahrscheinlich auch erst mal mit KI-Möglichkeiten geflutet werden in der Produktion, in der Produktionshinsicht. Also das Nachbauen von bestehenden Sounds, das finde ich sehr interessant, das kenne ich auch noch nicht so lang. Das KI das inzwischen schon kann oder wäre natürlich nicht 100 Prozent perfekt, aber es in die Richtung geht. Das ist so die eine Sache, also dass man eben bestehende Sounds eben aus Songs oder wo auch immer nimmt und die sich von

der KI nachbauen lassen kann und dann auch den Prozess dahinter sieht, wie ist dieser Sound aufgebaut. Wenn ich jetzt überlege in verschiedenen produzierenden Kreisen, sage ich jetzt mal, indem ich mich bewege, wie oft da die Frage fällt so, ich habe diesen Sound in einem Song gefunden, der ist super, wie wurde der gemacht? Solche Fragen können dann alle einfacher gelöst werden auf jeden Fall.

Klar, wir haben solche Sachen wie Stemseparation und eins hatte ich gerade noch, was ich zukünftig mehr sehe. Ja, tatsächlich dann auch in jetzt DAWs zum Beispiel prompt basierte, komplett prompt basiertes Sounddesign. Dass man reingehen kann und sagt, okay, ich brauch jetzt einen pluckigen Synthesizer-Sound, der modern klingen soll, metallisch mit vier Stimmen gleichzeitig, die aber leicht disharmonisch und durch einen Ringmod gejagt werden und alles. Man sich sozusagen diese ganzen Arbeitsschritte vereinfacht. Was auch, glaube ich, kommen wird, wenn es das nicht sogar schon gibt, diese Copilot-Programme, dass man beim Musik produzieren den AI Copilot mit drin hat, der mithört und sagt, hey, der Vorschläge gibt oder sagt, oh, das hier ist zu laut oder guck mal auf deinen Mix, du bist hier jetzt am Clippen und einfach so generelle Inputs gibt bei dem kompletten Workflow.

**Interviewer:** Okay, also es geht überwiegend auf Workflow-Optimierung.

**Vincent:** Ja, das würde ich sagen.

**Interviewer:** Und glaubst du, dass sich die Qualität der Musik in Zukunft dadurch steigern wird oder gleichbleibt oder verschlechtern wird, dadurch, dass ja quasi dann jeder, sagen wir mal, einfach nur ein paar Prompts eingibt und dann das erstbeste Ergebnis nehmen könnte im Vergleich zu den Sounds selber craften und implementieren. Glaubst du, dass die Qualität dann ein bisschen abnehmen wird in Zukunft? Oder zunehmen, also wenn die AI gut arbeitet, kann sie ja auch zunehmen.

**Vincent:** Als automatisches Speed-Matching für DJs kam, war plötzlich jeder DJ und es gab zehnmal so viele Leute, die gesagt haben, ich bin jetzt DJ, aber es gab dementsprechend auch sehr viel mehr, die einfach sehr schlecht sind oder eben unteres Mittelmaß. Und die, die standardmäßig schon besser waren, die haben das dann genutzt, um noch besser zu werden. Das heißt, ich würde sagen, es wird auf jeden Fall viel mehr Musik geben, noch mehr als jetzt. Es ist ja jetzt schon so viel geworden im Vergleich zu, wenn man sich überlegt vor zehn, zwanzig Jahren, weil einfach die Möglichkeiten für den Otto-Normalverbraucher einfach da sind.

Geiles Wort (lacht). Und ich denke, es wird auch dadurch, ich denke, es wird sich ähnlich entwickeln und es wird eine größere Kluft geben. Also die, die ohne KI-Tools schon krasse Musik machen, die aus welcher Ansicht auch immer besonders gut ist, sage ich mal, es ist ja dann wieder subjektiv, was gut ist und was nicht, die werden unter dem Einsatz noch besser werden und viele, die sonst nichts machen, werden wahrscheinlich nicht so tolle Sachen am Anfang damit erzeugen können. Ich will jetzt auch gar nicht produzieren sagen, aber grundsätzlich ist es schon eine gute Sache, weil je mehr Leute es machen, desto mehr ist die Wahrscheinlichkeit, dass was richtig cooles Neues und

aufregendes dabei ist. Deswegen, ich weiß nicht, wie ich es einschätzen würde, ob es im Durchschnitt besser oder schlechter wird.

**Interviewer:** Würdest du aber sagen, es würde mehr werden auf jeden Fall?

**Vincent:** Es würde auf jeden Fall mehr. Es würde auf jeden Fall mehr. Vielleicht gleicht sich das im Schnitt wieder aus, aber es wird dann wirklich subjektiv. Wenn ich jetzt sage, es kommen 1000 super schlechte Lieder raus, die ich mir nicht anhören kann, aber dafür kommt eins raus, was mich so unfassbar beeindruckt und bewegt, dann könnte ich schon wieder sagen, es ist gleichgeblieben wie jetzt der Stand ist, wo ich sage, 10 von 100 Liedern finde ich cool. Wie bewertet man das, wo legt man es an? Ist ja nicht messbar.

**Interviewer:** Nicht in dem Umfang, das stimmt. Dann hätte ich dazu noch mal was deine Meinung würde mich interessieren. Was glaubst du, muss man denn in Zukunft tun, um als menschlicher Producer, was muss man tun, um sich von der KI dann abzuheben? Was denkst du, was ist die Prognose für die Zukunft?

**Vincent:** Physisch live performen.

**Interviewer:** Etwas, was die KI nicht kann, ne?

**Vincent:** Ja (lacht), wenn wir dann wieder in der fernen Zukunft sind, Androiden und so..., so solche Sachen auf jeden Fall, dass man eben die menschliche Kompetente mit reinbringt. Wird dann auch wieder schwierig mit KI-generierten Musikvideos und sowas. Und es gibt auch Konzerte von so Hologrammen, 2Pac und so was. Ist ja dann auch schon so hm, okay?

**Interviewer:** Aber steht ja trotzdem immer noch oder stand mal eine Person dahinter. Also die Leute, die auf ein 2Pac-Hologramm-Konzert gehen, gehen glaube ich trotzdem, weil sie 2Pac schon mochten, bevor das Hologramm existiert hat.

**Vincent:** Das stimmt. Ah, Hatsune Miku?

**Interviewer:** Hatsune Miku, ja. Aber ich glaube, dass das ja auch so ein nischiger Bereich ist. Und da gibt es ja mehrere Leute dahinter, die das produzieren. Also, es ist ja nur der visuelle Aspekt und der...

**Vincent:** Ja, das stimmt, das stimmt. Aber ich glaube, ich sehe viele Parallelen zwischen dem vielleicht entstehenden Kulturkreis von Leuten, die KI-Musik feiern und die Hatsune Miku feiern aktuell. Jetzt, wo ich so drüber nachdenke, irgendwo ist das schon, irgendwo gibt es das schon. Ja, aber ich glaube also wirklich, dass der Hauptunterschied ist. Und die Frage ist halt, ob man das dann nur als reine, hörende Person mitbekommt, ist dann halt wirklich die physische und menschliche Komponente.

**Interviewer:** Es fehlt ja, also Hatsune Miku ist zwar ja nicht echt jetzt in dem Falle, aber sie hat ja trotzdem eine Persönlichkeit, die ihr zugeschrieben wurde von Menschen wieder. Das hat ja die KI, also Stand, jetzt noch nicht. Also ich glaube, niemand würde jetzt auf ein Konzert gehen von Suno AI.

**Vincent:** Das wäre schon ein... Ich glaube, das wäre interessant (lacht). So, wie ist das in Szene gesetzt? Was passiert da? Mehr aus den Beweggründen.

**Interviewer:** Das würde ja aber trotzdem wieder bedeuten, dass man hingehht für die Charakterzüge, also um zu sehen, wie die KI dargestellt wird.

**Vincent:** Ja, das stimmt.

**Interviewer:** Das heißt, man möchte ja trotzdem eine Verbindung aufbauen zu der KI irgendwie.

**Vincent:** Ja, ja. Also es ist schon... Nee, ich will... Ich möchte gerne menschliche Komponente einloggen, durch die man sich abhebt. Okay. Und das ist eben auch nicht... Das ist klar, das kann die KI dann auch wieder... ..emulieren, dass es auch nicht zu 100 Prozent perfekt ist. Weil Imperfektion ist ja dann auch wieder ein menschlicher Aspekt. Und dann, wenn ich da dann wieder die Brücke schlage zu Technos-Szene/Auflegen, wenn man sich ein DJ-Set anhört und man hört, dass da irgendwie der DJ einen Fehler macht, dann denkt man sich - oder ich zumindest - gleich so, ah, geil, der legt richtig raw auf, richtig live, manuelles Speed-Matching. Und dann macht er halt einen Fehler. Das macht es mir nur sympathischer, weil ich merke, da steckt dieser DJ, diese DJane - DJ macht Fehler. Versucht dann da wieder rauszukommen aus dem Fehler. Und das macht das ganz einfach sympathisch, weil es nahbarer ist. Und man sich da eben auch identifizieren kann mit. Und dasselbe bei Live-Performance eben auch. Und ich glaube, das ist dann das Problem bei Musik, die von KI produziert wurde, dass man da einfach keine direkte Verbindung hat. Dass man sich dann nicht mit identifizieren kann in irgendeiner Weise. Außer eben, wie wird es in Szene gesetzt von Leuten, die das in Szene setzen oder die Person, die die Musik gepromptet hat?

**Interviewer:** Ja, ja, also selbst da, auch beim Prompting kommt es ja dann doch wieder darauf an, was man promptet. Also auf die Person. Stand jetzt. Dann würde ich dich abschließend auch nur noch fragen, was denkst du, worauf sollte man für die Zukunft ein Auge haben in Richtung KI und Musik? Sei es Tools oder generativer KI, was denkst du, wird kommen und nochmal einen großen Einfluss haben?

**Vincent:** Also ich denke tatsächlich, was so für den Großteil der Muskschaffenden Einfluss haben wird, werden diese Co-Pilots sein, die einen einfach komplett begleiten. Was ja aktuell eh schon, ich sag mal, so eine ähnliche Entwicklung haben wir ja eh schon im Sinne von... blöd gesagt YouTube Tutorials. Nur da muss man halt wissen, nach was man sucht und dann Glück haben, dass es das auch gibt und verständlich ist. Aber das wäre wirklich so das, würde ich sagen, würde ich vermuten, ist das nächste große Ding. In den DAWs, einen Co-Pilot eingebaut, der genau sieht, was man macht und dazu noch Tipps und Infos geben kann und auch Vorschläge und Ideen.

**Interviewer:** Ja, okay. Ja, dann würde ich sagen, das war's auch dahingehend. Ich habe alle meine Fragen beantwortet bekommen so weit.

**Vincent:** Okay, gut (beide lachen).

**Interviewer:** Wenn du sonst, ja. Also wenn du nicht noch irgendwas hinzufügen möchtest, würde ich sagen, das war's.

**Vincent:** Nee, habe ich schon am Anfang gesagt, dass mich das nervt, dass... KI soll meine Steuern machen und für mich arbeiten gehen, damit ich mehr Kunst und Musik machen kann und nicht andersrum (lacht). Das ist die Quintessenz.

**Interviewer:** Nehme ich auf, alles klar. Gut, dann bedanke ich mich.

**Vincent:** Gerne, gerne.

### Interview 3

#### Collins Boateng

**Interviewer:** Okay, let's do this. So, da hätte ich erstmal ein paar Einleitende fragen. Wie heißt du?

**Collins:** Ich bin Collins, Collins Boateng. Das ist mein Name, ja.

**Interviewer:** Sehr schön, Collins. Freut mich, dass du da bist.

**Collins:** Ja, gerne.

**Interviewer:** Und deine aktuelle Position? Hier ist... was machst du?

**Collins:** Also ich bin... also klar, hauptberuflich hier an der Hochschule der Medien bin ich als technischer Angestellter, betreue ich hier den Technikpool. Das heißt Geräteverwaltung, Studios unter anderem auch im Blick haben, wenn Studenten irgendwelche Einführungen brauchen. Und dann bin ich noch freiberuflich als Producer unterwegs, produziere hauptsächlich Hip-Hop-Musik, lege auch ein bisschen auf, aber ja, das sind glaube ich so die Keywords, die man braucht in diesem Thema, jetzt sage ich mal, ein bisschen was sagen zu dürfen.

**Interviewer:** Ja, cool. Und wie bist du zur Produktion dann gekommen, zur Musikproduktion?

**Collins:** Zur Musikproduktion? Ganz kurz, so. Ja, immer schon krasser Hip-Hop-Fan gewesen. Also klar, als Kind natürlich dieses klassische Musikzeug, erst mit Flöte angefangen, dann Trompete gespielt, und irgendwann fängt halt mit 14, 15 so die Sozialisierung findet dann statt, und dann findet man halt Hip-Hop irgendwie cool, dann will man da irgendwie Teil davon haben und sein und dann fängt man halt an Musik zu produzieren, merkt, ah ja cool, macht Spaß, dann machst du ein Studium, was dich dann irgendwie noch so ein bisschen drin ausbildet, dass du mal so dieses ganze tontechnische Zeug drumherum verstanden hast, und dann läuft es mit der Mucke irgendwie so, und dann ja, produzierst du halt Musik, und bracht halt Freude und Spaß, und ja, hat mir eine Zeitlang auch ein bisschen Unterhalt versorgen können, und so ist es dann ganz schnell und knapp zusammengefasst, ja.

**Interviewer:** Oh, danke, danke. Du bist da drin tätig gewesen, du hast deine Expertise.

**Collins:** Genau.

**Interviewer:** Gut, dann würde ich jetzt zum Hauptteil kommen, zum KI-Teil, und der Entwicklung und dem aktuellen Stand. Was denkst du, welche aktuellen Entwicklungen und Technologien im Bereich KI und Musik und deren KIs, sind am vielversprechendsten? Was war denn das letzte?

**Collins:** Am vielversprechendsten, finde ich tatsächlich, das ist jetzt echt eine Sache. Also, na, es ist immer so eine Frage, was für ein Genre man produziert. Ich jetzt speziell im Hip-Hop fand es sehr spannend, diese, wie soll man sagen, dass ich jetzt quasi einfach eine Melodie in mein KI-Tool rein (...), oder komplette Songs irgendwie in mein KI-



Tool reinballern kann, sage ich jetzt mal ganz frei, und du dann halt quasi diese Stem-separation machst. Also, die finde ich am spannendsten, auch im Hinblick jetzt auf diese Auflegerei, DJ-Geschichte, fand ich jetzt für mich persönlich, weil man da halt einfach ganz schnell gemerkt hat, dass jetzt auf Partys vor allen Dingen, Jungs auf einmal, Aca-pellas mit Instrumentals gemischt haben, das früher halt einfach nie funktioniert hätte, und das finde ich jetzt so vielversprechend in dem Sinn, weil die Leute da draußen, es gibt ja immer so die technologische Seite, wo die ganzen Nerds denken, ja geil, super toll, und dann gibt es halt so diese öffentliche Seite, also wo bekommt der normale Kon-sument auch mit, oh, da passiert was. Und da fand ich jetzt so, nachdem ich das mal so ein Jahr lang beobachtet habe, dass wirklich diese Stempseparation-Geschichte, dass die am spannendsten ist, weil das halt auch so eine öffentliche Wirksamkeit hat, weil du auf Partys halt einfach diese Mash-ups und Remixes jetzt auf einmal hörst, die du so erstmal hättest vorproduzieren müssen, und jetzt ist es zwei, drei Knopfdrücke, und dann wird halt Acapella mit den Drums und dem Instrumental gemischt. Das ist so mein Ding, was ich so am spannendsten fand, bisher.

**Interviewer:** Okay, cool, und gut zu wissen. Dann wäre meine Frage, wie - ja, das hast du schon zum Teil beantwortet -wie relevant das jetzt gerade ist für die Musikindustrie, für die Allgemeinheit?

**Collins:** Für die Allgemeinheit ist das, ja, also für die Industrie, als solches aus Konsu-mentensicht, glaube ich, ist es jetzt noch nicht so wahrnehmbar, aber für die Produzen-ten, glaube ich, ist es jetzt, glaube ich, wichtig, dass sie jetzt mal schnell checken, wie sie diese Tools für sich benutzen und jetzt auch gar nicht alles verteufeln und noch bloß keine Angst bekommen, dass sie jetzt irgendwie denken, oh Gott, oh Gott, da kommt jetzt irgendwas Neues, was uns überrennt, sondern ich glaube, es muss jetzt so ein vernünftiger Umgang mit diesen Tools jetzt geschaffen werden und ein Weg und halt auch eine Reglementierung. Wie geht man jetzt damit um? Also, ab wann ist es irgend-wie eine Schöpfung von einem Menschen gewesen? Also, wo setzt man die Grenze? Man muss das, glaube ich, jetzt mal irgendwann mal klar definieren. Ab wann ist es wirklich handmade, wo man ein Tool benutzt hat oder ist es jetzt wirklich komplett KI generiert und deswegen ist es nicht mehr von einem Menschen gemacht. Also ich glaube, das sind jetzt so ein paar Sachen, die man so ein bisschen definieren muss.

**Interviewer:** Das wäre tatsächlich auch eine Frage von mir noch gewesen später, aber dann können wir die auch gerne jetzt reinnehmen. Was denkst du denn? Was müsste man tun? Wie könnte man das angehen, um KI zu reglementieren?

**Collins:** Ja, also es gibt ja schon zum Beispiel Ansätze, dass jetzt Leute versuchen, ihre Stimmen, sage ich mal, zu lizenzieren, wo ich auch glaube, das wird ein neuer spannen-der Markt sein, wo du als Künstler einfach deine Stimme lizenzierst und dann quasi frei gesagt jeder damit machen kann, was er will, aber du kriegst dann Bescheid gesagt - hey, der Werbespot möchte jetzt gerne mit der Stimme, mit dem Sound jetzt irgendwas machen - ist es A okay für dich so und B, das muss halt dann so irgendwie bezahlt werden. Das ist auf jeden Fall mal so ein Punkt. Dann muss halt auch wirklich diese, was ich gerade schon gesagt habe, diese Definition, ab wann ist es, wie gesagt, ist das KI-

Tool als Instrument benutzt worden, dass ich wirklich sage, okay, es war ein Tool und da war noch ein Mensch, der das Ganze ein Stück weit kuratiert hat und gesagt hat, okay, cool, ich habe das jetzt so gepromptet und ich wusste ganz genau, ich muss das so und so und so machen, um dieses Ergebnis zu bekommen und das muss man klar davon abgrenzen zu dem Beispiel, dass halt Leute halt irgendwas machen und das wird dann als neues Werk dann definiert. Und das wird dann Platz Nummer eins und kriegt dann irgendeinen fetten Award oder irgendwie so was. Das muss man definieren. Das kann ich jetzt auch nicht machen, aber das muss irgendwie definiert werden.

**Interviewer:** Fair, das heißt, du würdest da auch ganz klar unterscheiden zwischen rein KI-generierter Musik und KI, ich sage mal, unterstützter Musik?

**Collins:** Genau, muss man. Also es ist ja so ein bisschen ähnlich wie mit dieser Samplerei. Also ab wann ist es wirklich, okay, ich habe was genommen, habe mich davon inspirieren lassen, habe es aber soweit verändert, dass ein neues Werk entstanden ist oder ist es halt wirklich nur gesampelt und Drums drunter und fertig, so. In so einem Bereich befindet man sich jetzt bei der KI auch, wo man jetzt wirklich definieren muss. Ab wann hat da ein Mensch sich wirklich was sich dabei gedacht und wann war es einfach nur Zufallsgenerator und das hat jetzt funktioniert und deswegen ist es jetzt toll, das muss man noch definieren.

**Interviewer:** Ja, okay, ich stimme dir zu. Jetzt, wenn wir schon dann dabei sind, na ja, sagen wir bei der Schöpfungshöhe, na, bei der Schaffungshöhe, wie sieht es denn dann aus mit dem künstlerischen Aspekt dahinter? Wie weit könntest du aber dann sagen, dass die KI dich inspiriert hätte?

**Collins:** Ab wann ich das, also was da passieren muss oder wie man das sagen kann? Naja, nochmal, ich habe es ja gerade schon gesagt, ich glaube am Schluss geht es ja immer noch darum, das Ohr zu haben und entscheiden zu können, okay, ist das jetzt gut oder ist es schlecht? Also, ist es etwas, was ich jetzt für meine Produktion benutzen kann? Fangen wir mal ganz easy an, ich nehme jetzt einfach irgendeine Melodie und lasse mir jetzt von der KI mehrere Melodievorschläge bringen. Ich glaube, man erkennt schon schnell, ob das wirklich nur zusammengeklopft ist, weil es halt musikalisch miteinander funktioniert und deswegen sagt man, hey, man nimmt es oder ob sich jemand was dabei gedacht hat.

Das spürt man und das hört man und ich glaube auch bei gerade etablierten Künstlern, kriegt man da, weiß man glaube ich schon, ob die Jungs, sich da, Jungs und Mädels, sich da einfach nur, ja, wie soll ich sagen, einfach irgendwas rausgespült haben lassen oder sich wirklich was dabei gedacht haben. Also, ich denke, es geht immer noch darum, dass der Mensch, der vor der Maschine steht und entscheidet, hey, das benutze ich oder das benutze ich nicht, dass das immer noch der Key ist und das ist so der springende Punkt.

**Interviewer:** Also immer noch das Kuratieren?

**Collins:** Das Kuratieren, genau. Also, ja, auch vergleichbar, du kannst ja auch eine Studio-Session machen, dir einen Gitarristen holen und sagen, hey, jetzt spielen wir einfach eine halbe Stunde irgendwas ein. Am Schluss muss ich als Producer trotzdem dasitzen und entscheiden, okay, welche Parts waren jetzt nice und welche sind es nicht gewesen. Und so ähnlich sehe ich das jetzt gerade, wenn ich jetzt hier von diesen Melodiegeneratoren, wo ich jetzt dem sage, okay, ich möchte jetzt eine tragende Melodie in dem, in dem Akkord, in dem und dem und das. Das ist ja schon auch ein gewisses Wissen, was du da haben musst, damit du auch wirklich diese Stimmung bekommst, die du haben möchtest. Und ich glaube, das A zu wissen und B zu entscheiden, dass das jetzt wirklich meine Harmoniefolge ist, die ich haben möchte, ja, das ist so der Key.

**Interviewer:** Also immer noch der menschliche Faktor dahinter?

**Collins:** Genau, der entscheidet vor allen Dingen. Also mir geht es gerade nicht mehr darum, du musst es nicht selber machen. Du musst kein Instrument spielen, du musst es nicht machen, aber du brauchst das Ohr. Das war schon immer so. Und wenn man auch so die Entwicklung von jetzt von klassischen Produzenten aus den 60ern zu den Produzenten heute sieht, es ist schon eher so gewesen, dass man jetzt einfach sagen muss, okay, du bist immer weiter weg von diesem handmade gegangen und kamst immer mehr dazu, dass du halt selber einfach entscheiden musst. Weil selbst ein Sample, klar, ein Sample zu flippen kann, irgendwie jeder, aber zu entscheiden, welches Sample jetzt geil ist und funktioniert, da brauchst du immer noch einen Menschen.

**Interviewer:** Okay, ja. Und hast du das dann jetzt auch schon - ich käme da jetzt mal zur Interaktion - hast du das dann jetzt auch schon gemacht mit der Arbeit?

**Collins:** Ja, ja, sehr viel. Also natürlich, klar, im Zuge meiner Masterarbeit habe ich damit ein bisschen experimentiert, aber auch unabhängig davon habe ich für Produktion oft einfach mal versucht, also wie schon angesprochen, diese Stemseparation finde ich bockstark. Gerade beim Sampling, wo du dann wirklich sagst, okay, du hast da irgendeine Phrase, die du haben möchtest, aber jetzt liegen da doch irgendwelche Drums drunter und die nerven mich. Dann klar, dann hole ich mir das über die KI einfach raus und habe dann nur die Melodie, spiel dann natürlich noch eine andere Melodie dazu, eine andere Baseline, hole noch andere Drums drunter und habe gar keinen Schmerz damit, dass es jetzt irgendwie heißt, boah, ich habe jetzt mit KI gearbeitet. Aber auch in anderen Belangen, also jetzt so KI-Suchmaschinen, wo ich jetzt einfach sage, okay, ich suche Samples, gebe dann an, ich möchte ein asiatisches Jazzstück aus dem Jahr 1950 bis 1956 und das würde ich so niemals finden, da bin ich gottfroh darüber, dass ich jetzt so eine KI-Suchmaschine habe, die mir dann einfach kurz YouTube durchjagt und ich dann einfach genau das finde, was ich genau haben will, weil ich schon wie gesagt, ich muss ja trotzdem der KI sagen, was ich will. Und es ist für mich in dem Moment einfach nur eine geile, schnelle, gute Datenbank, wo ich einfach ganz schnell zu meinem Ziel komme. Sonst würde ich dafür eine Woche brauchen, jetzt mache ich es halt in fünf Minuten, das sind so ein paar Ways.

**Interviewer:** Also, überwiegend Workflow.

**Collins:** Workflow Optimierung besser gesagt, Workflow Optimierung.

**Interviewer:** Okay, okay. Und würdest du dadurch sagen, dass das die Qualität zwingend beeinflusst hat in deiner Produktion?

**Collins:** Naja, Qualität ist ja für mich manchmal auch ein bisschen Quantität. Also, ich sage ja immer so, okay, du machst fünf Songs, davon gefallen dir selber drei und irgendwie einen kriegst du platziert. So, jetzt werde ich halt schneller dadurch. Also, ich kann jetzt einfach 20 Songs irgendwie oder 20 Skizzen am Tag mal anfertigen und du musst am Schluss trotzdem immer noch entscheiden, okay, was kann ich jetzt wohin verteilen, wem könnte was gefallen. Und die Qualität bleibt gar nicht unter. Im Gegenteil, ich finde halt einfach, dass ich schneller und mehr machen kann und beim gleichen Qualitätslevel, also kann ich all in all sagen, so es wird besser und schneller.

**Interviewer:** Okay, okay. Du würdest die KI dann, aber jetzt in diesem Fall nicht als zweiten Composer ansehen?

**Collins:** Nee, das funktioniert auch nicht, weil... Also, wie soll ich sagen? Ich meine, klar kannst du dir irgendwie eine Melodie kurz von der KI ausspucken lassen. Easy, aber ein Song ist ja nicht nur eine Melodie. Also es geht ja erstmal darum, welchen Sound nämlich dazu. Dann wie bearbeite ich diesen Sound, wie effektiere ich diesen Sound, was habe ich für Anschlagdynamik in dem Sound? Deswegen, mit der Melodie alleine ist es noch nicht getan. Nein, ich finde dann auch, wie soll ich sagen, vielleicht benutze ich es tatsächlich als Composer, aber trotzdem glaube ich, sind dann noch so viele Schritte mehr, als jetzt nur dieses klassische Komponieren von Melodien, dass ich das jetzt gar nicht so schlimm finde. Also die Editierung und die Effektierung, die wirklich den Sound und das Klangbild ausmacht auch keine KI, das muss immer noch ich selber machen. Und deswegen schließt das eine das andere gar nicht aus, wenn du verstehst, was ich meine, ja?.

**Interviewer:** Ja, ja. Okay, und wenn die KI jetzt noch stärker werden würde, hättest du für dich dann eine Grenze, wo du sagst, das würde ich gar nicht mehr benutzen? Gäbe es das? Also, wie das jetzt klingt momentan, nutzt du es auf jeden Fall als Tool oder wie ein Synthesizer oder sonst was? Und wenn das jetzt darüber hinausgehen würde, sagen wir mal, die KI wäre in der Lage, dich auch noch zu unterstützen beim kompletten Mixing Process, bei allem automatisch, würdest du das in Anspruch nehmen?

**Collins:** Ich kann mir das gar nicht vorstellen, dass die KI mir so ein Ergebnis rausspuckt, wo ich wirklich sage, okay, geil. Also ich kann mir das bei bestem Willen nicht vorstellen.

Also dieses Mixing, es ist ja auch so eine komplette subjektive Sache, wann ist ein Mix geil, da gibt es kein richtig und falsch. Deswegen kann ich mir das gar nicht vorstellen. Ich meine, ich weiß auch auch, worauf du vielleicht hinaus willst, wenn ich jetzt wirklich der KI prompten könnte, mach mir einen Kanye West Beat, der genau so klingt wie this and this and this, und das Ding würde das genauso ausspucken. Also wirklich klingt krass und super und toll. Naja, selbst wenn es das dann könnte, dann würde ich der KI halt prompten, okay, dann mach mir ein Sample, das wie this and this and this klingt, und würde das dann halt weiterverarbeiten. Also, sagen wir es mal so, ich würde glaube

ich nie ein Ergebnis als solches für sich stehen lassen. Also ich fände es irgendwie spannend, wenn man das so machen könnte, dass ich sagen könnte, hey, mach mir this, das, exakt, das funktioniert ja jetzt noch gar nicht.

Dass du wirklich mit 3-4 Begriffen den promptest und der spuckt dir dann ein Stück raus, das funktioniert ja gar nicht. Aber wenn es das könnte, dann würde ich damit halt weiterarbeiten, weißt du? Wenn ich sage, ich will jetzt einen fetten Bläsersatz haben, und die KI macht das, ist okay, geil. Aber dann würde ich halt den Bläsersatz nochmal choppen und machen und tun und würde mir damit halt das Geld sparen, jetzt in ein Studio zu gehen und dann eine große Aufnahme dafür machen zu können. Deswegen, ich kann mir gar nicht vorstellen, dass es diesen Punkt gibt, wo ich dann sage, nee, das mache ich jetzt aber nicht.

**Interviewer:** Okay, ja, spannend zu hören, aber gut, dann hätte ich sonst, wenn wir schon so ein bisschen dabei sind, Thema Ethik, die Frage, glaubst du, also Musik, sagen wir mal so, Musik ist ja auch Kultur, definitiv, glaubst du, dass die KI-generierte Musik jetzt auch Kultur so abbilden kann, wie das Menschen könnten?

**Collins:** Absolut nicht. Also, gerade heutzutage lebt ja die Musik ja nicht nur von dem Song, das ist Drumherum, du musst als Künstler eine Marke sein, Social Media muss für dich funktionieren, und da, wie soll ich sagen, das fließt ja dann mit in den Song ein, also zum Beispiel Shakira hat irgendwann zur WM Waka Waka gemacht, war alles nett und so, aber es ist kein Vergleich, wenn ihr jetzt ein Burner Boy einen Afro-Song macht, weil das sind zwei verschiedene Personen, sie kommt aus Südamerika, die kann das nicht so rüberbringen, wie irgendein Dude, der halt wirklich aus Nigeria kommt und diese Kultur wirklich in sich trägt, und deswegen mit einem Song alleine, das geht nicht. Gerade heutzutage, das funktioniert nicht, dass du jetzt irgendwie sagst, ja, okay, da gibt es jetzt ein Song und ohne die Person dahinter funktioniert der Song auch gar nicht. Also davon bin ich komplett überzeugt, dass das jetzt unabhängig von KI, das geht nicht, das geht einfach nicht.

**Interviewer:** Okay, also nur KI alleine glaubst du nicht, auch dass es keine kulturelle Identität für sich selber aufbauen könnte?

**Collins:** Das ist jetzt wieder was anderes, also das ist eine KI-Kultur? Safe, also wenn ich das jetzt so auf das, was so mit diesen Bildern gerade passiert, so diese KI-generierten Bilder, das ist ja jetzt auch schon so eine eigene Kunst für sich geworden, ja? Genau, so eine eigene Ästhetik. Also ich kann mir jetzt nicht vorstellen, wie dieser KI-Kultur-Sound klingen soll, aber wenn ich das jetzt übertrage, was das mit diesen Bildern ist, so da kann ich mir das schon vorstellen, das ist eine eigene Kultur, aber das ist eine bestehende Kultur irgendwie kopiert und man ist dann irgendwie – nein. Aber eine eigene Kultur, da gebe ich dir Recht, das kann passieren, ja.

Okay, cool. Ja, dann sind wir ja auch schon bei der Zukunft eigentlich. Was siehst du dann noch so für Auswirkungen auf die Musikproduktion jetzt mal aus deiner Sicht? Was glaubst du, wird sich verändern, wenn KI immer stärker wird, die Tools werden mächtiger?

**Collins:** Ja, es wird halt zugänglicher, also man sieht ja schon in den letzten 10, 15 Jahren so jeder macht Musik gefühlt und jeder kann es ganz einfach machen und das wird stärker zunehmen mit KI, das ist einfach noch zugänglicher wird und dass jeder jetzt irgendwie Musik machen kann. Die Qualität der Musik wird natürlich auch drunter leiden, das sehe ich auf jeden Fall kommen. Die wird auf jeden Fall drunter leiden, aber andersrum sage ich auch, dass die Sachen, die aber gut sind, die werden dann schon besser, also das ist davon, wenn ich überzeugt, dass das was gut ist, ist gut, aber es wird einen ganz großen Brei an Scheiße geben, die eigentlich kein Mensch braucht, aber es trotzdem halt irgendwie auf den Markt gespielt wird, das glaube ich.

**Interviewer:** Wird ja jetzt schon zum Teil, ich weiß nicht, ob du das mitbekommen hast auf Spotify mit KI-Musik, die da hochgeladen wird. Das merkt man, also ich merke es jetzt schon

**Collins:** Okay krass, ich habe es jetzt noch nicht so gemerkt, also ich weiß es nicht, aber gucke auch gezielt nach meinen Künstlern, deswegen keine Ahnung, aber ja, ist eine logische Schlussfolgerung, das glaube ich, dass das passieren wird.

**Interviewer:** Okay, und du siehst das, siehst du das alles als positiv oder negativ? Gemischt?

**Collins:** It is what it is, also, wenn man irgendwie in der Musikbranche, sage ich jetzt mal ganz blöd, tätig ist, das ist part of the game. Das war immer im Wandel, es ist auch gut so, dass man jedes Mal vor neue Herausforderungen gestellt wird und also, ich sehe es wertfrei, also es ist wie es ist, deal with it, so. So sehe ich das.

**Interviewer:** Okay, ja. Man muss damit umgehen jetzt.

**Collins:** Man muss damit umgehen und das akzeptieren, dass es so ist und am Ball bleiben und gucken und halt einfach die Chancen darin sehen. Also, dieses generelle Bashing finde ich nicht cool. Ich finde es irgendwie nice, dass du damit Leuten einen Zugang zu Musik schaffen kannst, die es so vielleicht nie hinbekommen hätten. Und wenn die gut sind, dann gehen die ihren weg und es liegt dann nicht nur daran, dass sie jetzt mit KI irgendwas gemacht haben, vielleicht war das halt so, die erste Hürde wurde damit genommen, aber du wirst ganz schnell an den Grenzen stoßen und weißt dann selber, nee, das ist doch scheiße, wenn du gut bist, dann wirst du selber anfangen zu frickeln und zu machen und zu tun und deswegen, das ist okay, das ist gut.

**Interviewer:** Okay, das ist klar. Und für die rein KI-generierte Musik siehst du da eine Zukunft? Irgendwo, was die vielleicht ablösen könnte?

**Collins:** Ja, ich glaube halt in diesem Werbezeug sind so Industrie-Porn-Scheiß, wo du halt auf irgendwelchen Messen, irgendwelche Kolben siehst und da brauchst du irgendein Musikbett drunter. Okay, da schon, aber das ist jetzt keine Musik, wo ich jetzt irgendwie sage, die nehme ernst, das ist jetzt keine, also wenn ich von Musik rede, da rede ich halt von Künstlern, die was vertreten und da ist eine Subkultur dahinter und so und in diesem Bereich wird KI niemals eingreifen können.

**Interviewer:** Also keine aktive Musik?

**Collins:** Nee, genau, aber so diese Beschallungszeugs, was halt so ein bisschen nebenbei läuft so und das ist aber dann auch Musik für Konsumenten, wo ich jetzt auch sage, denen ist Musik auch nicht wichtig. Aber ich glaube so wirklich, wo es wirklich drum geht, okay, hey, wir machen jetzt richtig gute Musik. Was war die Frage?

**Interviewer:** Ob - was die Zukunft wäre für rein KI-generierte Musik.

**Collins:** Für KI, ah ja genau, in dem Bereich da sehe ich schon so, in dieser Massenabfertigung, so da kann ich mir das schon vorstellen, dass es da irgendwie seinen Platz finden wird.

**Interviewer:** Das heißt, weil der reinen KI-generierten Musik quasi die Authentizität fehlt?

**Collins:** Genau, und du brauchst es ja da auch gar nicht, also ja, das ist da ja scheißegal, da geht es halt wirklich nur um die Musik und dass du da irgendein Musikbett hast und natürlich ist es für so ganze Werbeagenturen natürlich schon einfacher, wenn sie da halt einfach kurz ihre Prompts reinhauen können. Ja, getragenes Gitarrenzeug, irgendwas und so und das in drei Minuten bitte rausspucken, verstehe ich dann auch, dass sie eher das nehmen, als dass sie halt einen Produzenten irgendwie 3-4 .000 Euro zahlen, dass der da jetzt komplett was für die produziert so, verstehe ich schon auch. Also, es ging ja schon damals los, als du dieses Stockzeugs losging, da dachte ich schon so, ah, scheiße und so, aber ja, gut, macht schon Sinn, weil die, ich verstehe das, dass für die ist Musik was anderes, es muss funktionieren, fertig so. Da muss das jetzt nicht irgendwie Soul haben, im Sinne von, hey, da hat sich jemand was bei gedacht und dramaturgisch und bla bla bla. Deswegen, um es kurz zu machen, ja, in der Werbung, Industriemessezeug, da kann ich mir schon vorstellen, dass KI-generierte Musik auf jeden Fall stattfindet.

**Interviewer:** Okay, dann hätte ich zur Zukunft wahrscheinlich nur noch eine Frage, was denkst du, worauf sollte man ein Auge haben für die Zukunft, was wird das nächste große KI-Musikding?

**Collins:** Boah, wenn ich das – was sowas angeht, bin ich echt schlecht, ich dachte auch damals, dass sich DVD nicht durchsetzen würde. Deswegen, was solche Zukunftsvisionen, das nächste KI-Ding, angeht... also ich glaube an diese Stimmenlizenzen, was ich schon einmal am Anfang erzählt habe, das finde ich irgendwie, das könnte funktionieren, dass ich als Künstler sage, okay, ich tue meine Stimme lizenzieren und ich muss eigentlich nichts mehr machen, sitze daheim und das wird halt alles KI-generiert und dann packe ich das halt da drauf, vielleicht dann auch, hey, auch so Modelle, die dann auch wirklich nur exklusiv für solche Geschichten benutzt werden, danach wird das auch nie wieder mehr benutzt und so. Dann kannst du halt auch easy Cash kurz nebenbei machen. Das könnte ich mir vorstellen, diese Stimmenlizenzen. Das, ja, aber ansonsten, keine Ahnung.

**Interviewer:** Okay, okay, gibt es denn dann etwas, jetzt persönlich gefragt, was du dir wünschen würdest, vielleicht was KI könnte? Jetzt zum Beispiel sei es noch weitere Workflow-Optimierung? Was würdest du dir wünschen, was die KI für dich übernehmen könnte?

**Collins:** Also dieses Ding, das ich wirklich mal, dieses Prompten und der mir wirklich das ausspuckt, was ich will, in einer Bombenqualität, also wirklich, dass ich wirklich sage, okay, ich möchte jetzt Streicher in A-Moll, dies-das soll klingen, als wurde es dort in der Kathedrale aufgezeichnet und so, in der Länge und dies und das, das wäre geil.

**Interviewer:** Also die echte Prompttreue?

**Collins:** Genau, die echte Prompttreue, ja, genau, guter Begriff. Genau, das fände ich spannend. Ja, das fände ich cool, das würde ich mir wünschen.

**Interviewer:** Okay, gut, ich glaube, die anderen Fragen haben wir zwischendurch schon beantwortet.

**Collins:** Ja, easy. Ich hoffe, es war irgendwie okay?

**Interviewer:** Ja klar, danke, war super. Danke dir!

**Collins:** Cool!