

HOCHSCHULE  
DER MEDIEN

---

Fakultät Electronic Media

Bachelorthesis

im Studiengang Audiovisuelle Medien

**Die KI-Stimme – Nutzung künstlicher Intelligenz  
in der Musikproduktion von Vocals**

vorgelegt von:

Philipp Sonntag

zur Erlangung des akademischen Grades

Bachelor of Engineering

Ausgabedatum: 03.05.2024

Abgabedatum: 02.08.2024

Erstgutachter: Prof. Oliver Curdt

Zweitgutachter: Prof. Dr. Andreas Koch

## Ehrenwörtliche Erklärung

Hiermit versichere ich, Philipp Sonntag, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Die KI-Stimme – Nutzung künstlicher Intelligenz in der Musikproduktion von Vocals“ 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 (§ 24 Abs. 2 Bachelor-SPO, § 23 Abs. 2 Master-SPO (Vollzeit)) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

Unterschrift:

A handwritten signature in black ink, appearing to read 'Sonntag', written in a cursive style.

## Zusammenfassung

Künstliche Intelligenz (KI) in der Musikproduktion ist ein aktuelles und kontrovers diskutiertes Thema. Bekannte Künstler wie Billie Eilish haben Bedenken geäußert, dass KI ihre berufliche Existenz gefährden könnte, da sie Musik erzeugen kann, die täuschend echt, wie die Werke anderer Musiker wirke. Trotz dieser Bedenken bietet KI bedeutende Vorteile für die Musikproduktion, insbesondere im Bereich der Gesangsproduktion.

Diese Arbeit gibt einen Überblick über die Nutzung von KI in der Musikproduktion bzw. im speziellen der Gesangsproduktion und analysiert verschiedene KI-Tools und -Technologien. Der Fokus liegt auf der Integration dieser Technologien in den kreativen Prozess und den neuen Möglichkeiten, die sie eröffnen.

Der Aufbau der Arbeit umfasst theoretische Grundlagen zur Stimme und Musikproduktion, zu KI und deren Geschichte in der Musik und eine detaillierte Betrachtung aktueller KI-Tools. Anschließend werden die sich daraus ergebenden Chancen und Herausforderungen betrachtet.

Im Fazit wird festgestellt, dass KI vielfältige neue kreative Möglichkeiten bietet, von der Text-Ideengenerierung bis zur vollständigen Songproduktion. Gleichzeitig betont die Arbeit die Notwendigkeit klarer Regelungen und ethischer Überlegungen, um die menschliche Kreativität zu schützen. Es bleibt eine gemeinsame Aufgabe von Entwicklern, Künstlern und der Musikindustrie, diese Technologien innovativ und verantwortungsvoll zu nutzen.

## **Abstract**

Artificial intelligence (AI) in music production is a current and controversially discussed topic. Renowned artists like Billie Eilish have expressed concerns that AI could threaten their professional existence, as it can generate music that deceptively resembles the works of other musicians. Despite these concerns, AI offers significant advantages for music production, especially in vocal production.

This paper provides an overview of the use of AI in music production, specifically in vocal production, and analyzes various AI tools and technologies. The focus is on integrating these technologies into the creative process and the new possibilities they provide.

The structure of the paper includes theoretical foundations of voice and music production, AI and its history in music, and a detailed examination of current AI tools. Subsequently, the resulting opportunities and challenges are considered.

The conclusion finds that AI offers a variety of new creative possibilities, from generating text ideas to full song production. At the same time, the paper emphasizes the need for clear regulations and ethical considerations to protect human creativity. It remains a joint task for developers, artists, and the music industry to use these technologies innovatively and responsibly.

# Inhaltsverzeichnis

<b>Ehrenwörtliche Erklärung .....</b>	<b>I</b>
<b>Zusammenfassung .....</b>	<b>II</b>
<b>Abstract .....</b>	<b>III</b>
<b>Inhaltsverzeichnis .....</b>	<b>IV</b>
<b>Abkürzungsverzeichnis.....</b>	<b>VI</b>
<b>Abbildungsverzeichnis.....</b>	<b>VII</b>
<b>Tabellenverzeichnis.....</b>	<b>VIII</b>
<b>1 Einleitung.....</b>	<b>1</b>
1.1 <i>Problemstellung .....</i>	1
1.2 <i>Ziele .....</i>	1
1.3 <i>Aufbau der Arbeit .....</i>	2
<b>2 Grundlagen der Vocal-Produktion .....</b>	<b>3</b>
2.1 <i>Die Stimme und deren Rolle in der Musik.....</i>	3
2.2 <i>Faktoren zur Bewertung von Gesang .....</i>	5
2.3 <i>Traditioneller Musik-/ Vokalproduktionprozess .....</i>	5
2.3.1 <i>Songwriting und Pre-Production .....</i>	5
2.3.1.1 <i>Herangehensweisen .....</i>	6
2.3.1.2 <i>Pre-Production .....</i>	7
2.3.1.3 <i>Analyse der eigenen Herangehensweise: .....</i>	7
2.3.2 <i>Produktion von Gesang .....</i>	11
2.3.2.1 <i>Recording.....</i>	11
2.3.2.2 <i>Editing .....</i>	12
<b>2.3.2.2.1 Comping .....</b>	12
<b>2.3.2.2.2 Tonhöhenkorrektur .....</b>	12
<b>2.3.2.2.3 Audiorestauration .....</b>	13
2.3.2.3 <i>Mixing und Mastering.....</i>	13
<b>3 Künstliche Intelligenz und Musik .....</b>	<b>15</b>
3.1 <i>Allgemeine begriffliche und theoretische Grundlagen zu KI .....</i>	15
3.1.1 <i>Machine Learning .....</i>	16
3.1.2 <i>Deep Learning und neuronale Netze .....</i>	17
3.1.3 <i>Starke und schwache KI .....</i>	19

3.2	<i>Geschichte von KI in der Musik</i> .....	19
<b>4</b>	<b>Analyse verschiedener KI-Applikationen in Bezug auf den Gesamtproduktionsprozess</b> .....	<b>22</b>
4.1	<i>ChatGPT</i> .....	22
4.1.1	Prompting.....	23
4.1.2	Songtexterstellung mit ChatGPT .....	25
4.1.3	Texterstellung zu einer Melodie .....	30
4.1.4	Zusammenfassung ChatGPT: .....	34
4.2	<i>KI-Musikgeneratoren</i> .....	35
4.2.1	Udio.....	36
4.2.2	Suno.....	44
4.2.3	Eleven Labs .....	45
4.2.4	Fazit Musikgeneratoren .....	47
4.3	<i>Singing Voice Synthesis mit ACE Studio</i> .....	47
4.3.1	Analyse ACE Studio.....	48
4.3.2	Voice Cloning mit ACE Studio .....	51
4.3.3	Fazit ACE Studio.....	53
4.4	<i>Audimee</i> .....	53
4.4.1	Nutzung von Audimee.....	54
4.4.2	Stem Separation .....	57
4.4.3	Fazit Audimee und Stem Separation .....	59
<b>5</b>	<b>Chancen und Herausforderungen mit KI in der Produktion von Vocals</b> .	<b>60</b>
5.1	<i>Chancen</i> .....	60
5.2	<i>Herausforderungen</i> .....	61
<b>6</b>	<b>Ausblick und Fazit</b> .....	<b>66</b>
	<b>Literaturverzeichnis</b> .....	<b>67</b>
	<b>Anhangsverzeichnis</b> .....	<b>72</b>

## Abkürzungsverzeichnis

ca.	circa
DAW	Digital Audio Workstation
d. h.	das heißt
ebd.	Ebenda
Engl.	Englisch
etc.	et cetera
ISMIR	International Society for Music Information Retrieval
KI	künstliche Intelligenz
o.Ä.	oder Ähnliches
RVC	Retrieval-Based Voice Conversion
SVS	Synthetic Voice Synthesis
usw.	und so weiter
z. B.	zum Beispiel

## Abbildungsverzeichnis

Abbildung 1: Melodie Slave or King .....	8
Abbildung 2: Aufbau eines einfachen neuronalen Netzes .....	18
Abbildung 3: Typische Mensch – Maschinen Interaktion bei algorithmischen Kompositionen bei denen der Komponist gleichzeitig Nutzer oder Programmierer eines Programms ist.....	20
Abbildung 4: Startseite ChatX.....	24
Abbildung 5: ChatGPT Texterzeugung .....	31
Abbildung 6: Melodie mit Originaltext .....	33
Abbildung 7: Bild Diffusion .....	35
Abbildung 8: Audio Diffusion.....	36
Abbildung 9: Startseite Udio .....	37
Abbildung 10: Songs im Reiter „My Creations“ .....	39
Abbildung 11: Piano Roll in ACE Studio .....	48
Abbildung 12: GUI ACE Studio .....	49
Abbildung 13: Sängerbeschreibung Audimee.....	54
Abbildung 14: Sängerauswahl Audimee .....	55
Abbildung 15: GUI Audimee .....	55
Abbildung 16: Anzeige bei Prompting mit Künstlernamen.....	62
Abbildung 17: Anzeige bei Prompting mit Künstlernamen.....	64



## **Tabellenverzeichnis**

Tabelle 1: Ranking der wichtigsten Faktoren für Zuhörer.....	4
Tabelle 2: Ranking der wichtigsten gesangsspezifischen Faktoren für Zuhörer.....	5
Tabelle 3: Veränderbare Parameter im Creation Mode von Udio.....	37

# 1 Einleitung

## 1.1 Problemstellung

*„AI as a tool in music-making is fine, but it's always going to be the humanity in music that makes people want to listen to it.”*

*- Jacob Collier*

Erst diesen April wurde ein offener Brief verfasst, der an die Techunternehmen und KI-Firmen gerichtet ist. Fast 250 Künstler, darunter Bekanntheiten wie Billie Eilish, Stevie Wonder oder Katy Perry unterschrieben diesen Brief und sprachen sich so gegen die Verwendung von Künstlicher Intelligenz (KI) in der Musikindustrie aus. Der unverantwortliche Umgang mit KI stelle eine Gefahr für die Musiker und ihre Arbeit dar. KI ist in der Lage Stücke zu kreieren, die klingen, als wären Sie von Kylie Minogue oder ABBA, obwohl diese Künstler nie an diesen Stücken beteiligt waren.<sup>1</sup>

Diese sogenannten Deepfakes verbreiten sich über die ganze Social Media Welt und es scheint so, als werden sie von den Fans der echten Künstler tatsächlich als eher positiv aufgenommen.<sup>2</sup>

Fakt ist, dass Künstliche Intelligenz die Grenzen des Möglichen erweitert und eine immer größer werdende Rolle in der Produktion von Gesang und Musik spielt.<sup>3</sup> Fraglich ist, wie man diese Technologie verantwortungsvoll nutzen kann und welche Vorteile sie für Sänger oder Producer bieten könnte.

## 1.2 Ziele

Das Ziel dieser Arbeit ist es, dem Leser einen Einblick in die Nutzung von KI in der Musikproduktion bzw. im speziellen der Gesangsproduktion zu verschaffen. Besonders angehenden Producern und Musikschaaffenden soll diese Arbeit als Inspiration dienen und Wege zeigen, wie KI in ihren eigenen Schaffensprozess miteingebunden werden kann und was damit möglich ist.

Hierzu werden unterschiedliche Möglichkeiten und Tools betrachtet, die auf dem Markt zur Verfügung stehen und genauer analysiert. Der Leser soll dazu angeregt sein, seinen eigenen Workflow zu überdenken und mögliche KI-gestützt Alternativen in Betracht zu ziehen oder auch zu verwerfen.

---

<sup>1</sup> Vgl. Jörg, D. (2024).

<sup>2</sup> Vgl. Danesi, M. (2024).

<sup>3</sup> Vgl. Beshko (2023).

### **1.3 Aufbau der Arbeit**

Im ersten Kapitel wird auf die theoretischen Grundlagen zum Thema Stimme und Musikproduktion eingegangen. Es wird die Frage geklärt, was einem Zuhörer bei Musik wichtig ist und welche Rolle die Stimme in diesem Kontext spielt. Danach wird der traditionelle Musikproduktionsprozess, im speziellen die Vokalproduktion elaboriert, um ein Verständnis davon zu erlangen, welche Aufgaben bei der Erschaffung von Musik existieren. Um dem Leser eine Idee zu geben, wie dies in der Praxis genau aussieht, wird die Herangehensweise eines eigenkomponierten Stückes detaillierter beleuchtet und analysiert.

Im zweiten Teil werden dem Leser allgemeine Grundlagen bezüglich Künstlicher Intelligenz nahegebracht. Nach einer Definition des Begriffs und einiger damit in Zusammenhang stehenden Schlüsselwörter wird auf die Geschichte von KI in der Musik eingegangen.

Anschließend werden verschiedene KI-Tools betrachtet, die aktuell verfügbar sind und für die Vokal-Produktion nützlich sein könnten. Neben dem KI-gestützten Schreiben eines Songtextes werden auch andere Möglichkeiten wie zum Beispiel die komplette Stimersetzung getestet. Die dabei entstandenen Audioclips werden dem Leser per Google-Drive-Link zur Verfügung gestellt.

Im letzten Teil dieser Arbeit wird auf die Chancen und Herausforderungen bezüglich der Nutzung von KI in der Musikproduktion eingegangen.

Die gewonnenen Erkenntnisse werden nachfolgend in einem Fazit zusammengefasst.

## 2 Grundlagen der Vocal-Produktion

### 2.1 Die Stimme und deren Rolle in der Musik

Die menschliche Stimme ist das Resultat eines vielschichtigen Zusammenspiels der Artikulationswerkzeuge im Vokaltrakt, dem Kehlkopf und der Atmung. Sie ist ein zentrales Kommunikationsmittel und wird auf verschiedenste Weisen gebraucht. Neben dem Sprechen und Singen können auch affektbasierte Laute wie schreien, lachen oder weinen damit erzeugt werden.<sup>4</sup> Jeder Mensch ist im Prinzip dazu fähig zu singen und viele tun dies auch in verschiedensten Situationen: zu Geburtstagsfeiern, bei Gottesdiensten oder im Fußballstadion. Sobald die Stimme jedoch in einem professionelleren Rahmen verwendet wird, bedarf es meistens viel Zeit und Übung. Opern, anspruchsvolle Kunstlieder und ähnliches verlangen der Stimme viel ab. Die Stimme muss über eine längere Zeit belastbar sein. Ebenso ist es erforderlich, dass sie schnelle Tonabfolgen und dynamische Abstufungen beherrscht, sowie über einen großen Tonumfang verfügt.<sup>5</sup>

Um herauszufinden, wie wichtig die Stimme in der Musik ist und auf was genau ein Publikum achtet, wird eine Studie der International Society for Music Information Retrieval (ISMIR) betrachtet. Bei der Studie wurden 50.000 potenzielle Teilnehmer in Kanada und den USA angeschrieben. Von diesen füllten 626 Teilnehmer die Onlineumfrage komplett aus, die unter anderem folgende Fragen enthielt:<sup>6</sup>

*Q1: When you listen to music, what things about the music do you notice? Please list as many as you can.*

*Q2: Please rank how important the aspects you listed are to your musical preference, where 1 is the most important.*

Die Antworten dieser Fragen wurden anschließend von den Forschern in semantische Gruppen kategorisiert. Unter „Songtext“ wurden beispielsweise alle Antworten gezählt, die die Präsenz des Songtextes, die Verständlichkeit, die „Message“, den lyrischen Inhalt o.Ä. betrafen. Eine zweite Testgruppe von 531 Teilnehmern sollte anschließend diese Kategorien nun nach Wichtigkeit ordnen.<sup>7</sup>

---

<sup>4</sup> Vgl. Richter, B. (2016). S. 584

<sup>5</sup> Vgl. Seedorf, T. (2023). S. 75

<sup>6</sup> Vgl. A.; Jansson, A.; Kumar, A.; & Bittner, R. M. (2018). S. 516ff.

<sup>7</sup> Ebd.

Die folgende Tabelle zeigt die Ergebnisse der Studie. Eine höhere Borda Score bedeutet hier im Prinzip, dass mehr Teilnehmer entsprechende Kategorie in ihrem persönlichen Ranking hoch eingestuft haben. Der P-Wert gibt Aufschluss zur statistischen Relevanz. Ein P-Wert von 1.0 bedeutet hierbei, dass das Ergebnis statistisch gesehen nicht relevant ist und es keinen Unterschied gibt zu einer zufälligen Anordnung. Anzumerken ist bei der folgenden Tabelle, dass einige den Rang betreffend statistisch unrelevante Kategorien, wie Melodie, Sound, spezifischer Künstler, Genre, Tempo weggelassen wurden.<sup>8</sup>

**Tabelle 1: Ranking der wichtigsten Faktoren für Zuhörer**

<b>Semantische Kategorie</b>	<b>Beschreibung</b>	<b>Borda Score</b>	<b>P-Wert</b>
Emotion/ Gefühlslage	Was fühlt der Zuhörer? – Emotion/ Gefühlslage	4641	<0.001
Stimme	Stimme/Gesang	3688	<0.001
Songtext	Songtext	3656	<0.001
Beat/Rhythmus	Beat/Rhythmus	3460	<0.001
Struktur/ Komplexität	Wie es komponiert ist, der Hook, die Struktur?	2677	1.0
Musikalität	Das Können der Musiker, Musikalität	2583	1.0
...	...	...	...

Quelle: Eigene Darstellung nach Demetriou, A.; Jansson, A.; Kumar, A.; & Bittner, R. M. (2018). S. 518

Die Studie zeigt, dass es den Teilnehmern bei einem Song am wichtigsten ist, dass er überhaupt Emotionen hervorruft. Direkt danach wird die Stimme und der Songtext aufgelistet, gefolgt vom Beat/ Rhythmus.<sup>9</sup>

In der Musik werden Sänger oft als das Herz und die Seele eines Songs angesehen. Wie kein anderes Instrument ist die Stimme des Menschen dazu in der Lage ein Publikum zu fesseln, Emotionen auszulösen und es zu inspirieren.<sup>10</sup>

Folglich besteht zwischen dem wichtigsten Faktor in der Musik (Emotion) und der Stimme zweifelsohne ein Zusammenhang.

<sup>8</sup> Vgl. A.; Jansson, A.; Kumar, A.; & Bittner, R. M. (2018). S. 517f.

<sup>9</sup> Vgl. A.; Jansson, A.; Kumar, A.; & Bittner, R. M. (2018). S. 518

<sup>10</sup> Vgl. Bricio, C. (2023).

## 2.2 Faktoren zur Bewertung von Gesang

Nachdem die Teilnehmer die Kategorien geordnet hatten, wurde das Ganze vokalspezifisch wiederholt.<sup>11</sup>

**Tabelle 2: Ranking der wichtigsten gesangsspezifischen Faktoren für Zuhörer**

Vokale semantische Kategorie	Borda Score	P-Wert
Singing Skill (Können, Fähigkeit)	3423	<0.001
Wie gut die Stimme zum Rest der Musik passt	3380	<0.001
Lyrisches Können, Cleverness	3145	<0.001
Die „Message“, die Bedeutung des Textes	3038	0.048
Authentizität	2884	<0.001
Einzigartigkeit	2780	<0.001
Gefühl in der Stimme/ Emotionalität	2771	0.006
....	...	1.0

Quelle: Eigene Darstellung nach Demetriou, A.; Jansson, A.; Kumar, A.; & Bittner, R. M. (2018). S. 519

Einige weitere nennenswerte Faktoren, die zwar einen P-Wert von 1.0 haben, aber dennoch während der ersten Umfrage genannt wurden und damit für eine Gesangsbewertung nicht irrelevant sind, waren Intensität/Stärke, Natürlichkeit, melodisch klingend, Verständlichkeit, nasal klingend oder Nachempfindbarkeit.<sup>12</sup>

## 2.3 Traditioneller Musik-/ Vokalproduktionprozess

Im folgenden Abschnitt wird auf den üblichen Prozess einer Produktion eines Songs eingegangen, um eine Übersicht davon zu erhalten, an welchen Stellen eine KI bezüglich des Gesangs nützlich sein könnte.

### 2.3.1 Songwriting und Pre-Production

Das Songwriting und die Musikkomposition stehen an erster Stelle. Dabei gibt es unterschiedliche Methoden, wie man einen Song schreiben kann. Einige Songwriter beginnen mit einer Melodie, andere mit einem Text, wiederum andere starten mit dem Instrumental

---

<sup>11</sup> Vgl. A.; Jansson, A.; Kumar, A.; & Bittner, R. M. (2018). S. 517

<sup>12</sup> Vgl. A.; Jansson, A.; Kumar, A.; & Bittner, R. M. (2018). S. 519

und schreiben später erst einen Text dazu.<sup>13</sup> Kurz gesagt, am Anfang braucht man irgendeine Art von Idee, auf der man aufbauen kann.<sup>14</sup>

### 2.3.1.1 Herangehensweisen

Die Idee zu einem Song kann auf unterschiedliche Weise entstehen.

Durch bewusste Ablenkung wie Bücher oder Filme oder durch Gespräche mit Freunden kommt man auf Themen, auf die man allein nicht gekommen wäre.<sup>15</sup>

Paul McCartney behauptet, dass er die Melodie von Yesterday ursprünglich geträumt hätte und sich sicher gewesen wäre, dass sie von jemand anderem stammt. Erst als er sie anderen Leuten vorspielte und diese ihm versicherten, dass sie die Melodie nicht kannten, stand fest, dass sie nur in McCartneys Kopf existierte.<sup>16</sup>

In anderen Liedern werden wiederum persönliche Probleme und Erfahrungen verarbeitet. So sagt Matt Tuck von der Metalband Bullet for my Valentine über das Album Gravity: "As clichéd as it probably sounds, this is my break-up record. (...) I've gone through a really hard time behind the scenes the last couple of years with the breakdown of my marriage, and all the turmoil and emotional baggage that comes with that. (...) I slipped into depression – nothing too heavy, thankfully – and April came around and I started to try and make music. So me being me, I wanted to do something different."<sup>17</sup>

Wie Matt Tuck reflektiert auch der Weltstar Ed Sheeran seine persönlichen Erfahrungen in seinen Songtexten und bleibt so sich selbst treu. Diese emotionale Ehrlichkeit sorgt dafür, dass man sich als Zuhörer sehr gut mit seinen Texten identifizieren kann und Ed Sheeran selbst ausgesprochen authentisch wirkt, was wiederum ein Schlüsselement seines Erfolgs ist.<sup>18</sup>

Sheeran selbst macht beim Songwriting Gebrauch vom Konzept der sogenannten Wasserhahn-Theorie. Diese besagt, dass man analog zu einem Wasserhahn, der lange nicht benutzt wurde, am Anfang eines kreativen Prozesses sehr wahrscheinlich ein schlechtes Ergebnis bekommt. Lässt man den Wasserhahn jedoch laufen und seine Ideen immer weiter sprudeln, wird das Wasser irgendwann klar, bis zu dem Punkt, wo der Output seinen Standards entspricht. Disziplin und regelmäßiges Üben helfen seine Kreativität aufrechtzuerhalten. So ist es ihm möglich einen Song pro Tag zu schreiben.<sup>19</sup>

---

<sup>13</sup> Vgl. O.V. peak-studios.de (2024).

<sup>14</sup> Vgl. Dumont, L. (o.D.).

<sup>15</sup> Vgl. Fresh, D. (2017).

<sup>16</sup> Vgl. Kurz, P. (2021).

<sup>17</sup> Goodman, E. (2018).

<sup>18</sup> Vgl. Butler, G. (2023).

<sup>19</sup> Ebd.

Andere Hilfsmittel können Stift und Papier sein, um in Alltagssituationen seine Ideen festhalten zu können und diese nicht zu vergessen. Das Aufschreiben von Gegensätzen oder Mindmaps können durch Assoziationsbildung zu einem Schlüsselwort ebenfalls hilfreich sein, um neue Ideen zu entwickeln.<sup>20</sup>

### 2.3.1.2 Pre-Production

Ist die Grundidee vorhanden, werden neue Songparts geschrieben, ausprobiert, eventuell ausgetauscht, bis die grobe Struktur steht. Die ungefähre Länge des Songs ist nun klar und es kann an die Feinarbeit und das Arrangement gehen. Melodien, Akkorde und Textpassagen werden angepasst. Ebenso kann man sich nun überlegen, welche Stellen man im Gesang mehrstimmig singen kann oder an welchen Stellen das Schlagzeug vielleicht doch noch etwas anderes spielt.<sup>21</sup>

In dieser Phase ist das Ziel am Ende eine Art Skizze zu haben, anhand derer man den Song schließlich final aufnehmen kann.<sup>22</sup>

Im Folgenden wird eine selbsterstellte Demo eines eigenkomponierten Songs gezeigt. Anschließend wird etwas detaillierter die eigene Herangehensweise betrachtet und analysiert, um zu zeigen welche Arbeit und welche Gedanken hinter der Pre-Production stecken.

Beispiel 1: Slave or King Demo

[Hier anhören](#)

### 2.3.1.3 Analyse der eigenen Herangehensweise:

Die Grundidee des Songs „Slave or King“ war die Rhythmusgitarre aus dem Intro. Diese wurde mit einem Handy aufgenommen und laut abgespielt, damit man darüber eine Idee für die Lead-Gitarre entwickeln konnte. Das Ergebnis war eine einfache Tonfolge, die sich über die wechselnden Akkorde immer wieder wiederholte. Danach wurde die Rhythmusgitarre für den Vers, Pre-Chorus und Chorus geschrieben. In diesem Fall entstand also zuerst die Musik (bzw. die Gitarrenstimme), bevor es irgendeine Art von Gesangsmelodie oder Text gab.

Nachdem ein grobes Harmoniegerüst und ein Rhythmus gegeben war, wurde an der Melodie gearbeitet.

---

<sup>20</sup> Vgl. Fresh, D. (2017).

<sup>21</sup> Vgl. Dumont, L. (o.D.).

<sup>22</sup> Vgl. O.V. peak-studios.de (2024).



**Abbildung 1: Melodie Slave or King**

Quelle: Eigene Abbildung erstellt mit Guitar Pro 7

Die Melodie vor dem Text zu schreiben, wird vom Autor als einfacher wahrgenommen, um eingängigere Phrasen zu schreiben. Diese Methode kommt allerdings mit der Herausforderung einen Text zu finden, dessen Silben genau auf die Noten passen. Hierzu wurden von mir folgende Methoden verwendet:

Manche Wörter kann man z.B. auf zwei unterschiedliche Weisen segmentieren:

„ev-e-ry“ oder „ev-ery“.

Durch Abkürzungen oder Apostrophe lässt sich ebenfalls die Silbenanzahl verändern:

„I'm“ statt „I am“; „sec“ statt „se-cond“

Ebenso kann mit Ellipsen, also dem Weglassen von Wörtern (z.B. Pronomen) arbeiten:

„Have a ticking time bomb in every cell“ oder „I have a ticking time bomb in every cell“

Gegensätzlich dazu kann man Wörter nutzen, um Lücken aufzufüllen:

„It's just hysteria.“ statt „It's hysteria.“

Vokale können durch gebundene Noten verlängert werden.

„Don't feel my fe-ee-et“

Wenn nichts der oben genannten Dinge funktioniert, kann man versuchen Sätze umzuformulieren oder nach Synonymen mit der passenden Silbenanzahl zu suchen. Durch diese Methoden kann man flexibler mit einer potenziellen Textzeile umgehen und diese an die gewünschte Melodie anpassen.

Als die Melodie des Songs feststand, wurde versucht frei heraus einzelne Phrasen mit einem zufälligen Text zu singen. Es ist möglich, dass hierbei direkt ein Text zu einer ganzen Phrase entsteht. Es kann jedoch auch sein, dass der Ausgangspunkt nur das Endwort der Phrase ist und man den Satz dann teils mit oben genannten Methoden erst von hinten konstruiert. Gerade bei letzterem Fall probiert man verschiedene Wörter mit

verschiedenen Vokalen aus und arbeitet an dem weiter, was persönlich am meisten zu-sagt. Exemplarisch wurde das Wort „feet“ im Refrain über dem Wort „foot“ präferiert, wahrscheinlich auch aus dem Grund, da es zusammen mit dem Anfang des Refrains „Don't feel“ einen unreinen Reim bildet.

Reime wirken auf unser Gehirn stimulierend und lassen uns Melodien besser merken. Bei sich wiederholenden melodischen Phrasen bietet es sich an, besonders am Ende Reime einzubauen.<sup>23</sup>

Als Hilfsmittel wurde Teils sogar die Website Rhymezone.com verwendet. Eine Liste mit möglichen Wörtern zu haben, lässt einen auf weitere Ideen kommen. Hilfreich kann es sein auch unreine Reime in Betracht zu ziehen, um den Pool an möglichen Alternativen zu vergrößern.

Vers 1 mit markierten Reimen:

***My master*** is pulling the strings in a wrong ***direction***

*Have I ever had control over myself?*

***Disaster*** can't be avoided, my life is ***reckless***

*Have a ticking time bomb in every cell* – *Causing the Avalanche (Philipp Sonntag)*

Auf diese Weise findet man Themen und baut sich letztendlich den Songtext wie ein Puzzle zusammen. Der Refrain wurde mehrfach wiederholt und wird durch die Verse Storyline-technisch miteinander verbunden.

Inhaltlich geht es in dem Song „Slave or King“ darum, dass man sich entweder wie das Opfer seines Schicksals sehen kann oder dieses selbst in die Hand nimmt. Es geht um das Gefühl der Machtlosigkeit, aber auch darum loslassen zu können und die Situation zu genießen. Der Refrain stellt diesen Gedanken in den Kontext eines Konzertes, in dem man nur Teil einer Masse ist und sich fühlt, als hätte man die Kontrolle über sich selbst verloren. Gleichzeitig ist der Chorus vage genug formuliert um Interpretationsspielraum zu lassen und ihn auf andere Lebensbereiche beziehen zu können.

Nachdem Text und Melodie fertig waren und zusammen mit den Gitarrenstimmen grob in einer DAW aufgenommen wurden, wurde mithilfe eines Drum Plugins ein Schlagzeug skizziert. Auf den Bass wurde bei der Demo absichtlich verzichtet, da er rhythmisch und harmonisch meistens dasselbe spielt wie die Rhythmusgitarre und keine eigenständige Melodie hat.

---

<sup>23</sup> Vgl. Reinartz, L.S. (o.D.).

Anschließend wurden Harmonies an bestimmten Stellen im Gesang aufgenommen und teils schon mit Effekten versehen (siehe Demo 1:02 - 1:06). Auch bei diesem Schritt hat man erst verschiedene Möglichkeiten ausprobiert und sich dann für eine Option entschieden.

Auf richtige Backing Tracks die man additional zu einer Lead-Stimme links/ rechts part wurde verzichtet. Nur im ersten Vers wurde mit einer zweiten Unisono-Stimme etwas versucht und getestet.

Zuletzt wurde für die Demo ein Synthesizer an bestimmten Stellen eingefügt, um sich leer fühlende Stellen klanglich aufzufüllen.

Zusammenfassend kann man sagen, dass einiges an Arbeit schon in die Pre-Production und ins Songwriting fließt. Nicht alles, was hier entschieden wird, schafft es jedoch in die finale Aufnahme. Im Studio und in der Postproduktion wurden bei Slave or King noch immer Dinge entschieden, ausprobiert und verändert, sei es Bass, Synthesizer, Backings oder sogar Lead Vocals und Lyrics.

Die Demo stellt einen groben Fahrplan dar, bei dem es wichtig ist zu wissen, was musikalisch gemeint ist und was man in der finalen Aufnahme anders macht. Es ist wichtig, dass falls ein Producer vorhanden ist, dieser den Fahrplan kennt und sich vorher mit der Band oder dem Sänger und seinen Ideen auseinandergesetzt hat. Vokales Arrangement, gewünschter Stil und das Gefühl, dass der Künstler dem Publikum übermitteln möchte, werden ebenfalls vor der Produktion besprochen. Anhand der Stimme und dem gewünschten Sound kann der Producer dann schließlich eine Mikrofonauswahl treffen. Dieser Schritt ist nicht zu unterschätzen, denn das Mikrofon kann einen signifikanten Unterschied im generellen Sound der finalen Aufnahme machen.<sup>24</sup>

Für den Song Slave or King hat der Schlagzeuger der Band die Rolle des Produzenten übernommen. Nach einem kleinen Mikrofon Shootout, wurde sich für das SM 7 B von Shure entschieden, da dieses innerhalb des verfügbaren Budgets die besten klanglichen Eigenschaften aufwies und die Stimme des Sängers besonders gut hervorhob.

---

<sup>24</sup> Vgl. Bricio, C. (2023).

### 2.3.2 Produktion von Gesang

Die Musikproduktion kann in die Schritte Recording, Editing, Mixing und Mastering, Export unterteilt werden.<sup>25</sup>

Der nächste Teil wird dabei sich vor allem auf das Recording und Editing bei Gesangsproduktionen fokussieren. Das Mixing und Mastering wird der Vollständigkeit wegen nur kurz angeschnitten.

„Vocal production is the art of capturing, refining and enhancing the human voice for presentation on a musical recording.“<sup>26</sup>

#### 2.3.2.1 Recording

Nachdem ein Mikrofon ausgesucht wurde, geht es nun darum, die Performance eines Sängers festzuhalten. Diese sollte sich authentisch anfühlen und von Herzen kommen. Ein guter Producer schafft es den Sänger in einen Zustand zu bringen, in dem er sich wohlfühlt, sich in den Song einfühlen und das Beste aus sich rausholen kann. Es sollte eine Umgebung geschaffen werden, in der es dem Sänger möglich ist, sich frei auszudrücken und seinen Emotionen freien Lauf zu lassen.<sup>27</sup> Gesang ist einer der schwierigsten Disziplinen in der Musik und die Performance ist eng mit dem Selbstbewusstsein und den Emotionen des Sängers verbunden.<sup>28</sup>

Producer sind also nicht nur für den technischen Part zuständig, sondern agieren zusätzlich als Coach und Motivator. Das beinhaltet es unter anderem auch Timing oder Tonhöhenprobleme anzusprechen, andere Gesangstechniken zu empfehlen oder ein Doppeln der Stimme bzw. Harmonien vorzuschlagen. Gesangsparts werden teils phrasenweise wiederholt, um einen perfekten Take zu bekommen.<sup>29</sup>

Studiotechniken wie Punch-Ins erhöhen dabei die Chance für den Sänger schwierige Passagen erfolgreich darzubieten.<sup>30</sup>

Do-it-yourself (DIY) – Artists müssen während eines Recordings sowohl in die Rolle des Producers als auch in die des Sängers schlüpfen. Beide Rollen benötigen verschiedene Mindsets und es ist wichtig eine gute Balance zu finden zwischen dem kontrollierenden, technischen und dem emotionalen, darstellenden Teil.<sup>31</sup>

---

<sup>25</sup> Vgl. Dumont, L. (o.D.).

<sup>26</sup> Vgl. Hahn, M. (2024).

<sup>27</sup> Vgl. Bricio, C. (2023).

<sup>28</sup> Vgl. Hahn, M. (2024).

<sup>29</sup> Vgl. Bricio, C. (2023).

<sup>30</sup> Vgl. Hahn, M. (2024).

<sup>31</sup> Ebd.

Erwähnenswert bezüglich des Recordings ist weiterhin, dass bei der Aufnahme von „Slave or King“ festgestellt wurde, dass man mit einem guten Mix auf dem Ohr und etwas Hall auf der eigenen Stimme bessere Ergebnisse bezüglich der Intonation erzielen konnte. Ein Click bzw. ein vorher geeditetes/quantisiertes Schlagzeug sorgen ebenfalls für Vocals, die mehr in-time sind.

### **2.3.2.2 Editing**

Die Audiotbearbeitung bzw. das Editing fällt meist in den Aufgabenbereich des Tontechnikers. Dieser wählt die besten Aufnahmen und Takes aus, schneidet diese zu einer Spur zusammen und entfernt ungewünschte Atmer und sonstige Störgeräusche.<sup>32</sup> Zusätzlich werden Tonhöhe<sup>33</sup>, Timing und Groove korrigiert, sodass am Ende eine perfekte trockene Vocalspur zu hören ist.<sup>34</sup>

#### **2.3.2.2.1 Comping**

Das Zusammenschneiden zu einem Take wird in der Fachsprache auch als Comping bezeichnet. Ein Sänger singt in der Regel mehrere Takes des Songs oder einer Passage ein. Diese unterschiedlichen Versionen werden dabei in einem separaten Ordner oder in Playlists abgespeichert.<sup>35</sup>

Nach einer Recording Session werden diese Playlists dann durchgehört und die besten Stellen und Segmente markiert oder direkt in eine neue finale Spur gezogen und zusammengeschnitten.<sup>36</sup>

#### **2.3.2.2.2 Tonhöhenkorrektur**

Nicht immer ist die Performance die am emotionalsten Klang, die bei der die Töne am besten getroffen wurden. Ebenfalls kommt es manchmal auch vor, dass aus der Recording Session einfach zu wenig gute Takes entstanden sind. In beiden Fällen ist die Tonhöhenkorrektur nützlich.<sup>37</sup>

Unter Tonhöhenkorrektur denkt man eventuell als erstes an Auto-Tune oder ähnlich funktionierende Software, bei der die Stimme einen roboterartigen Effekt erhält.<sup>38</sup>

---

<sup>32</sup> Vgl. Dumont, L. (o.D.).

<sup>33</sup> Vgl. Hahn, M. (2024).

<sup>34</sup> Vgl. Dumont, L. (o.D.).

<sup>35</sup> Vgl. Hahn, M. (2024).

<sup>36</sup> Vgl. Hahn, M. (2024).

<sup>37</sup> Ebd.

<sup>38</sup> Ebd.

Dieser Effekt, den man aus Songs wie „Believe“ von Cher oder diversen T-Pain Songs kennt, ist jedoch nur ein extremes Beispiel für die Nutzung derartiger Software. Viele Anwendungen zielen darauf ab, kleine Intonationsfehler zu beheben und dabei eine natürlich klingende Stimme zu bewahren. Entsprechende Software kann also auch weitaus subtiler genutzt werden.<sup>39</sup>

Manche DAWs wie Steinbergs Cubase oder Logic haben Pitch Correction Plugins schon integriert. Das Third-Party-Plugin Melodyne von Celemony gehört zu den besten in dem Bereich, wenn es um natürlichen Klang geht.<sup>40</sup>

### **2.3.2.2.3 Audiorestaurierung**

Unabhängig davon, wie gut die Aufnahmebedingungen sind, weisen Vocal Takes oft noch kleinere Unvollkommenheiten auf. Hintergrund-, Atem- oder Mundgeräusche können jedoch ebenfalls mit entsprechender Software entfernt werden und so eine Aufnahme auf ein anderes Level heben.<sup>41</sup>

### **2.3.2.3 Mixing und Mastering**

Nach der Audibearbeitung aller Spuren werden diese gemixt und gemastert. Producer sorgen, während dem Mixing dafür, dass sich Instrumente nicht im Weg stehen und dass sich jedes wichtige Element im Song durchsetzen kann. Vor allem der Gesang sollte jederzeit gut hörbar und verständlich sein, damit die Worte zum Zuhörer durchdringen und dieser emotional angesprochen wird.<sup>42</sup> Kompressoren, EQs, Reverbs und Delays sind typische Werkzeuge mit denen Toningenieure in diesem Schritt arbeiten.<sup>43</sup>

Beim nachfolgenden Mastering wird das Audiomaterial auf die entsprechende Lautheit gebracht und eventuell noch kleine Anpassungen mit dem EQ gemacht. Gerade bei mehreren Songs bzw. Alben ist es wichtig, die Songs aufeinander anzupassen und den Zuhörer auf eine Reise mitzunehmen. Das Mastering bereitet das Audiomaterial nach allen Standards für die CD-Pressung etc. vor.<sup>44</sup>

In folgendem Hörbeispiel wird die finale Version des Songs Slave or King gezeigt, um den klanglichen Unterschied zur Demo zu verdeutlichen. Es wurde sich deutlich mehr

---

<sup>39</sup> Vgl. Jones, A. (2022).

<sup>40</sup> Ebd.

<sup>41</sup> Vgl. Hahn, M. (2024).

<sup>42</sup> Vgl. Dumont, L. (o.D.).

<sup>43</sup> Vgl. Hahn, M. (2024).

<sup>44</sup> Vgl. Dumont, L. (o.D.).

Zeit als bei der Demo gelassen, um Takes einzusingen bzw. einzuspielen. Jede Spur wurde geedited, verfeinert und danach mit den anderen Spuren gemixt.

Im Refrain wurden Backings/ Dopplungen eingesungen, um dem Gesang hier mehr Kraft zu verleihen. Harmonies wurden ebenfalls in mehreren Spuren aufgenommen.

Beispiel 2: Slave or King finale Version

[Hier anhören](#)

### 3 Künstliche Intelligenz und Musik

Um zu verstehen, wie Künstliche Intelligenz (KI) in einem Musikproduktionsprozess eingesetzt werden kann, werden im folgenden Abschnitt zunächst einige begriffliche Grundlagen zum Thema KI erklärt. Anschließend werden einige KI-Applikationen betrachtet.

#### 3.1 Allgemeine begriffliche und theoretische Grundlagen zu KI

Die Auffassung von Künstlicher Intelligenz, im Englischen als Artificial Intelligence (AI) bekannt, variiert sehr stark und ist kompliziert. Eine allgemeingültige Definition fehlt momentan, da sich das Verständnis davon, was Künstliche Intelligenz ausmacht, ständig und dynamisch wandelt.<sup>45</sup>

In Anbetracht dessen, ist es zunächst sinnvoll den Begriff in seine wesentlichen Bestandteile aufzuspalten: künstlich und intelligent.

Künstlich ist ein Objekt oder ein Verhalten, wenn es „nicht natürlich, sondern mit chemischen und technischen Mitteln nachgebildet, nach einem natürlichen Vorbild angelegt, gefertigt, geschaffen“ wurde.<sup>46</sup>

Ebenso kann man etwas als künstlich bezeichnen, wenn es natürliche Vorgänge nachahmt, nicht auf natürliche Weise vor sich geht oder wenn etwas gekünstelt oder unnatürlich wirkt.<sup>47</sup>

Intelligenz kann wird von der Dudenredaktion hingegen auf folgende Weise definiert: Sie ist die „Fähigkeit [des Menschen], abstrakt und vernünftig zu denken und daraus zweckvolles Handeln abzuleiten“. Des Weiteren kann es sich bei einer Intelligenz um ein „vernunftbegabtes/ intelligentes (Lebe-)Wesen“ handeln.<sup>48</sup>

Um zu beurteilen, ob eine Maschine oder ein Computer Intelligenz besitzt, entwickelte Alan Turing 1950 einen Test, bei dem ein Proband als Fragesteller mittels eines Bildschirms und einer Tastatur (d.h. ohne Sicht- und Hörkontakt) in eine Unterhaltung zwischen zwei Gesprächspartnern geworfen wird. Nach der Unterhaltung muss der Proband entscheiden, welcher der Gesprächspartner der Computer und welcher der Mensch war. Kann er dies nicht mit Sicherheit beantworten oder tippt womöglich sogar falsch, könne der Computer intelligent genannt werden. Allerdings wurde dieser Test

---

<sup>45</sup> Vgl. Mockenhaupt, A. (2021). S.49

<sup>46</sup> Duden.de künstlich (o.D.)

<sup>47</sup> Vgl. Duden.de künstlich (o.D.)

<sup>48</sup> Duden.de Intelligenz (o.D.)



immer wieder kritisiert, da er nur in einem gewissen Umfang das testet, was menschliche Intelligenz ausmacht. Bewusstsein, Emotion oder Intentionalität werden beispielsweise gar nicht in dem Test berücksichtigt. Die Idee hinter dem Turing-Test ist trotzdem interessant, denn es wird damit folgende Frage aufgeworfen: Ist ein Computer dazu in der Lage, einen Menschen zu imitieren?<sup>49</sup> Sowohl das Fraunhofer Institut als auch das Europäische Parlament betrachten dies als Kernaspekt, wenn es darum geht KI zu definieren.<sup>50</sup>

„Künstliche Intelligenz ist die Fähigkeit einer Maschine, menschliche Fähigkeiten wie logisches Denken, Lernen, Planen und Kreativität zu imitieren.“<sup>51</sup> Diese kognitiven Fähigkeiten, sind das Resultat aus der Sortierung und Erkennung von Informationen aus Eingabedaten.<sup>52</sup> Technischen Systemen ist es durch KI möglich, ihre Umwelt zu erkennen, auf diese zu reagieren und Aufgaben zu lösen, um spezifische Ziele zu erreichen. Der Computer erhält Daten über beispielsweise eine Kamera oder sonstige Sensoren, verarbeitet diese und reagiert entsprechend.<sup>53</sup>

Künstliche Intelligenz kann auf zwei verschiedene Weisen erzeugt werden. Zum einen, kann sie auf Abläufen, die programmiert wurden, basieren, zum anderen kann sie das Ergebnis maschinellen Lernens (Engl.: Machine Learning) sein. In diesem Gebiet wurden besonders in den letzten Jahren immense Fortschritte gemacht, da die verfügbare Rechenleistung und Datenmengen, welche für die komplizierten Berechnungen des maschinellen Lernens unabdingbar sind, stetig gewachsen sind.<sup>54</sup>

### 3.1.1 Machine Learning

Maschinelles Lernen zeichnet sich dadurch aus, dass mit Hilfe von Trainingsdaten ein Modell abstrahiert wird. Diese Trainingsdaten werden selbständig ohne den Eingriff des Menschen analysiert.<sup>55</sup> D.h. im Gegensatz zu traditionellen Algorithmen wird kein Lösungsweg vorgegeben. Der Computer erkennt selbständig die Struktur der Daten und lernt durch wiederholte Ausführung allein eine Aufgabe zu erfüllen.<sup>56</sup>

---

<sup>49</sup> Vgl. Scheuer, D. (2020). S.7f.

<sup>50</sup> Vgl. Fraunhofer Institut (o.D.); Europäisches Parlament (2020).

<sup>51</sup> Europäisches Parlament (2020).

<sup>52</sup> Vgl. Fraunhofer Institut (o.D.).

<sup>53</sup> Vgl. Europäisches Parlament (2020).

<sup>54</sup> Vgl. Fraunhofer Institut (o.D.).

<sup>55</sup> Vgl. Scheuer, D. (2020). S.19f.

<sup>56</sup> Vgl. Fraunhofer Institut (o.D.).

Wurde das Modell, welches im Prinzip einer Hypothese ähnelt, generiert, wird es auf die tatsächlichen Felddaten (Input Daten) angewendet und ein Output generiert.<sup>57</sup>

Anwendungsmöglichkeiten sind beispielsweise Roboter, die ein Objekt von A nach B bringen sollen. Dabei sind A und B die Input Daten. Der Roboter lernt aber selbständig durch Trail-and-Error, wie er ein Objekt greifen kann und dieses transportiert.<sup>58</sup>

### 3.1.2 Deep Learning und neuronale Netze

Ein weiterer Begriff, den man oft im Zusammenhang mit KI und Machine Learning wahrnimmt, ist das Deep Learning oder auch die neuronalen Netze. Neuronale Netze sind ein Untergebiet des Machine Learnings. Es handelt sich hierbei um Lernalgorithmen, die vom menschlichen Gehirn und dessen Nervenzellenverbindungen inspiriert sind. Statt aus Synapsen und Neuronen sind künstliche neuronale Netze jedoch aus Datenknoten zusammengesetzt. Diese wiederum sind durch Verbindungen, welche unterschiedlich gewichtet sein können, vernetzt.<sup>59</sup>

Aufgebaut sind neuronale Netz aus verschiedenen Layers bzw. Schichten. In folgender Abbildung ist ein Beispiel skizziert. Links ist die sogenannte Input-Layer, mit der das Netz Umweltinformationen enthält. In der Mitte befindet sich die Aktivitäten-Layer (auch hidden layer genannt) und rechts schließlich die Output Layer, die ein Ergebnis zeigen soll.<sup>60</sup>

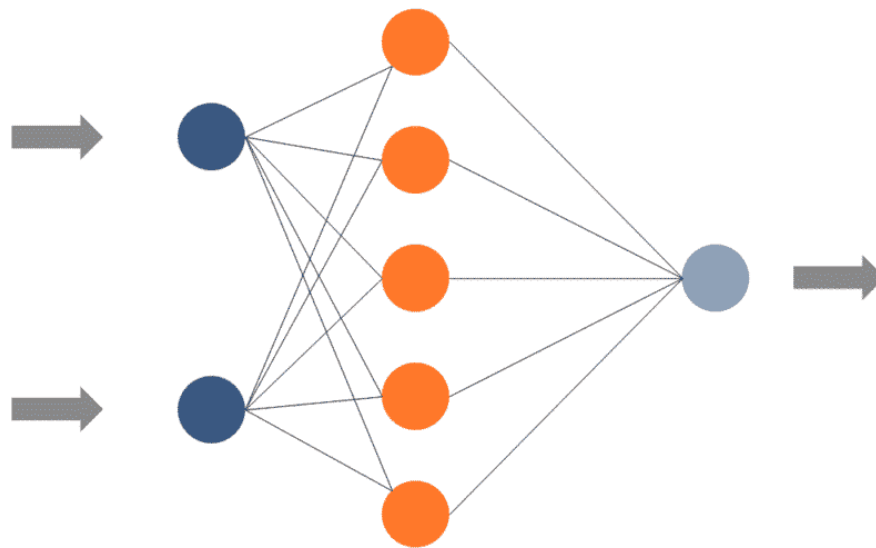
---

<sup>57</sup> Vgl. Scheuer, D. (2020). S.19f.

<sup>58</sup> Vgl. Fraunhofer Institut (o.D.).

<sup>59</sup> Ebd.

<sup>60</sup> Vgl. Tiedemann, M. (2018).

**Abbildung 2: Aufbau eines einfachen neuronalen Netzes**

Quelle: Tiedemann, M. (2018).

Der Unterschied zwischen einem einfachen neuronalen Netz und einem Deep Learning Algorithmus ist der, dass es in der Mitte statt einer Zwischenschicht mehrere hidden layers existieren.<sup>61</sup> Die Zahl dieser Schichten kann hierbei bis in die Hunderttausende oder sogar Millionen gehen.<sup>62</sup>

Problematisch sind die hidden layers in der Hinsicht, dass für Außenstehende es nicht möglich ist, zu betrachten was eigentlich darin passiert. Zwischen Input- und Output-Schicht befindet sich dementsprechend eine Black Box. Somit ist es notwendig, den Ergebnissen ein gewisses Maß an Vertrauen entgegenzubringen.<sup>63</sup>

In den letzten Jahren wurde Deep Learning zu einem der wichtigsten Motoren in der KI-Forschung und -Entwicklung. Durch diese Algorithmen wurde es möglich geistige und kognitive Prozesse darzustellen, von denen man lange dachte, dass sie nur dem Menschen möglich sind. Der zweite Grund sind die sehr guten Ergebnisse, die erzielt werden, wenn Deep Learning Algorithmen mit Big Data, also sehr großen Datenmengen, trainiert werden.<sup>64</sup>

---

<sup>61</sup> Tiedemann, M. (2018).

<sup>62</sup> Vgl. Fraunhofer Institut (o.D.).

<sup>63</sup> Vgl. Scheuer, D. (2020). S.21

<sup>64</sup> Vgl. Tiedemann, M. (2018).

### 3.1.3 Starke und schwache KI

Um lernende Systeme einzuordnen werden häufig die Begriffe starke und schwache KI verwendet. Mit schwacher KI ist dabei ein System gemeint, das sich nur auf bestimmte Anwendungsprobleme fokussiert. Starke KI ist hingegen eine künstliche Intelligenz, die der menschlichen Intelligenz ebenbürtig oder sogar voraus ist. Umgebungswahrnehmung, sinnvolle Handlungen und ein Verständnis für identifizierte Komponenten sind einige Kennzeichen einer starken KI. Forscher rechnen damit, dass letztere Form erst in über 20 Jahren realisierbar sein wird. Bislang existiert sie weitestgehend lediglich im Forschungsumfeld.<sup>65</sup>

Bei allen KIs die in dieser Arbeit vorgestellt werden handelt es sich folglich um schwache KIs.

## 3.2 Geschichte von KI in der Musik

Frühe Versuche Musik generieren zu lassen bzw. nicht in einem traditionellen Stil zu komponieren, belaufen sich zurück ins 18. Jahrhundert. Eine Möglichkeit den Zufall in der Musik zu inkorporieren ist der Gebrauch eines einfachen Würfels. Das musikalische Würfelspiel, welches 1792 veröffentlicht wurde und Mozart zugeschrieben wird, lässt den Spieler einen 16-taktigen Wiener Walzer komponieren. Dabei wird für jeden Takt des Stücks gewürfelt und so zufällig entschieden welcher vorkomponierte musikalische Schnipsel in den Takt eingefügt wird. Eine der Prämissen des Spiels ist, dass jedes dieser vordefinierten Schnipsel in Kombination mit jedem möglichen Schnipsel des nächsten Taktes musikalisch gesehen Sinn ergibt. Bei dem Spiel sind für 14 der 16 Takte 11 verschiedene Möglichkeiten gegeben. Die anderen zwei Takte haben eine und zwei mögliche Varianten. Dies ergibt in Summe  $1 * 2 * 11^{14} = 759\,499\,667\,166\,482$  verschiedene mögliche Wiener Walzer.<sup>66</sup> Es handelt sich in diesem Beispiel nicht um KI, wie sie von uns definiert wurde, doch man kann das Würfelspiel von Mozart als Vorstufe, als „symbolische KI“ betrachten. Es besteht ein regelbasierter Ansatz und Zufall spielt eine Rolle, dennoch sind gespeicherte Informationen menschenlesbar vorhanden und Ergebnisse jederzeit nachvollziehbar.<sup>67</sup>

Fortschrittlichere Versuche mit KI in Musik gibt es in den 50er Jahren.<sup>68</sup> In dieser Zeit haben Lejaren Hiller und Leonard Isaacson zusammen an der „Illiac Suite“ gearbeitet, einem klassischen Musikstück für Streichquartett, welches sogar noch heute ab und zu

---

<sup>65</sup> Vgl. Mockenhaupt, A. (2021). S.52

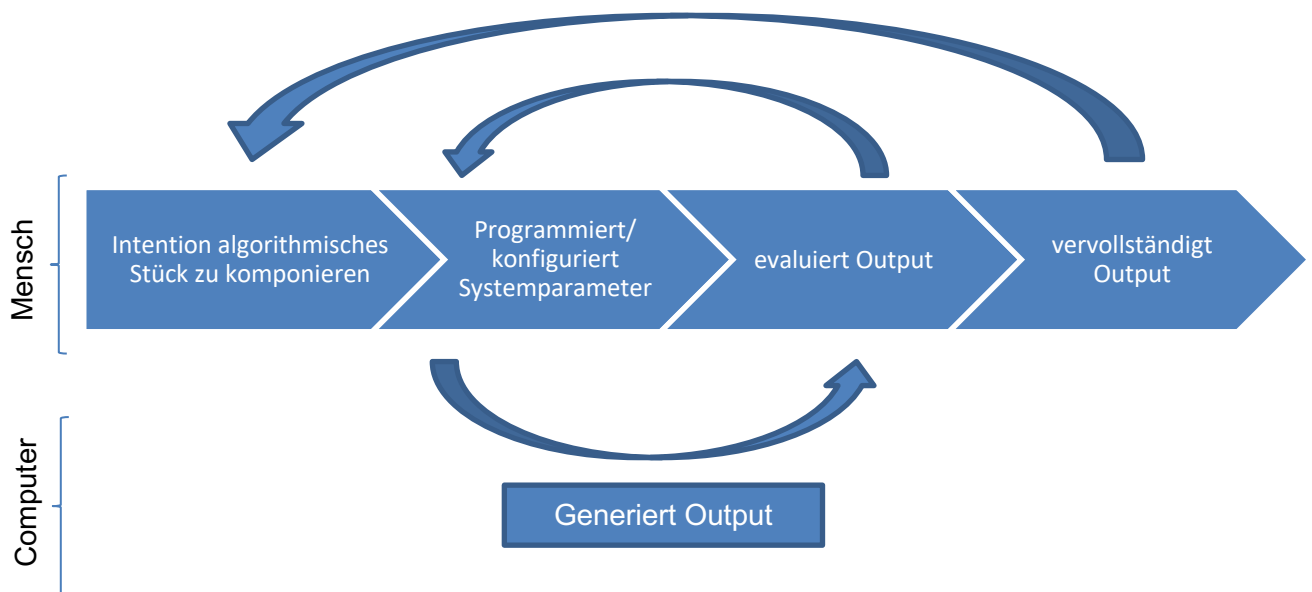
<sup>66</sup> Vgl. Guo, R. (2023). S. 57

<sup>67</sup> Vgl. Klawonn, T. (2023). S. 30 f.

<sup>68</sup> Vgl. Brown, O. (2021). S. 2

aufgeführt wird.<sup>69</sup> Die Partitur der Illiac Suite gilt weitgehend als die erste, die mithilfe des kreativen Inputs von Algorithmen auf einem Computer komponiert wurde und ist damit ein wichtiger Meilenstein. Jede der vier Sätze wurde hierbei mit einer anderen Methode von algorithmischer Komposition erstellt (z.B. durch stochastische Generation mithilfe des Markov Modells oder durch handcodierte Regeln). Beim Anhören der Illiac Suite könnte man meinen, dass automatisierte Musikgeneration kein großes Problem darstellt, denn die Suite klingt harmonisch, emotional und gut strukturiert. Tatsächlich ist die Suite Produkt eines mühsamen iterativen Trail-and-Error-Prozesses, bei dem der Output des Computers immer wieder von neuem evaluiert wurde und das Programm angepasst wurde. Dieser Ansatz sollte sich für das restliche Jahrhundert als prädominante Form bei Experimenten mit Computer generierter Musik herausstellen.<sup>70</sup>

**Abbildung 3: Typische Mensch – Maschinen Interaktion bei algorithmischen Kompositionen bei denen der Komponist gleichzeitig Nutzer oder Programmierer eines Programms ist**



Quelle: Eigene Darstellung nach Brown, O. (2021). S.3

Neben dem Markov Modell wurden in der Vergangenheit auch andere Methoden vorgeschlagen, um Musik zu erzeugen.<sup>71</sup>

David Cope führte 1997 ein Stück auf, das von seinem Programm „Experiments in Musical Intelligence“ (EMI) geschrieben wurde. Das Programm beruht auf einem grammatikalischen Ansatz, bei dem besonders die Struktur der Musik im Mittelpunkt steht. Es ist

<sup>69</sup> Vgl. Klawonn, T. (2023). S. 33

<sup>70</sup> Vgl. Brown, O. (2021). S. 2

<sup>71</sup> Vgl. Guo, R. (2023). S. 59

dazu in der Lage Teile und deren Funktion in einem Musikstück zu erkennen, in Kernelemente zu zerlegen und diese neu anzuordnen. Die Ergebnisse sind dabei täuschend echt und kommen dem Stil des Komponisten, mit dessen Stücken das Programm gefüttert wurde, sehr nahe. Allerdings kann das Programm nicht erkennen, ob der generierte Output gut ist, noch ist es dazu in der Lage neue Stile zu erschaffen. Das Stück, das letzten Endes aufgeführt wurde, wurde von Cope von Hand aus Tausenden von Ergebnissen ausgewählt.<sup>72</sup>

Die vorgestellte Auswahl an Techniken haben sich vor allem aufgrund des hohen menschlichen Mehraufwandes und der zum Teil schlechten Ergebnisse auf dem Markt nicht durchsetzen können. Dennoch sind sie ein wichtiger Bestandteil der KI-Entwicklung und heben die heutzutage genutzten Deep Learning Systeme vor.<sup>73</sup>

---

<sup>72</sup> Vgl. Klawonn, T. (2023). S. 34 f.

<sup>73</sup> Ebd.

## 4 Analyse verschiedener KI-Applikationen in Bezug auf den Gesamtproduktionsprozess

### 4.1 ChatGPT

ChatGPT (kurz für Chat Generative Pre-trained Transformer) wurde von OpenAI im November 2022 veröffentlicht. Bei dieser auf einem Machine Learning basierenden Modell handelt es sich um eine generative KI, die menschliche Sprache versteht und natürlich wirkende Antworten geben kann. Texte können damit zusammengefasst, Gedichte oder Programme geschrieben werden.<sup>74</sup>

ChatGPT gehört zu einer der größten Innovationen im Bereich Natural Language Processing, KI und neuronalen Netzen.<sup>75</sup> Die Transformer-Architektur, auf der das Modell beruht, gehört zu den fortschrittlichen Architekturen von Deep Learning. Ausschlaggebend bei dieser Art von Architektur ist die sogenannte Self-Attention-Layer. Diese kann bewerten, welche Elemente eines Umfelds eines betrachteten Tokens wichtig sind. Ein Token kann hierbei alles Mögliche sein: ein Pixel im Foto, ein Wort eines Satzes, eine Note einer Melodie. Durch diese Bewertung kann der Kontext jedes Elements berücksichtigt werden, was wiederum zu natürlicheren Ergebnissen führt. Besonders bei der Spracherzeugung und Spracherkennung konnten mittels dieser Methode ausgesprochen gute Ergebnisse gezeigt werden.<sup>76</sup>

Die aktuelle Version von ChatGPT wurde mit einer enormen Datenmenge trainiert. 570GB an Daten von unterschiedlichsten Quellen wie Büchern, Forschungsartikeln, Wikipedia-Einträgen und mehr wurden in das System eingespeist. In Wörtern gezählt sind das rund 300 Milliarden.<sup>77</sup> Andere Chatbots wie zum Beispiel der direkte Konkurrent Google Gemini (ehemals Google Bard AI) liefern im direkten Vergleichstest in einigen Bereichen bessere Ergebnisse.<sup>78</sup> Allerdings besitzt ChatGPT einen größeren Wortschatz<sup>79</sup>, kann mehr Sprachen sprechen und scheint allgemein wortgewandter zu sein, wenn es darum geht, kreativen Output, wie zum Beispiel Gedichte zu liefern.<sup>80</sup> Genau aus diesen Gründen, wird versucht mit diesem KI-Tool ein Songtext zu generieren.

---

<sup>74</sup> Vgl. Berens, A.; Bolk, C. (2023). S. 43f.

<sup>75</sup> Vgl. Singh S.K.; Kumar, S.; Mehra P. S.1

<sup>76</sup> Vgl. Klawonn, T. (2023). S. 61

<sup>77</sup> Vgl. Hughes, A. (2023).

<sup>78</sup> Vgl. Drapkin, A. (2024).

<sup>79</sup> Vgl. Singh S.K.; Kumar, S.; Mehra P. S.6

<sup>80</sup> Vgl. Drapkin, A. (2024).

### 4.1.1 Prompting

Generative KIs wie ChatGPT benötigen Prompts, um Output zu generieren. Beim Prompting gibt man der KI im Prinzip Anweisungen, was sie genau tun oder erstellen soll. Je besser der Prompt ist, desto effektiver kann die KI genutzt werden. Eine Eingabeaufforderung kann neben einer Anweisung, Beispiele, Fakten, Fragen, Eingabedaten und mehr enthalten. Die Kombination verschiedener Elemente lässt die Wahrscheinlichkeit wachsen, dass genau das Bild, das Video oder der Text, den man sich gewünscht hat, generiert wird.<sup>81</sup> Kurz gesagt, kann man sagen: „Je besser der Prompt, desto besser ist das Ergebnis.“<sup>82</sup>

Ein guter Prompt sollte mehrere Kriterien erfüllen:

1. Er sollte klar und spezifisch sein.<sup>83</sup>  
Ein klar definierter Prompt gewährleistet, dass die KI versteht, was gemeint ist.
2. Er sollte prägnant sein.<sup>84</sup>
3. Er sollte viele kontextbezogene Informationen beinhalten, denn KI-Tools fällt es ab und zu schwer, den Kontext eines Prompts einzuordnen.<sup>85</sup>

Die Fähigkeit gut zu prompten, wird inzwischen als eine der in Zukunft womöglich wichtigsten Schlüsselqualifikationen angesehen. Dabei ist es ratsam, seine Prompts oft zu variieren und sich so iterativ bis zum gewünschten Ergebnis heranzuarbeiten.<sup>86</sup> Inzwischen gibt es sogar sogenannte Prompt-Marketplaces, wie beispielsweise ChatX, PromptBase oder PromptSea, die hochklassige Prompts anbieten.<sup>87</sup>

---

<sup>81</sup> Vgl. Berens, A.; Bolk, C. (2023). S. 43f.

<sup>82</sup> Emmerich, N. (2023).

<sup>83</sup> Vgl. Berens, A.; Bolk, C. (2023). S. 45

<sup>84</sup> Ebd.

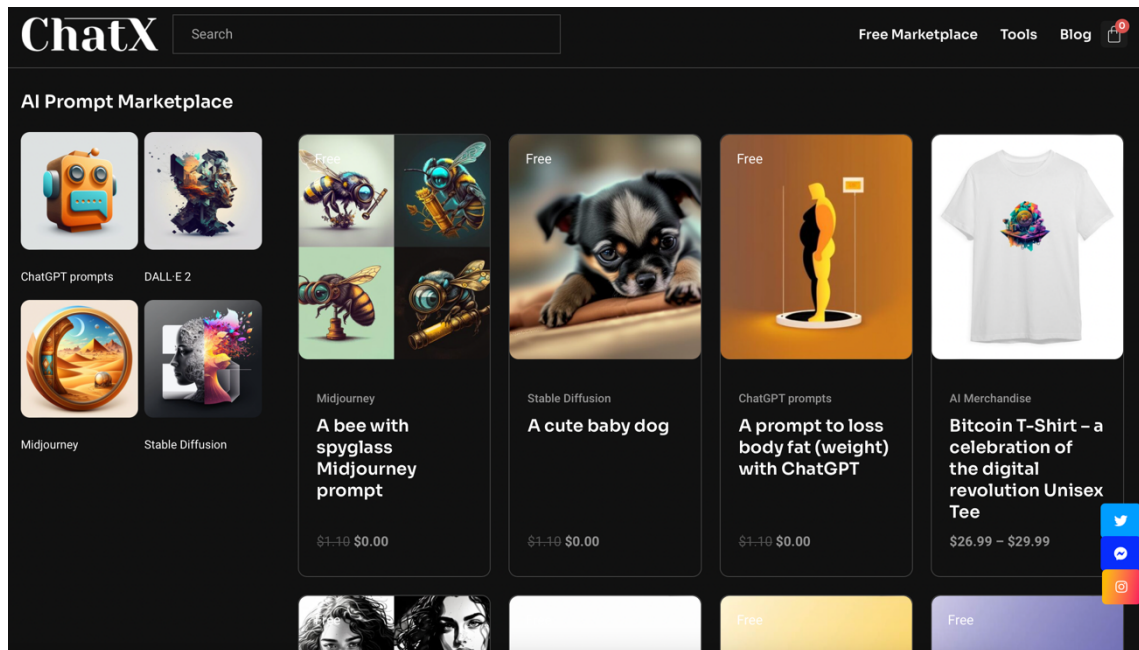
<sup>85</sup> Ebd.

<sup>86</sup> Vgl. Emmerich, N. (2023).

<sup>87</sup> Vgl. Berens, A.; Bolk, C. (2023). S. 46f.



Abbildung 4: Startseite ChatX



Quelle: chatx.ai (o.D.).

Bei dem kostenlosen Marketplace ChatX, konnten leider keine Prompts zum Thema Songwriting oder Musik gefunden werden, weswegen im folgenden Teil ein Prompt für das Vorhaben generiert werden soll. Berens und Bolk (2023) schlagen in Ihrem Buch dafür folgenden Text vor, den man in eine generative KI eingeben kann:

*„Ich möchte, dass du mein Prompt Ersteller wirst. Dein Ziel ist es, mir dabei zu helfen, den bestmöglichen Prompt für meine Bedürfnisse zu erstellen.*

*Der Prompt wird von dir, ChatGPT, verwendet. Du wirst dem folgenden Prozess folgen:*

*1. Deine erste Antwort wird sein, mich zu fragen, um was es in dem Prompt gehen soll. Ich werde dir darauf meine Antwort geben, aber wir müssen sie durch kontinuierliche Iterationen weiter verbessern. Dafür durchlaufen wir die nächsten Schritte.*

*2. Basierend auf meinen Angaben erstellst du 3 Abschnitte für die Ausgabe.*

*a) Überarbeiteter Prompt (gib deinen überarbeiteten Prompt an. Er sollte klar, prägnant und leicht verständlich für dich sein).*

*b) Vorschläge (gib Vorschläge, welche Details in dem Prompt enthalten sein sollten, um ihn zu verbessern).*

*c) Fragen (stelle alle relevanten Fragen, die sich auf zusätzliche Informationen beziehen, die benötigt werden, um den Prompt zu verbessern).“<sup>88</sup>*

<sup>88</sup> Berens, A.; Bolk, C. (2023). S. 48

Folgender Text wurde als erster Schritt eingegeben:

*„Das Ziel soll ein englischer Songtext werden, der die Menschen emotional berührt und das Potential hat, ein Welthit zu werden. Er soll den Zuhörer auf eine Reise mitnehmen und eine Story haben mit Kniff. Viele Reime auch innerhalb der Zeilen hören sich besser an als keine Reime.“*

Nachdem ChatGPT einen überarbeiteten Prompt erstellt hat, gibt uns die KI weitere Verbesserungsvorschläge und Fragen bezüglich der Anweisung. Dieser Schritt wird mehrfach wiederholt. Nach der dritten Iteration wurde folgender Prompt von ChatGPT erstellt:

*„Schreibe einen englischen Songtext im Genre Alternative Pop Rock, der gute Vibes, Sommerstimmung, Belebung und Chill vermittelt. Der Text soll eine fesselnde Geschichte über Freundschaft, Leben, Liebe und Nostalgie erzählen, wobei eine Szene eines Abends am Strand im Mittelpunkt steht. Das Gefühl der Freiheit und das Bewusstsein, da zu sein, wo man sein sollte, sollen hervorgehoben werden. Die Emotionen sollen das Gefühl widerspiegeln, mit einem Lächeln auf das zurückblicken zu können, was war. Die Zuhörer sollen sich gut in die Rolle hineinversetzen können, und der Text kann aus der Ich- oder Du-Perspektive geschrieben sein. Der Tonfall soll ernst, aber auch spielerisch sein. Der Songtext soll emotionale Themen wie Liebe, Nostalgie und Freiheit betonen. Verwende viele Reime, auch innerhalb der Zeilen, und andere rhetorische Mittel, um einen fließenden und eingängigen Text zu erstellen. Der Songtext sollte ohne Schimpfwörter auskommen und einen starken Flow haben. Die Struktur soll Chorus, Vers, Chorus, Vers, Chorus, Bridge, Chorus, Outro umfassen.“<sup>89</sup>*

#### **4.1.2 Songtexterstellung mit ChatGPT**

Der erzeugte Prompt wird nun in ChatGPT eingegeben, um uns einen Songtext zu erstellen. Um das Ergebnis genauer analysieren zu können, bzw. um besser einschätzen zu können, wie die KI mit dem Prompt umgeht, wird derselbe Prompt zwei weitere Male eingegeben. Es sind nun drei verschiedene Texte entstanden. (siehe Anhang 1: ChatGPT Prompt 1).

Als groben Anhaltspunkt und Vergleichsbasis wurden im Anhang die Texte einiger bekannter Künstler hinzugefügt. Die Songs „I like the way you kissed me“ von Artemas (s. Anhang 2), „R U Mine“ von den Arctic Monkeys (s. Anhang 3), „Pumped up kicks“ von

---

<sup>89</sup> Chatgpt.com (2024).

Foster the People (s. Anhang 4) und „Don't“ von Ed Sheeran (s. Anhang 5) sind oder waren alle in den Charts.<sup>90</sup>

Zuallererst fällt auf, dass verglichen zu den Songs aus den Charts bei allen drei Chat-GPT-Texten sehr viele Wörter bzw. Zeilen vorhanden sind und diese Menge pro Zeile auch relativ konstant bleibt. Folglich bedeutet das, dass es weniger Abwechslung gibt und man den Text entweder schneller singen muss oder die Taktanzahl der einzelnen Parts verlängern muss, um die Wörter alle unterbringen zu können. Letzteres könnte den Song sehr in die Länge ziehen, ersteres könnte dagegen auf Kosten der Melodie gehen und ihn mehr zu einer Art Rap-Song machen.

In Ed Sheerans Don't wechseln sich gerappte Verse mit dem melodiös gesungenen Refrain ab. Hier kann man eindeutig erkennen, welchen Effekt das auf den Text und die Menge an Wörtern hat (s. Anhang 5).<sup>91</sup>

Das Reimschema ist bei allen drei generierten Texten sehr einfach gehalten. Es handelt es sich so gut wie ausschließlich um Paarreime. Paarreime sind zwar auch bei den Songs aus den Charts vorhanden, jedoch wird immer wieder daraus ausgebrochen oder andere komplexere Reimschemata verwendet. Unreine Reime sind ebenfalls in den Chart-Songs wesentlich öfter vertreten:

*“I'm a puppet on a string*

*Tracy Island (A), time-travelin' diamond (A) cutter-shaped heartaches (B)*

*Come to find you four in some velvet mornin' years too late (B)*

*She's a silver linin' (A), lone ranger ridin' (A) through an open space (B)*

*In my mind (C), when she's not right (C) there beside me (...)” – Arctic Monkeys*

Inhaltlich gesehen hat sich die KI gut an den erstellten Prompt gehalten. Alle drei Texte handeln von Liebe, Freiheit, Nostalgie und beinhalten Elemente wie ein Lagerfeuer am Strand oder die Wellen des Meeres. Die Bridge stellt dabei einen Bruch in der Story dar, in der der Protagonist sich in schweren Zeiten wieder an diese schöne Zeit zurückerinnert und daraus Kraft schöpft. (s. Anhang 1)

---

<sup>90</sup> Vgl. O.V. Officialcharts.com (2024).

*“And when the world feels heavy, and the night is too long,  
I’ll close my eyes and remember, the place where we belong.”<sup>92</sup>*

Immer wieder wird in sehr bildlicher Sprache, Metaphern und Vergleichen gesprochen:

*“Under the stars, where the waves kiss the sand,  
We’re chasing the night with our hearts in our hands.”<sup>93</sup>*

*“Time’s flown like a seagull, soaring high and wide,  
But those moments we shared are never left behind.”<sup>94</sup>*

Dennoch sind ein paar weitere Unterschiede zu den Songs aus den Charts festzuhalten. Bei den Songs aus den Charts gibt es häufige textliche Wiederholungen, gerade im Refrain.

*“I like the way you –, uh  
I like the way you kiss me, I can tell you miss me  
I can tell it hits, hits, hits, hits  
Not tryna be romantic, I’ll hit it from the back  
Just so you don’t get attached (yes, yes, yes)” - Artemas*

Ebenso wirken manche der von ChatGPT generierten Passagen auf unechte Weise gefühlvoll und rührselig-sentimental, wodurch sich die Frage nach Authentizität stellt.

*“The breeze whispered secrets, tales of old and new,  
With every crashing wave, our dreams were coming true.”<sup>95</sup>*

Im Folgenden wird versucht durch einen neuen komplett selbst erstellten Prompt ein Ergebnis zu erzielen, dass einem Text der Chart Songs näherkommt. Es soll unterschiedliche, komplexere Reimschemas geben, weniger Text, mehr Authentizität. Auf Wörter wie „viele Reime“ oder „Flow“ wird im Prompt verzichtet, da das eventuell zu einem Rap-ähnlicherem Text führen könnte. Die Story wird dahingehend verändert, dass es kein Happy-End gibt.

---

<sup>92</sup> Chatgpt.com (2024).

<sup>93</sup> Ebd.

<sup>94</sup> Ebd.

<sup>95</sup> Ebd.

Neuer Prompt:

„Schreibe einen Englischen Text über einen Protagonisten, der am Strand entlang wandert in der Ich-Perspektive. Er erinnert sich an die schöne Zeit, die er hier hatte, mit seinen Eltern in seiner Kindheit, seiner Freundin in seiner Jugend. Doch nun ist er allein und schwelgt in der Vergangenheit. Der Protagonist soll ab der Bridge traurig darüber sein, dass diese Zeit vorbei ist. Er wünscht sich die Zeit zurück und geht im Outro ins Kalte Wasser, um sich davon treiben zu lassen. Der Text soll gut zu singen sein. Achte darauf, dass ich später eine eingängige Melodie für den Text schreiben muss. Im Refrain können sich Phrasen oder Wörter wiederholen, damit die Melodie besser ins Ohr geht. Auch sollten im Refrain weniger Silben als im Vers benutzt werden. Benutze keine kitschigen Metaphern, der Text muss authentisch sein. Benutze komplexe Reimschemas wie ABAABABA oder ähnliches im Vers. Die Struktur soll Vers, Chorus, Vers, Chorus, Bridge, Chorus, Outro umfassen.“

Output von ChatGPT:

**Title: "Waves of Yesterday"**

**Verse 1:** *Walking down this sandy shore, where the past comes flooding back,  
Footprints from my childhood days, in the warmth of summer's track.*

*I recall the laughter, and the sun on our faces bright,  
Moments with my parents, feeling so right.*

*Days of building castles, dreams of endless summer skies,  
Now the echoes linger here, where the sea meets my eyes.*

**Chorus:** *Waves of yesterday, waves of yesterday,  
Where the past and present play, where the memories stay.*

*Waves of yesterday, waves of yesterday,*

*In the stillness, I replay, wishing I could stay.*

**Verse 2:** *With her in my youth, walking hand in hand by the sea,*

*We talked of dreams and futures, and the world was ours to be.*

*Now the tide brings back those whispers, in the quiet of the night,*

*I'm left with fading echoes, of a love that felt so right.*

*The sand beneath my fingers, cool and smooth as I recall,*

*Moments lost in time, and the memories that fall.*

**Chorus:** *Waves of yesterday, waves of yesterday,  
Where the past and present play, where the memories stay.*

*Waves of yesterday, waves of yesterday,*

*In the stillness, I replay, wishing I could stay.*

**Bridge:** *Now the shore is empty, and the night feels so cold,  
I wish I could return to the warmth of days untold.*

*The laughter's turned to silence, and the warmth to a distant dream,  
All I have are these old echoes, drifting down the stream.*

**Chorus:** *Waves of yesterday, waves of yesterday,  
Where the past and present play, where the memories stay.  
Waves of yesterday, waves of yesterday,  
In the stillness, I replay, wishing I could stay.*

**Outro:** *Stepping into the chilly sea, letting go of yesterday,  
Drifting with the cold waves, hoping they'll wash the past away.  
Waves of yesterday, waves of yesterday,  
In the cold embrace, I sway, as the memories fade away.<sup>96</sup>*

Inhaltlich hat ChatGPT die Geschichte sehr gut umgesetzt und in die Struktur des Songs integriert. Während im ersten Vers noch über die Zeit mit den Eltern und die Kindheit gesprochen wird, geht es im zweiten Vers um eine Liebe in der Jugend. Verbunden wird das Ganze durch die Refrains, in denen Phrasen wie gewollt wiederholt werden. Auch Bridge und Outro runden den Songtext gut ab. In letzterem wird sogar wieder ein Teil des Refrains aufgegriffen. Der Text wirkt durch die veränderte Story etwas authentischer.

Zur Melodiefindung kann man sagen, dass die unterschiedliche Silbenanzahl der Zeilen im Vers diese auf jeden Fall erschwert. Besonders in der dritten und vierten Zeile müssten wahrscheinlich mehrere Noten auf bestimmte Vokale gesetzt werden, um eine einheitliche Melodie zu ermöglichen.

Anweisungen bezüglich des Reimschemas wurden nicht eingehalten. Es handelt sich durchgehend wieder um Paarreime. Im Refrain sind nun aber bemerkenswerterweise anders als zuvor auch innerhalb der Zeilen Reime zu finden.

In weiteren Tests konnte dennoch festgestellt werden, dass ChatGPT einige Schwierigkeiten damit hat, Prompts bezüglich Reimschemas und ähnlichem einwandfrei zu befolgen. In folgendem Versuch wurde der Prompt fast gar nicht befolgt:

Prompt:

*Benutze dieses Reimschema im Vers: ABBABAB, benutze auch unreine Reime.*

---

<sup>96</sup> Chatgpt.com (2024).

Output:

**Verse 1:** *Walking down this haunted coast,  
Footprints fade like distant dreams,  
Where the echoes are my ghosts.  
Whispers carried by the streams,  
Stars above in silent flight,  
Moonlight's glow and silent cries,  
Mocking every sleepless night,  
Reflecting secrets in your eyes.*<sup>97</sup>

#### 4.1.3 Texterstellung zu einer Melodie

In diesem Teil soll betrachtet werden, wie ChatGPT im Songwriting Prozess helfen kann, wenn zuerst eine Melodie gegeben ist. ChatGPT muss also passend zu den Noten bzw. einer bestimmten Silbenanzahl einen Text generieren.

Im neuesten Modell ChatGPT 4o ist es möglich Audios hochzuladen und dazu Prompts zu schreiben. Im Folgenden wird die Melodie aus Abbildung 1 ohne Text auf den Laut „da“ eingesungen, exportiert und in ChatGPT als mp3-Datei hochgeladen. Der dazu erstellte Prompt lautet folgendermaßen:

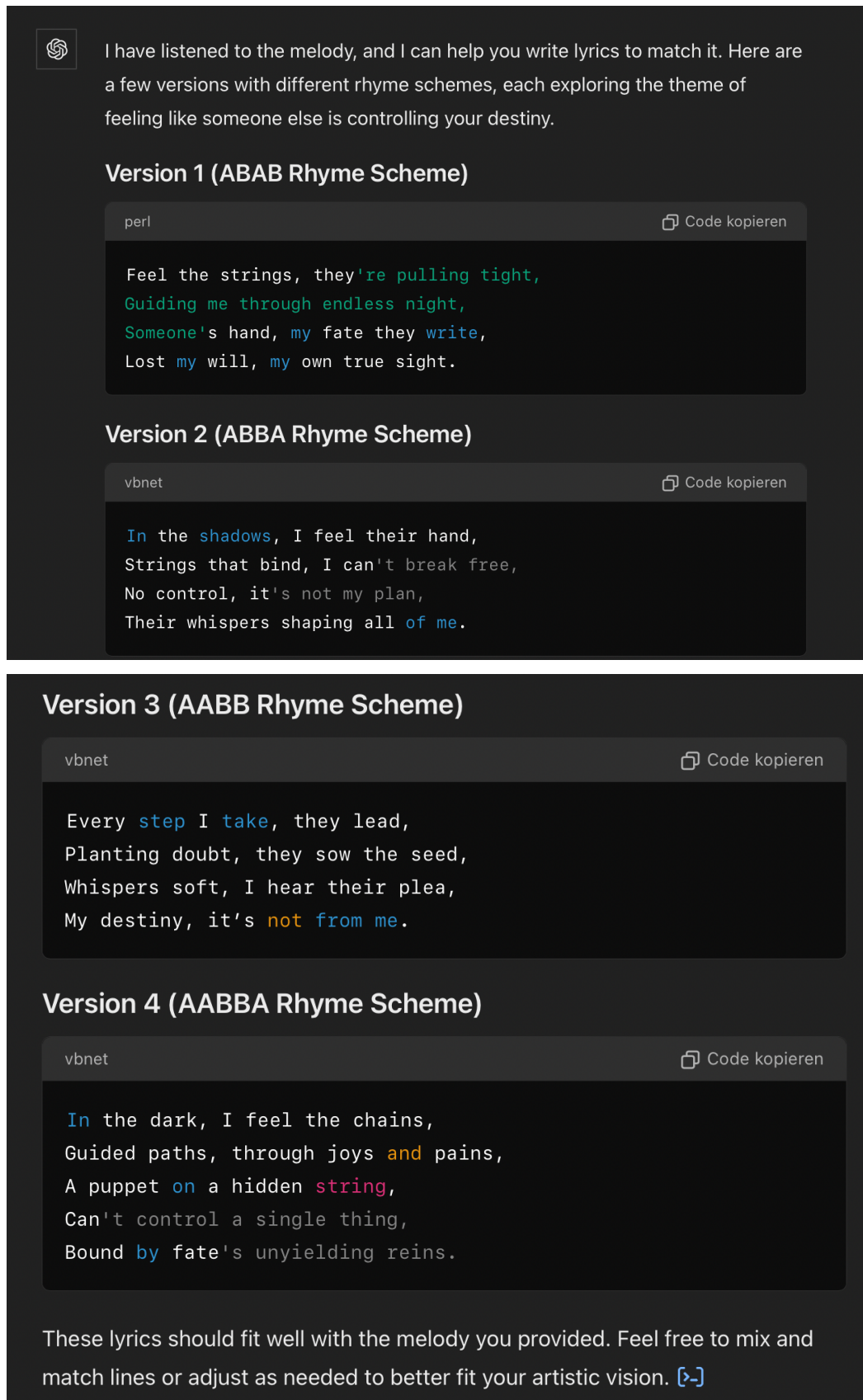
*„Can you help me to write lyrics for this verse melody? Every "da" which is sung is supposed to be one syllable. The lyrics should be about feeling like someone else is controlling my destiny. Can you give me some versions? Please try also different rhyme schemes like ABAB or ABBA. It should be written down in 4 or 8 lines.“*


ChatGPT erzeugt daraufhin folgenden Text, wie man ihn in Abbildung 5 sieht:

---

<sup>97</sup> Chatgpt.com (2024).

Abbildung 5: ChatGPT Texterzeugung



 I have listened to the melody, and I can help you write lyrics to match it. Here are a few versions with different rhyme schemes, each exploring the theme of feeling like someone else is controlling your destiny.

### Version 1 (ABAB Rhyme Scheme)

```
perl Code kopieren
```

Feel the strings, they're pulling tight,  
Guiding me through endless night,  
Someone's hand, my fate they write,  
Lost my will, my own true sight.

### Version 2 (ABBA Rhyme Scheme)

```
vbnet Code kopieren
```

In the shadows, I feel their hand,  
Strings that bind, I can't break free,  
No control, it's not my plan,  
Their whispers shaping all of me.

### Version 3 (AABB Rhyme Scheme)

```
vbnet Code kopieren
```

Every step I take, they lead,  
Planting doubt, they sow the seed,  
Whispers soft, I hear their plea,  
My destiny, it's not from me.

### Version 4 (AABBA Rhyme Scheme)

```
vbnet Code kopieren
```

In the dark, I feel the chains,  
Guided paths, through joys and pains,  
A puppet on a hidden string,  
Can't control a single thing,  
Bound by fate's unyielding reins.

These lyrics should fit well with the melody you provided. Feel free to mix and match lines or adjust as needed to better fit your artistic vision. [↔]

Quelle: ChatGPT (2024).



ChatGPT hat es diesmal geschafft verschiedene Reimschemas zu erzeugen. Auch enthält der erzeugte Text zum Teil große Ähnlichkeiten zum Originaltext:

*„My Master is pulling the strings in a wrong direction*

*Have I ever had control over myself?*

*Disaster can't be avoided my life is reckless*

*Have a ticking time bomb in every cell” – Causing the Avalanche (Philipp Sonntag)*

ChatGPT war es allerdings nicht möglich die Silben anhand des beigefügten Audiomaterials zu erkennen und einen zur Melodie passenden Text zu schreiben. Es wird nun nochmals versucht ohne Input-Audio einen Verstext generieren zu lassen. Dabei wird angemerkt, dass ChatGPT die Silben zählen soll.

Prompt:

*Please try again and give me a text, that meets the following criteria:*

*Line 1 should have 14 syllables*

*Line 2 should have 11 syllables*

*Line 3 should have 13 syllables*

*Line 4 should have 10 syllables*

*Please write a “-“ between each syllable, if a word contains more than one. Write in brackets behind every line the number of syllables that line contains.*

*For example: My des-tin-y, it's not from me. (8 syllables)*

Wieder erzeugt ChatGPT vier Versionen. Da die Ergebnisse ähnliche Fehler aufweisen, wird vorwiegend nur auf die erste Version eingegangen:

*I-feel-their-grip-on-my-soul,-like-a-pup-pet-on-a-string (14 syllables)*

*My-mind-it's-no-long-er-free (11 syllables)*

*Con-trol-ling-my-path,they-de-cide-my-ev-er-y-thing (13 syllables)*

*This-is-not-meant-to-be (10 syllables)<sup>98</sup>*

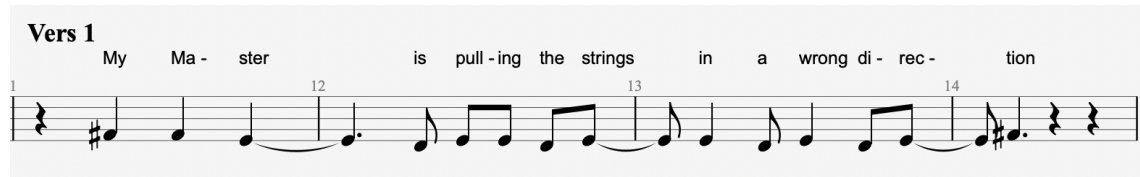
ChatGPT ist es gelungen die Wörter nach Silben zu trennen, doch war es der KI nicht möglich zuverlässig diese Silben zu zählen. Es macht sich zudem ein weiteres Problem bei der Textgenerierung bemerkbar. Die erste Zeile besteht zwar tatsächlich aus 14 Silben, allerdings passen die Betonungen der Wörter nicht zu den Betonungen der Melodie.

---

<sup>98</sup> Chatgpt.com (2024).

In der betrachteten Versmelodie sind alle Noten auf dem D beispielsweise unbetont, dementsprechend würde ein betontes Wort wie „grip“ an der Stelle nicht gut klingen.

#### Abbildung 6: Melodie mit Originaltext



Quelle: Eigene Abbildung erstellt mit Guitar Pro 7

Vice versa würde es seltsam klingen, wenn bei dem Wort „pup-pet“ oder der Wortkombination „soul-like“ auf der zweiten statt der ersten Silbe die Betonung liegt, was in unserem Beispiel bzw. mit dem Text von ChatGPT der Fall wäre.

In Version drei dieses Versuchs hat ChatGPT folgende Zeile generiert:

*“Feel-ing-trapped-in-this-world con-trolled-by-a-si-lent-com-mand. (13 syllables)”<sup>99</sup>*

Diese wurde zwar fälschlicherweise mit 13 statt mit 14 Silben erkannt und in Zeile drei gesetzt, doch passt hier die Silbenanzahl und die Betonung gut zum ersten Teil der Melodie.

In einem letzten Test mit ChatGPT werden zunächst einige Notizen zu einem unfertigen Song in die KI eingegeben (s. Anhang 6). Es fehlt hier der Text zu einem ersten Vers. Daraufhin wird folgender Prompt mit dem schon existierenden zweiten Vers eingegeben:

*„Can you give me a verse a 1 to this? Amount of syllables should be similar to this:*

*Sky is violet, where's my pilot*

*towards uncharted places*

*Thought I'm all set, make me forget*

*lies we tell to ourselves“*

---

<sup>99</sup> Chatgpt.com (2024).

Der Output von ChatGPT lautet folgendermaßen:

*Of course! Here's a Verse A1 with a similar syllable count:*

**Verse A1**

*Morning's golden, dreams embolden*

*Chasing hidden traces*

*Paths we wander, make me ponder*

*Truths we tell to ourselves*

*Feel free to tweak any part to better fit your melody or the theme of your song!<sup>100</sup>*

Nach der Bitte weitere Versionen zu erzeugen, kreierte ChatGPT die Verse in Anhang 7. Diesmal könnte man fast alle fünf Versionen, einschließlich der oberen verwenden. Manchmal sind dazu minimale Anpassungen nötig. Bei "*Chasing hidden traces*" müsste man z.B. das Wort *hidden* eine Silbe bzw. Achtelnote länger ziehen, damit die Melodie gleichbleibt.

Ab und zu hat die letzte Zeile eine Silbe zu viel, um sie auf die Melodie anpassen zu können (s. Anhang 7 Version 2 und Version 3).

#### **4.1.4 Zusammenfassung ChatGPT:**

Mit ChatGPT ist es durchaus möglich sinnvolle Songtexte zu erzeugen. Einzelne Zeilen wirken dabei manchmal etwas floskelhaft oder unauthentisch, andere jedoch sehr gut brauchbar. Mit einigem Ausprobieren kann man befriedigende Songtexte generieren lassen. Diese scheinen in Ihrer Komplexität jedoch begrenzt zu sein.

ChatGPT scheint besonders mit Reimschemas und Silben Probleme zu haben. Das macht die Songtextgenerierung zu einer bereits vorhandenen Melodie extrem schwer. Das neueste Model ChatGPT 4o konnte einige verschiedene Schemas anwenden und verschiedene Versionen erzeugen, die Silbenanzahl der Zeilen war allerdings meistens nicht wie vorgegeben. Somit waren die Generierungen zum Großteil unbrauchbar.

Eine weitere Erkenntnis ist, dass die Silbenanzahl und Reime nicht die einzigen wichtigen Parameter sind. Wortbetonungen sind mindestens genauso wichtig, um einen passenden Text zu einer Melodie zu schreiben.

ChatGPT kann auf diese eingehen, wenn ein Beispieltext oder Vers geschrieben wird. Zumindest war es der KI möglich, einen fehlenden ersten Vers eines Songs zu schreiben. Wenn man noch keinen Text zu Beginn hat, könnte ein möglicher Workaround so aussehen, dass man irgendeinen evtl. sinnlosen Text schreibt, bei dem jedoch

---

<sup>100</sup> Chatgpt.com (2024).

Reimschema und Silbenanzahl stimmen. ChatGPT könnte dann zu einem neuen Thema, das im Prompt erwähnt wird, einen neuen Text schreiben.

Fraglich ist, ob dies wirklich schneller ist als der normale Songwriting Prozess.

Festhalten kann man jedenfalls, dass man ChatGPT sehr gut zur Inspiration verwenden kann oder als zusätzlichen Hilfstool zu Seiten wie Rhymezone oder Ähnlichem. Die Möglichkeit sich viele Versionen generieren zu lassen befähigt den User dazu, sich das Beste herauszusuchen und ähnlich wie beim Comping so zu einer Endversion zu gelangen.

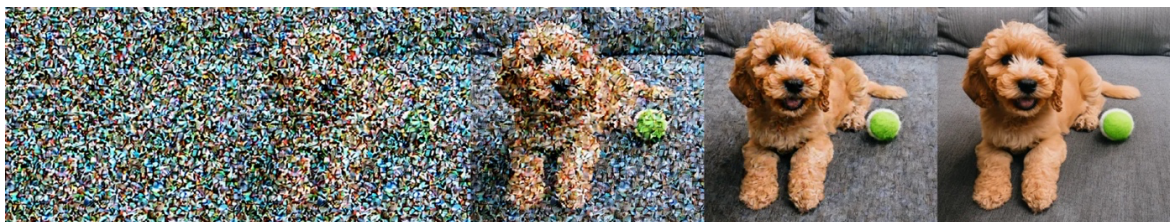
## 4.2 KI-Musikgeneratoren

KI-Musikgeneratoren sind dazu in der Lage eigenständig Musik oder Melodien auf Basis von Eingaben und Parametern zu erzeugen.<sup>101</sup> Die zwei Hauptmethoden, die dazu genutzt werden, sind die MIDI-Generierung und die Audio Wellenform Generierung. Ersteres benötigt sehr viel weniger Rechenleistung, da die generierten MIDI-Noten durch ein schon existierendes virtuelles Instrument durchlaufen und auf diese Weise ein Sound erzeugt wird. Die KI ersetzt also in diesem Fall nur den Schritt des Eintippens der Noten in eine Piano-Roll. AIVA und Seeds sind Beispiele für Tools, die auf diesem Prinzip basieren.<sup>102</sup>

Audio Wellenform Generierung ist in der Lage ohne externe Technologie Sound zu erzeugen. State-of-the-Art Modelle hierfür sind Riffusion, MusicGen, Stable Audio, AudioLM und noch weitere. Vieler dieser Modelle benutzen für die Generierung irgendeine Form von sogenannter Diffusion.<sup>103</sup>

Der Prozess der Diffusion wurde ursprünglich für die Generierung von Bildern entwickelt, kann jedoch auf Audio genauso angewandt werden. Der Begriff bezieht sich im Kontext von KI dabei auf den Prozess einem Signal Rauschen hinzuzufügen oder zu entfernen.<sup>104</sup>

**Abbildung 7: Bild Diffusion**



Quelle: Landschoot, C. (2024).

---

<sup>101</sup> Vgl. Kaufhold, B. (2024).

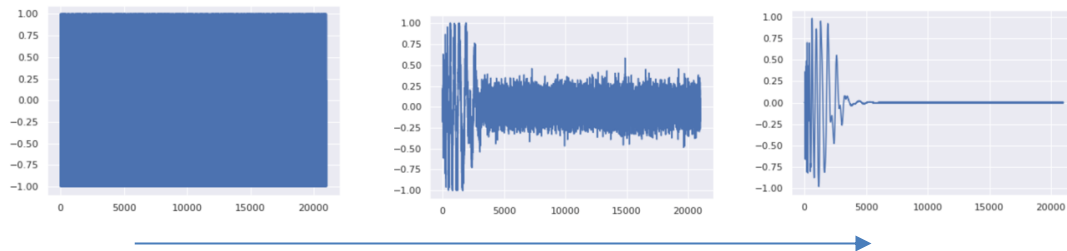
<sup>102</sup> Vgl. Landschoot, C. (2024).

<sup>103</sup> Vgl. Landschoot, C. (2024).

<sup>104</sup> Ebd.

Im Audiokontext wird mit White Noise angefangen und durch schrittweises Denoising so lange verändert, bis etwas erkennbares wie ein Song oder Sample entsteht.<sup>105</sup>

**Abbildung 8: Audio Diffusion**



Quelle: Eigene Darstellung nach Rouard, S. (o.D.).

Die nachfolgenden generativen Tools basieren vermutlich aus einer Kombination aus den schon erwähnten Transformer-/ Large Language Modells und einem Diffusion Modell wie z.B. Stable Audio.<sup>106</sup>

#### 4.2.1 Udio

Udio ermöglicht es den Benutzern komplette Songs anhand von einfachen Text Prompts zu erstellen. Themen, Genres und andere beschreibende Adjektive können eingegeben werden, um die Musik zu kreieren, welche der Benutzer hören möchte. Die Idee hinter Udio ist es, jedem mit einer Idee oder einem Text die Chance zu geben sich musikalisch auszudrücken.<sup>107</sup> Entwickelt wurde Udio von ehemaligen Google DeepMind Forschern mit dem Ziel Musik so leicht wie möglich zu generieren. Gerade mal 40 Sekunden oder weniger braucht die Website, um einen gemasterten Track zu produzieren.<sup>108</sup>

Das User Interface erinnert im ersten Moment dabei stark an das UI von Spotify. So werden im Discover-Reiter Staff Picks und trending KI-Songs vorgeschlagen, sowie die Top-Songs in verschiedenen Kategorien und Genres gezeigt. Den Erstellern der KI-Songs kann man folgen und somit die neuesten Kreationen der Prompt-Schreiber sehen und anhören. In dem extra Reiter „Liked Songs“ kann man seine Favoriten schließlich wiederfinden, die man entsprechend markiert hat. Interessant ist hier zu beobachten, dass die User von Udio ebenfalls von KI generierte Bilder als Songcover verwenden.

---

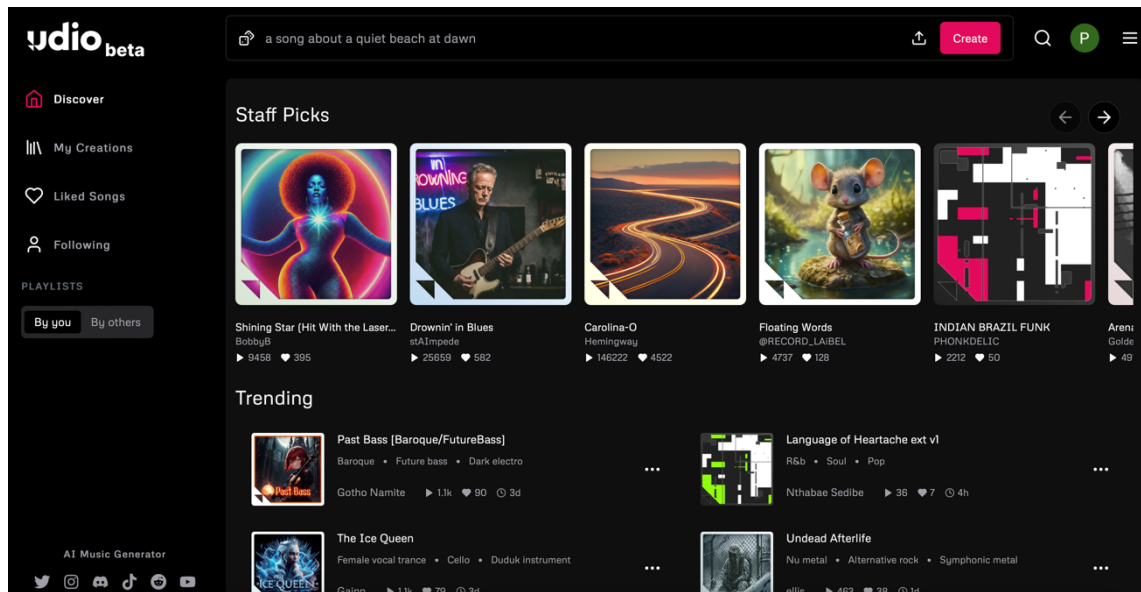
<sup>105</sup> Ebd.

<sup>106</sup> Vgl. Edwards, B. (2024); Rember, R. (2024).

<sup>107</sup> Vgl. O.V. *About*. [ud.io](https://ud.io) (2024).

<sup>108</sup> Vgl. Tencer, D. (2024).

Abbildung 9: Startseite Udio



Quelle: udio.com (2024).

In der oberen Leiste ist es nun möglich, eigene Prompts zu schreiben. Es öffnet sich ein Fenster in dem Tags (zum Großteil Genres) vorgeschlagen werden und man verschiedenen Parameter einstellen kann. Folgende Parameter lassen sich nun verändern:

Tabelle 3: Veränderbare Parameter im Creation Mode von Udio

Manual Mode	Im Manual Mode kann das Model ohne Prompt Rewriting verwendet werden. Standardmäßig verändert das System die eingegebenen Prompts, um die Durchschnittsqualität des Outputs zu verbessern. Ist der gewünschte Output sehr spezifisch, kann man in den Manual Mode wechseln und direkt mit dem Model Interface interagieren.
Lyrics	eigene Lyrics/ Instrumental / Automatisch generiert
Prompt Strenght	Einflussstärke der eingegebenen Prompts auf den Output; eine höhere Prompt Strenght sorgt für eine höhere Einhaltung der Prompts, könnte aber dazu führen, dass das Ergebnis weniger natürlich klingt.
Lyrics Strenght	Einflussstärke der eingegebenen Lyrics auf den Output; geringere Werte können natürlicher klingen, sorgen aber dafür, dass die Lyrics eventuell ignoriert werden.
Seed	Ein Wert, der den Inhalt des Clips definiert. Im Manual Mode sorgt der gleiche Seedwert bei gleichbleibenden Prompts immer für denselben Output.
Clip Start	Startpunkt des generierten Clips (für das Songverlängerungsfeature relevant)

Generation Quality	Schnellere Generierung ist möglich, wenn man im Gegenzug auf Qualität des Outputs verzichtet. Dieses Feature kann auch experimentell eingesetzt werden.
Context Length (nur bei Extensions)	Bestimmt, wie viel Sekunden des aktuellen Tracks für die Generierung des Outputs berücksichtigt werden. Eine längere Context length erhöht die Kohärenz des Outputs und führt zu konsistenteren Songparts. Kürzere Zeiten sind flexibler bezüglich Genre- oder Stilveränderungen.

Quelle: udio.com (2024)

Klickt man nun auf Create werden dem Benutzerkonto zwei Credits abgezogen und Udio erstellt zwei 32-sekündige Tracks, die in „My Creations“ abgespeichert werden. Hier ist es möglich, die Kreation mit der Community zu teilen, einen Remix zu erstellen oder den 32-sekündigen Clip zu verlängern. Auch in diesen Modi gibt es wiederum verschiedene Parameter, wie zum Beispiel im Remix Modus den Variance-Regler, der beschreibt, wie stark der Remix vom Original abweichen soll. Außerdem ist es möglich, den Songtext nochmal zu bearbeiten oder zu bestimmen, ob ein neuer Teil vor oder nach dem produzierten Clip angehängt werden soll. So lassen sich Clips zu vollen Songs aneinanderreihen oder bereits bestehende Songs aus der Community verlängern.

In nachfolgendem Beispiel wurde ein bereits existierender Song von Monz mit folgenden Prompts erweitert:

Tags: female, a cappella, opera

Extension Placement: Add Intro

Lyrics: Auto-generated

Prompt Strength: 100%

Lyrics Strength: 50%

Seed: -1

Clip Start: 0% (Beginning)

Generation Quality: Ultra

Context Length: 20s

Beispiel 3: 9-5 Blues ext V.1

[Hier anhören](#)

Ursprünglicher Ausgangsclip:

„9-5 Blues (Insane vocals I'm literally blown away!)“ von Monz

Tags: soulful blues, female vocalist, slow-tempo

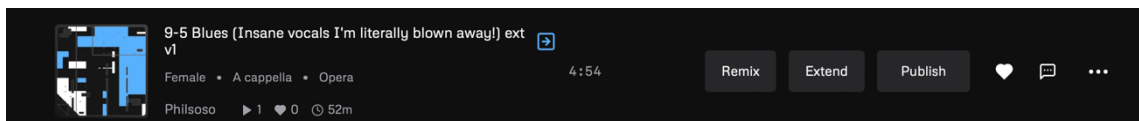
Länge: 4:22

Udio hat den Ausgangsclip um 33 Sekunden verlängert. Sowohl das hinzugefügte Intro als auch der bereits existierende Clip haben eine ausgesprochen hohe Qualität und zeigen das Potential von Udio. Es lässt sich durch das bloße Hören nicht erkennen, dass der Song KI-generiert ist. Sowohl die Stimmen als auch die Gitarrensoli hören sich auffallend menschlich und emotional an. Ebenso passt der neugenerierte Songtext sehr gut in den Kontext eines Blues.

*Weary tread from bed to floor  
Life is but a daily chore  
Sunrise paints, the sky so grand  
Yet my spirit feels so bland  
Yearning for a breath of ease  
In this symphony of pleas<sup>109</sup>*

Analysiert man genauer die ursprünglichen Prompts fällt auf, dass unter anderem das Tag *female vocals* darin enthalten ist. D.h. auch bei dem Clip von Monz (0:32 – 4:55), hat das Model einen anderen Output geliefert als geplant. Allerdings muss man hier dazu sagen, dass man nicht erkennen kann, wie gepromptet wurde (Manual Mode: Ja/Nein etc.) und ob Monz den kompletten Song allein generiert hat oder ebenfalls, wie hier in diesem Beispiel nur der Ausgangsclip verlängert wurde. So wird unsere verlängerte Version im Benutzerkonto unter „My Creations“ angezeigt mit der Möglichkeit diese unter unserem Benutzernamen zu veröffentlichen. Es werden die Tags *female*, *a cappella* und *opera* angezeigt, obwohl das nur die Tags waren, die für das neue Intro verwendet wurden. Rückschließend kann man bei Udio nicht erkennen, wie viele Menschen an einem kompletten Song gearbeitet haben und welche Tags für die einzelnen Songparts genutzt wurden.

#### Abbildung 10: Songs im Reiter „My Creations“



Quelle: udio.com (2024).

Die zweite verlängerte Version, die von Udio erstellt wurde, ist ebenfalls 4:55 min lang, beinhaltet jedoch im Intro anders als vorgegeben Instrumente. Gleichwohl ist hierbei anzumerken, dass man nur die ersten zehn Sekunden eine Gitarre hört, die danach äußerst

<sup>109</sup> udio.com (2024).



nahtlos und unauffällig durch eine Bassstimme ersetzt wird. Ansonsten hört man nur einen Jazzbesen, der einen Takt vorgibt. Rein musikalisch und lyrisch passt das erzeugte Intro wieder relativ gut zum Rest des Songs. Als Text wurde der Text aus dem Chorus erneut generiert, die Melodie im Intro ist jedoch eine andere.

Beispiel 4: 9-5 Blues ext V.2

[Hier anhören](#)

Ein weiteres Beispiel, welches die Flexibilität von Udio zeigt, ist der trending Song „African Groove Rhythm On Call“ von smithyommitte. Folgende Prompts wurden genutzt: afrobeat, afrobeats, afro house, afro trap, aftro, international afrobeat, amapiano, afro-pop, west african music, afrobeats - amapiano blend.

Beispiel 5: African Groove Rhythm On Call

[Hier anhören](#)

Hier hat der generierte Sänger sogar einen Akzent bzw. Slang, der passend zum Genre die richtige Stimmung vermittelt. Ein Blick auf den Songtext verrät, dass geschickt geprompted wurde, um das Ziel zu erreichen:

*[Verse]*

*Ah, feel the vibe wey dey scatter flow*

*Body move when the beat just go tun tun tun*

*When the groove hit, e be like say you dey fly solo*

*The rhythm sweet, catch the beat, no be so so*

*Hello hello...*

Wählt man nun einen Ansatz, der eher dem eines Producers oder Songwriters ähnelt, kann man einen eigenen Songtext in das Model eingeben. Udio schreibt in seinem Guide, dass das Modell verschiedene Sprachen, darunter Englisch, Deutsch, Japanisch, Chinesisch, Polnisch, Russisch, Französisch und Italienisch erkennen kann.<sup>110</sup> Für das Exempel wird anschließender Text aus „Slave or King“ verwendet:

---

<sup>110</sup> Vgl. O.V. *How do I make music with Udio?* udio.com (2024).

*[Chorus]*

*Don't feel my feet*

*I'm just drifting around*

*Let my Body float*

*Through the masses*

*Am I a slave or a king?*

*Screaming blood out of my throat – Causing the Avalanche (Philipp Sonntag)*

Im folgenden Versuch, wird neben dem Songtext Genre, Tonart und Tempo der KI vorgegeben. Es wird ebenfalls vorgegeben, dass der Text zu einem Refrain gehören soll. Um zu gewährleisten, dass auch wirklich ein Refrain entsteht, wird der Clip Start Parameter auf 40% gesetzt.

Folgende Prompts wurden für das nachfolgende Beispiel verwendet:

Tags: Modern Rock, Pop-punk, Rock, alternative rock, male vocalist, energetic, drums (drum set), guitar, 140bpm, B minor

Lyrics: siehe oben

Prompt Strength: 100%

Lyrics Strength: 100%

Seed: 4208 (randomize)

Clip Start: 40%

Generation Quality: Ultra

Beispiel 6: Slave or a King Text Prompt

[Hier anhören](#)

Das Ergebnis hier ist eher ernüchternd. Der gewünschte Output mit 140bpm und der Tonart B-moll wurde nicht geliefert. Die Stimme und die Komposition klingt etwas ungewöhnlich und nicht unbedingt massentauglich. Anzumerken ist jedoch, dass Udio in seinen Guidelines selbst dazu schreibt:

“No matter how detailed text prompts cannot fully define an actual piece of music - the same text describes an infinite number of possible audio tracks. That’s why to get close to the music idea you have in your head you might want to create multiple clips with identical prompts (and other settings).”<sup>111</sup>

---

<sup>111</sup> udio.com (2024).

Eine weiteres relativ neues Feature von Udio ist der sogenannte Audio Upload. Mit diesem ist es möglich, eigene Clips hochzuladen und diese durch die KI verlängern zu lassen.<sup>112</sup> Ob man dabei rhythmisch auf Haushaltsgegenstände schlägt, mit einem Instrument etwas einspielt oder etwas einsingt, ist dabei egal. Die KI kann den Audio Clip verlängern oder remixen. Es ist möglich, das Genre zu wechseln oder die Lyrics zu ändern. Dieses Feature verschafft dem Nutzer dementsprechend ein weitaus höheres Level an Flexibilität und kreativer Kontrolle.<sup>113</sup>

In folgendem Beispiel wird ein Teil des Songs „Slave or King“ in die KI hochgeladen und verlängert. Der Inputclip, welcher für das Vorgehen verwendet wird, ist folgender:

Beispiel 7: Slave or King Audio Upload Input

[Hier anhören](#)

Prompts:

Tag: keep the style

Lyrics Strength: 69%

Prompt Strength: 69%

Seed: -1

Generation Quality: Ultra

Beispiel 8: Slave or King Audio Upload extension

[Hier anhören](#)

Es wurden in diesem Versuch zwei Verlängerungen generiert, einmal vor und einmal nach dem Input Clip (0:33 – 0:50). Der Vers ist eingängig und der Übergang zum Input Clip ist in den Instrumenten kaum zu hören. Bei 0:27 wird sogar das Gitarrenriff aus dem Input Clip generiert. Sowohl bei der Stimme auch als in der Melodie hört man aber den Bruch. Der KI war es somit nicht möglich anhand des Inputclips die Stimme zu klonen und fortzuführen. Stattdessen scheint Udio eine ähnliche Stimme aus einem Pool von vorhandenen Stimmen auszuwählen. Nach 0:50 gibt es einige unerwartete Chord changes (0:58; 1:18) und es wird ein komplett neuer Teil generiert statt ein zum ersten Teil ähnlich klingender zweiter Vers.

---

<sup>112</sup> Vgl. Dredge, S. (2024).

<sup>113</sup> Vgl. Kabir, S. (2024).

Die Context Length wird nun so angepasst, dass nur noch der Refrain (0:07 – 0:17) berücksichtigt wird. Es wird eine Verlängerung generiert, die nach dem Input Clip hinzugefügt werden soll.

Beispiel 9: Slave or King Audio Upload extension 2

[Hier anhören](#)

Überraschenderweise wird in diesem Beispiel die hochgeladene Originalstimme mehrfach wieder aufgegriffen und wechselt sich mit der generierten Stimme ab. Zwar erkennt man auch hier den Bruch, dennoch kommt die KI-Stimme der Originalstimme etwas näher als im vorherigen Beispiel. Der Refrain wird mehrfach wiederholt und vor dem letzten Chorus wurde ein Aufbau eingefügt. Ab 0:34 singt die KI-Stimme schließlich terzversetzt zur originalen Lead-Vocal mit. Ebenso wurde eine neue Lead-Gitarrenstimme generiert, die den letzten Refrain sehr gut unterstützt. Aus Songwriter-Sicht ist dies ein äußerst ansprechendes Ergebnis.

Der Versuch wird wiederholt. Diesmal wird jedoch nur ein Vocal Snippet aus dem Song verwendet:

Beispiel 10: Slave or King just Vocals

[Hier anhören](#)

Als Tags werden die Wörter *solo* und *no instruments* eingegeben. Seed, Generation Quality und Prompt Strength bleiben gleich. Bei den Lyrics wird die Funktion „autogenerated“ angewählt.

Beispiel 11: Slave or King Vocal ext Text autogenerated

[Hier anhören](#)

Ab 0:17 beginnt der generierte Teil. Der autogenerierte Text hat nichts mit dem Text davor zu tun („I sing this melody just for you...uuh...Feel the notes I'm riding.“). Udio war es nicht möglich, den Text zu erkennen und ihn in passender Form fortzusetzen. Entgegen der Erwartung wurde er sogar zum Teil gesprochen. Die Tags *solo* und *no instruments* wurden zwar eingehalten, das Ergebnis ist aber recht unmusikalisch.

Für Beispiel neun wird der Text des ersten Verses als Verlängerung verwendet. An den Prompts wurden die Tags verändert: modern rock, post punk, hardcore, alternative rock Seed, Prompt und Lyrics Strength.

Beispiel 12: Slave or King Vocal ext mit Text

[Hier anhören](#)

Obwohl diesmal auf die Tags *no instruments* und *solo* verzichtet wurde, hat uns die KI trotzdem eine a-cappella-Spur generiert. Die KI-Stimme klingt nicht gleich, aber zumindest ähnlicher zur Stimme des Input Clips als in Beispiel 8. Sie bleibt in derselben Tonart und singt eine Melodie. Das Wort „Disaster“ wird mit viel Distortion in der Stimme und das Ende hingegen eher sanft, aber kratzig gesungen, wodurch sie emotionaler klingt.

#### 4.2.2 Suno

Udio AI ist nicht die einzige KI seiner Art. Ein direkter Konkurrent ist Suno, mit dessen Model sich ebenfalls auf Knopfdruck komplette Songs generieren lassen. Das Magazin Rolling Stone bezeichnete Suno sogar als ChatGPT für Musik.<sup>114</sup>

Funktionsmäßig steht es Udio in nichts nach. Auch hier lassen sich Tags und Lyrics eingeben, Audio hochladen und verlängern. Die Promptparameter sind nicht so detailliert wie in Udio, jedoch lassen sich statt 32 Sekunden Clips volle 2-4 Minuten Lieder direkt generieren. Um einen kurzen Einblick in dieses Tool zu geben, wurde nachfolgend ebenfalls der Text aus Beispiel 4 in Suno mit denselben Tags eingeben. Prompt Strength, Lyrics Strength, Seed, Clip Start und Generation Quality sind in Suno nicht einstellbar.<sup>115</sup>

Beispiel 13: Suno AI Slave or King Text Prompt

[Hier anhören](#)

Der Song der generiert wurde ist 2:33 lang und enthält keine neugenerierten Lyrics, d.h. der Text wiederholt sich wieder und wieder. Tonart und BPM wurden auch hier nicht eingehalten, weiterhin hört man Artefakte in der Stimme, wodurch sie relativ schnell als KI-Stimme zu erkennen ist. Die Songstruktur kommt der eines normalen Songs relativ nahe: Nach einem kurzen Drumfill startet es direkt in den Refrain, es folgt ein eingängiger Instrumental Hook, eine Art Vers, der wieder in den Refrain übergeht. Hiernach folgen eine Bridge und ein weiterer Refrain/Hook. Bei 1:50 scheint der Song vorbei zu sein, doch es wurde ein zweites Outro aus 2 verschiedenen Teilen generiert.

---

<sup>114</sup> Vgl. Fischer, B. (2024).

<sup>115</sup> Vgl. O.V. suno.com (2024).

Bessere Ergebnisse ergeben sich durch genaueres Prompting. Mit vollständigen Lyrics und Angaben in eckigen Klammern zur Struktur innerhalb des Songtextes scheint Suno besser umgehen zu können. (siehe Anhang 8: Songtexteingabe in Suno)

Style of Music: energetic, rock, alternative rock, post-punk, minor, male voice

Beispiel 14: Suno AI Slave or King Text Prompt vollständiger Song

[Hier anhören](#)

Die Instrumentals auch als die Gesangsparts sind sehr eingängig. Der Text in der Bridge wurde allerdings durch den Text vom zweiten Pre-Chorus ersetzt. Statt dem Refrain wurde die Bridge in einer entspannten, tieferen Lage gesungen und die Tonart wäre auch eher Cis-Dur statt Moll zuzuordnen. Für den Refrain hat die KI einen 12-Takter statt 16-Takter generiert, was dazu führt, dass der Refrain etwas unvollständig klingt.

Alles in allem ist die Songstruktur aber sehr gelungen und Suno hat die Erwartungen übertroffen. Feinheiten wie der Filter auf der Stimme in der ersten Hälfte des Verses, kleinere Pausen, um Spannung zu erzeugen oder der Wechsel der Drums von Half in Fulltime im letzten Refrain sind Entscheidungen, die auch ein ausgebildeter Producer eventuell getroffen hätte.

#### 4.2.3 Eleven Labs

Die Firma Eleven Labs hat angekündigt ebenfalls ein Musik-generierendes Tool ähnlich wie Suno und Udio zu veröffentlichen. Eleven Labs ist besonders für ihr ausgesprochen genaues Voice Cloning-Tool bekannt, baute aber in der letzten Zeit einige neue KI-Features in ihre Website ein wie zum Beispiel die Generierung von Sound Effekten. Bisher ist das Musikgenerierungstool nur für die Mitarbeiter von Elevenlabs zugänglich. Samples, die auf der Social Media Plattform X geteilt wurden, weisen dennoch darauf hin, dass das neue Tool qualitätsmäßig über viele Genres hinweg, Udio voraus ist.<sup>116</sup>

In folgendem Versuch wird das Soundeffekt Tool benutzt, um einen Vocal Chop, also eine kurze Gesangsphrase, und Ad-Libs, zu generieren. Die Prompts, die dafür eingegeben wurden sind:

“Vocal Chop: vocal chop, female voice, key: d-minor, sopran” und “ad-lib aggressive male”.

---

<sup>116</sup> Vgl. Morrison, R. (2024)

Beide Male soll ein 3-sekündiger Audioclip generiert werden. Ähnlich wie bei Udio gibt es bei Eleven Labs eine Art Prompt Strength Slider, wobei hier 0% bedeutet, dass es die Eingabe kreativ umsetzt und 100%, dass die KI den Prompts folgt. Für den Versuch werden an dieser Stelle 69% eingegeben.

Beispiel 15: Eleven Labs Vocal Chop

Beispiel 16: Eleven Labs Ad-Lib

[Hier anhören](#)

Der generierte Vocal Chop ist sehr opernhafte und klassisch. Wahrscheinlich war das Wort *sopran* das Schlüsselwort, welches zu diesem Ergebnis führte.

Die vier aneinander geschnittenen Ad-Libs, die herausgegeben wurden, sind so gut wie unbrauchbar. Als Output wurde zum Teil unerkennbare Sprache, zum Teil die Prompts selbst als gesprochene Wörter generiert.

Beide Prompts werden angepasst, um ein Ergebnis zu generieren, welches für aktuelle Populärmusik passender ist. Die Ausgabelänge und der follow prompt/ creative-Slider wurden dafür nicht verändert.

Angepasste Prompts:

“Vocal Chop, female, key: d-minor, RnB, pop”

“ad-lib aggressive male shouting "what?!"”

Beispiel 17: Eleven Labs Vocal Chop 2

Beispiel 18: Eleven Labs Ad-Lib „What?!“

[Hier anhören](#)

Das Problem, das vorher bei den Ad-Libs aufgetreten ist, ist nun auch bei dem Vocal Chop vorhanden. Dennoch kommen wir dem gewünschten Ergebnis näher, denn der Stil entspricht schon etwas mehr dem, der generiert werden sollte.

Die erzeugten Shouts könnten als Ad-Libs verwendet werden. Allerdings clippen die generierten Audiofiles diesmal und können deswegen nicht flexibel eingesetzt werden. Im Kontext eines aggressiven Rapsongs könnte man die Clips dennoch sicherlich verwenden.

In einem letzten Versuch wird der Prompt für den Vocal Chop erneut verändert zu: „Vocal Chop, female singing „oh yeah“, d-minor, RnB, pop“.

Beispiel 19: Eleven Labs Vocal Chop 3 "Oh yeah"

[Hier anhören](#)

Stil und Sprache sind dieses Mal zufriedenstellend. Allerdings sind die Clips eher den Tonarten C-Moll und D-Dur zuzuordnen. Dazukommt, dass sie hier nun ebenfalls clippen.

#### 4.2.4 Fazit Musikgeneratoren

Zusammenfassend kann man sagen, dass man mit Udio und Suno mit den richtigen Prompts zu erstaunlichen Ergebnissen kommen kann. Die generierten Melodien und Songstrukturen sind zum Teil sehr gut. Es lassen sich ganze Songs mit diesen Tools generieren, doch ist es genauso möglich diese als Inspiration zu nutzen. Besonders mit dem Audio-Upload kann man schon aufgenommene Ideen erweitern und verändern. Im Pre-Production-Prozess oder bei Schreibblockaden könnte dies sehr nützlich sein und Abhilfe schaffen. Die Vocals weisen teils Artefakte auf, teils sind sie jedoch nicht als KI-Stimmen erkennbar und echten Menschenstimmen zum Verwechseln ähnlich. Suno und Udio war es nicht möglich die eigene Stimme zu klonen und neue Melodien damit zu erzeugen.

Eleven Labs ist für die Vocal Production auf dem jetzigen Stand nicht wirklich geeignet. Für Ad-Libs im Hintergrund könnte man damit Sounds produzieren, jedoch ist es für Vocal Chops o.Ä. sinnvoller auf Sample Libraries zuzugreifen und nach entsprechenden Files zu suchen. Der zukünftige Musikgenerator von Eleven Labs lässt jedoch viel versprechen. Falls ihr Voice-Cloning Tool integrierbar und auf Gesang anwendbar wäre, könnte Eleven Labs einen Vorsprung gegenüber Suno und Udio erlangen.

### 4.3 Singing Voice Synthesis mit ACE Studio

ACE Studio ist eine Software mit der es möglich ist per MIDI Lead Vocals, Harmonies, Backings und Chöre zu generieren. Mehr als 41 verschiedene, lizenzierte Stimmen lassen sich nach der Eingabe der MIDI Noten nutzen, um die eingegebene Melodie bzw. den eingegebenen Text zu singen. Dabei werden nur die Sprachen Englisch, Chinesisch und Japanisch unterstützt. ACE Studio wirbt ebenfalls damit weitere Parameter (Atmung, Luft, Falsetto, Spannung, Energie) verändern zu können, um so Emotionen besser ausdrücken zu können. Verschiedene Stimmen ist es außerdem möglich zu mixen und so neue Stimmfarben zu erzeugen. Ebenso ist es möglich gegen einen Aufpreis eine



Custom Voice, also ein eigenes KI-Stimmen Model zu erstellen, indem man Samples der zum Beispiel eigenen Stimme hochlädt.<sup>117</sup>

Technisch gesehen steckt hinter der ACE AI Voice Engine die sogenannte Gesangsstimm synthese (SVS = Singing Voice Synthesis). Die herkömmliche verkettenden SVS beruht auf statistischen Methoden und klassischem Programmieren. ACE nutzt ungleich dazu tiefe neuronale Netze, um realistisch klingende Stimmen zu erzeugen. Hierzu werden zahlreiche Gesangsproben eines Sängers analysiert, um die akustischen Eigenschaften der Stimme wie Tonhöhe, Aussprachemuster und Klangfarbe zu erfassen. Diese Merkmale werden dann genutzt, um neue Gesangsstimmen zu synthetisieren. Während des Trainings eines Stimmmodells werden zusätzlich zu den Audioaufnahmen der Zielstimme bzw. des Sängers Stimmen von verschiedenen Sängern aus der ACE Quellbibliothek genutzt. Das befähigt das resultierende Stimmmodell dazu Sprachen oder Stile zu synthetisieren, die nicht in den Aufnahmen des Sängers vorhanden waren.<sup>118</sup>

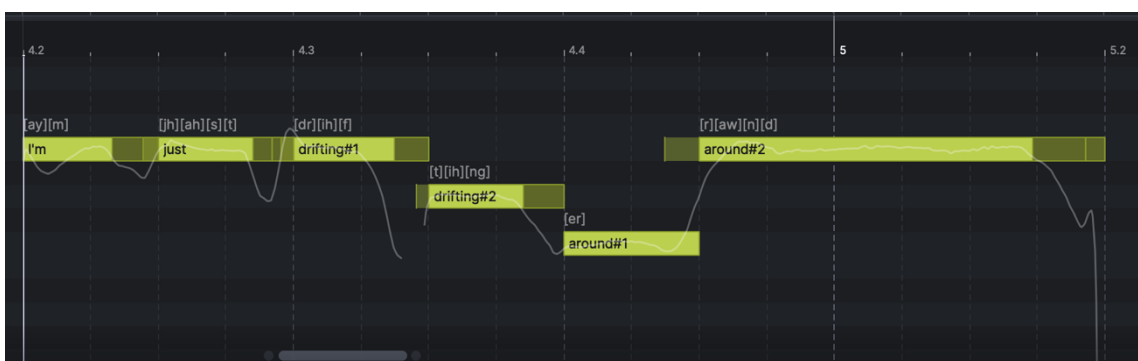
In folgendem Abschnitt wird ACE Studio genauer betrachtet und getestet.

#### 4.3.1 Analyse ACE Studio

ACE Studio ist ähnlich aufgebaut wie eine DAW. Es gibt verschiedene Arten von Spuren, ein Edit-Fenster, sowie ein Mixer Fenster mit dem die Grundparameter wie Pan und Lautstärke jeder Spur einzustellen sind.

Es gibt sogenannte „Singer Tracks“ die mit Instrumentenspuren in z.B. Pro Tools gleichzusetzen sind. In diese Singer Tracks kann man MIDI Clips einfügen und in einer Piano Roll wie gewohnt Noten einfügen, die dann von der KI-Stimme gesungen werden.

Abbildung 11: Piano Roll in ACE Studio



Quelle: ACE Studio (2024)

<sup>117</sup> Vgl. O.V. [cestudio.ai](https://www.cestudio.ai) *Pro AI Vocalist For All Your Vocal Creations*. (2024).

<sup>118</sup> Vgl. [support.cestudio.ai](https://support.cestudio.ai) (2024).

Anders als bei einer normalen Piano Roll lassen sich hier Wörter bzw. Silben in die eingefügten Noten schreiben. Dabei werden mehrsilbige Wörter aufgeteilt und mit einer Raute und einer Nummer versehen. Über den Noten werden die Wörter in Lautschrift angezeigt. Dies ermöglicht den genaueren Eingriff, falls Wörter von der Software nicht korrekt unterteilt oder ausgesprochen werden.

Neben den Singer Tracks gibt es normale Audio Tracks, in die man WAVs hochladen kann bzw. mit denen es möglich ist, etwas aufzunehmen. Dieser aufgenommene Audio Track lässt sich nun per Drag and Drop auf die MIDI-Spur ziehen und konvertieren. ACE Audio erkennt dabei den gesungenen Text und fügt ihn der Piano Roll automatisch hinzu. Im Control Panel jeder Spur lässt sich der Input bzw. bei den Singer Tracks die KI-Stimme/n auswählen und anpassen. Timbre und Style können unabhängig voneinander verändert werden.

In der Singer Library hat man die Auswahl zwischen den 41 verschiedenen KI-Stimmen die mit Tags wie #male, #female, #rock, #R&B #English #smooth etc. versehen sind, um sie schneller zu finden.

Die nachfolgende Abbildung zeigt das GUI von ACE Studio mit der Singer Library (ganz links), dem Track Control Panel (rechts davon), dem Mixer-/ Edit- Fenster (rechts oben) und der Piano Roll (rechts unten).

Abbildung 12: GUI ACE Studio



Quelle: ACE Studio (2024).

Im ersten Versuch mit ACE Studio wird per Hand jede Note einzeln eingegeben. Die Noten werden daraufhin markiert und mit einem Text versehen. Hierzu wird die Melodie und der Text des Songs „Slave or King“ verwendet:

[Chorus]

*Don't feel my feet*

*I'm just drifting around*

*Let my body float*

Schnell machen sich schon bei der Texteingabe einige Probleme bemerkbar. Wörter, die in der Melodie auf mehrere Noten aufgeteilt werden, werden nicht von der Software entsprechend aufgeteilt. Dem User bleibt nichts anderes übrig als phrasen- oder sogar wörterweise vorzugehen und teils in Lautschrift Verbesserungen durchzuführen.

Um die Möglichkeiten des Tools besser aufzuzeigen, wurden in diesem Beispiel einige weitere Parameter verändert. Breath, Air, Falsetto, Tension, Energy und Formant werden unter der Piano Roll angezeigt und sind für jede einzelne Note anpassbar. In diesem Beispiel wurde der Falsetto-Regler bei dem ersten Wort „Don't“ hochgezogen. Weiterhin wurde bei „feel my feet, I'm just drifting around“ die Tension und schließlich für die letzte Zeile das Energy-Level erhöht. Als KI-Stimme wurde „Naples“ verwendet, eine Stimme mit den Tags *rock, pop, powerful, gritty, searing, male* und *english*. Zuletzt wurden Einatmer vor jeder Phrase unter dem Reiter Breath hinzugefügt, um das Ergebnis menschlicher klingen zu lassen.

Beispiel 20: ACE Studio Texteingabe

[Hier anhören](#)

Trotz des Tags *english* scheint unsere KI-Stimme einen leichten Akzent zu besitzen. Wörter werden nicht vollständig oder nicht richtig ausgesprochen.

Wie schon vorher erwähnt, erweist sich die manuelle Texteingabe als zeitintensiv und problemreich. Deswegen wird im nächsten Beispiel ausgetestet, wie gut die Audio-konvertierung der Software funktioniert und ob sich dadurch eventuell auch das Akzentproblem beheben lässt. Als Input wird der Anfang von „Beispiel 7: Slave or King just Vocals“ herangezogen. Parameter wie Energy etc. werden diesmal nicht angepasst.

Beispiel 21: ACE Studio Audio to MIDI

[Hier anhören](#)

Durch die direkte Konvertierung kommt man um einiges schneller zum Ergebnis. Trotzdem scheint auch hier ein Akzent vorhanden zu sein und die Stimme klingt nicht ganz zufriedenstellend. Bei den Wörtern „around“ und „float“ wurde außerdem ein s

hinzugefügt. Das heißt auch hier muss man manuell in Lautschrift nachbessern, um zum gewünschten Ergebnis zu kommen. Einen echten Grit oder Distortion in die KI-Stimme zu bekommen ist nach mehrmaligen Versuchen, bei denen man die Energy und Tension der Noten verstellt hat, nicht gelungen.

#### 4.3.2 Voice Cloning mit ACE Studio

Statt eine der 41 von ACE Studio zur Verfügung gestellten Sänger zu benutzen, kann man auch einen Custom Sänger oder Sängerin erstellen, indem man auf der Website entsprechende Vocal Samples hochlädt. ACE legt Wert darauf, dass man nur Vocal-Dateien von Sängern hochlädt, deren Einverständnis man bekommen hat.<sup>119</sup> Einen Vertrag zwischen Sänger und Producer oder ähnliches muss man jedoch nicht hochladen. Ca. 30 Songs, die in einer angemessenen Aufnahmeumgebung entstanden sind, sollte man hochladen, damit die KI umfassend die Zielstimme nachbilden kann. Hierbei sollte es sich um dry Vocals handeln, ohne jegliche Noise in hoher Samplerate und nicht überlappend, um optimale Ergebnisse zu erzielen.<sup>120</sup> Leider liegt der Aufpreis bei mindestens 279USD, um dieses Feature zu benutzen, weswegen sich im Folgenden auf die Beispiele von dem Musiker und Youtuber Thomas Warburton bezogen wird, der für diese Bachelorarbeit zwei Audioclips zur Verfügung gestellt hat. Warburton hat Samples seiner eigenen Stimme hochgeladen und einen seiner Songs durch die KI-Version seiner selbst ersetzt. Im Folgenden hört man also einen kurzen Part eines vollproduzierten Songs, einmal mit der echten Stimme und einmal mit dem KI-Klon.<sup>121</sup>

Beispiel 22: Thomas Warburton echte Stimme

[Hier anhören](#)

Quelle: Warburton, T. (2024). Make YOUR voice sound AWESOME: Trying out ACE Custom Voice. 4:42 - 5:02

Beispiel 23: Thomas Warburton AI-Klon

[Hier anhören](#)

Quelle: Warburton, T. (2024). Make YOUR voice sound AWESOME: Trying out ACE Custom Voice. 5:04 - 5:22

Der Unterschied wird hier deutlich. Der KI mangelt es verglichen zur echten Stimme an Expression und Grit und sie wirkt auch hier um einiges softer.

---

<sup>119</sup> O.V. ACE Studio (2024).

<sup>120</sup> O.V. ACE Studio (2024).

<sup>121</sup> Vgl. Warburton, T. (2024). Make YOUR voice sound AWESOME: Trying out ACE Custom Voice.

Trotzdem konnte Warburton auch einige sehr gute Ergebnisse mithilfe von ACE Studio erzielen. Dabei geht er ähnlich wie bei einem normalen Vocal-Recording vor. Einzelne Phrasen werden erst in einer DAW nach Wahl aufgenommen und in unterschiedlichen Varianten eingesungen. Wichtig ist hier, dass die Takes in unterschiedlichen Weisen, d.h. mal normal, mal überbetont eingesungen werden, um ACE möglichst viel Material zu geben und die Wahrscheinlichkeit zu erhöhen, dass etwas verwertbares als Output generiert wird. Nachdem noch in der DAW die Vocals manuell getunt wurden, werden die verschiedenen Takes in ACE Studio in ein Projekt mit dem selben Tempo geladen und dort in MIDI konvertiert.<sup>122</sup>

In einem Video sagt er hierzu: „Because this tech isn’t perfect, there is still gonna be a lot of imperfections (...) but by giving it multiple takes, it increases the chances of getting just enough good material to frankenstein together one great take.“<sup>123</sup>

Nach einer ersten Hörprobe in ACE Studio wird nun entschieden, ob bestimmte Phrasen neu aufgenommen werden oder ob das Input Material ausreichend war und die KI-Takes exportiert werden können. Kleinere Aussprachefehler sind in diesem Schritt mit der manuellen Lautschrifteingabe eventuell korrigierbar. Nachdem die KI-Takes wieder in die DAW geladen wurden, folgt eine klassische Comping-Vorgehensweise. Warburton geht Wort für Wort durch die Takes und zieht das beste Material in einen Mastertake, der dann in der Produktion verwendet wird. Für den folgenden Clip wurde eine Custom Voice von Warburton verwendet.<sup>124</sup>

#### Beispiel 24: Thomas Warburton Produktion mit ACE Studio

[Hier anhören](#)

Quelle: Warburton, T. (2024). *How to make Your BAD Vocals GREAT with ACE Studio*.

4:08 – 4:21

Auch mit den von ACE Studio zur Verfügung gestellten KI-Stimmen lassen sich gute Ergebnisse erzielen:

#### Beispiel 25: Thomas Warburton Produktion mit ACE Studio weibliche KI-Stimme

[Hier anhören](#)

Quelle: Warburton, T. (2024). *Make YOUR voice sound AWESOME: Trying out ACE Custom Voice*. 5:43 – 6:00

---

<sup>122</sup> Vgl. Warburton, T. (2024). *How to make Your BAD Vocals GREAT with ACE Studio*.

<sup>123</sup> Warburton, T. (2024). *How to make Your BAD Vocals GREAT with ACE Studio*. (2:56 – 3:10)

<sup>124</sup> Vgl. Warburton, T. (2024). *How to make Your BAD Vocals GREAT with ACE Studio*.

### 4.3.3 Fazit ACE Studio

Sowohl bei Beispiel 16 als auch Beispiel 17 kann man nicht heraushören, dass es sich um eine KI-Stimme handelt. Gerade für poppige, weiche Stimmen gibt es gute KI-Modelle in der Software. Sind viele und gute Vocal Takes vorhanden, die in MIDI konvertiert werden können, bietet ACE Studio eine große Flexibilität. Für eine professionelle Produktion ist der Aufwand mit dieser Software jedoch mindestens genau so groß wie eine Produktion ohne KI.

Die reine Texteingabe ist aufwendig und aufgrund des Akzents höchstens für Demos interessant, um verschiedene Stimmen oder Harmonien auszutesten. Weiterhin weisen die KI-Stimmen von ACE Studio gerade im Bereich harsh/ extreme Vocals Defizite auf. So ist es schwierig auf einzelnen Noten eine Art Grit oder Distortion zu erzeugen, wie man es oft z.B. bei Rockmusik hört.

## 4.4 Audimee

Audimee ist eine Website, die damit wirbt Stimmen in die Stimmen anderer lizenzfreier Sänger konvertieren zu können. Des Weiteren ist es möglich lizenzfreie Gesangscovers zu erstellen oder das Modell auf die eigene Stimme zu trainieren.

Hinter Audimee steckt ein RVC Modell und eine Vielzahl an hochklassigen Studioaufnahmen, mit denen die KI trainiert wurde.<sup>125</sup>

RVC steht hierbei für Retrieval-based Voice Conversion. RVC ist eine Open-Source-Stimmwandlungstechnologie, die maschinelles Lernen nutzt, um ausgesprochen echt klingende und anpassbare künstliche Stimmen zu erzeugen.<sup>126</sup> Mit dem Aufkommen von RVC hat die KI-Stimmtechnologie besonders in den Bereichen Text-to-speech, voice cloning und Echtzeitstimmveränderung in den letzten Jahren einen großen Sprung gemacht. Plattformen wie GitHub oder HuggingFace haben bei dieser Entwicklung und vor allem der Distribution dieser Technologie eine wesentliche Rolle gespielt.<sup>127</sup>

Bei RVC wird zunächst der phonetische Inhalt, die Tonhöhe und die Dauer des Inputsignals bzw. des Eingangssprechers analysiert. Anschließend wird in der Datenbank des Zielsprechers nach passenden Segmenten gesucht und die extrahierten Merkmale verglichen. Als Output gibt das Modell eine konvertierte Stimme aus. Das Timbre, die

---

<sup>125</sup> Vgl. Audimee.com (2024).

<sup>126</sup> Vgl. Maria. (2024).

<sup>127</sup> Vgl. Weitzman, C. (2023).

Intonation und die stimmlichen Merkmale des Eingangssprechers werden dabei beibehalten.<sup>128</sup>

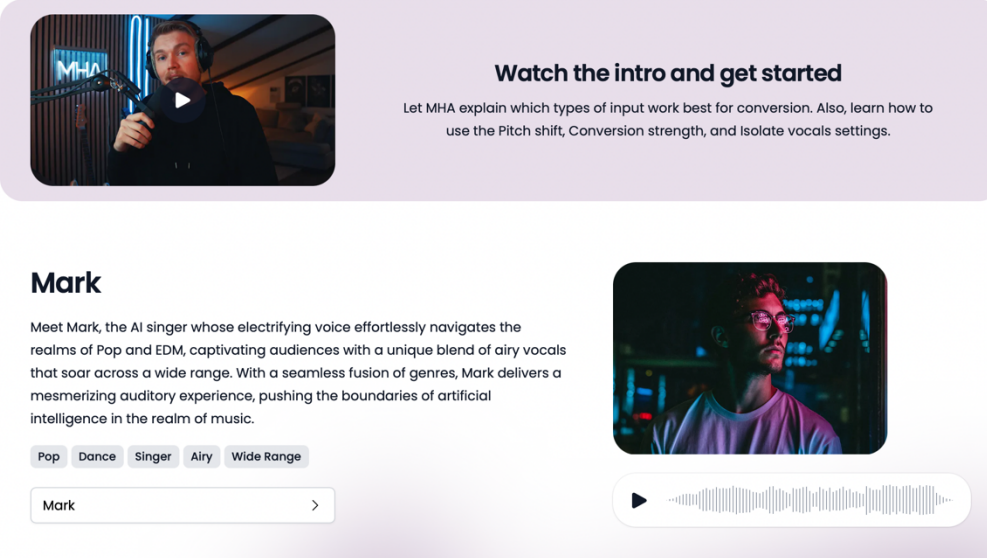
Im Folgenden wird ein RVC Modell via Audimee genutzt.

#### 4.4.1 Nutzung von Audimee

Nach dem Login wird man direkt zur Seite für die Audiokonvertierung weitergeleitet. Diese ist sehr übersichtlich aufgebaut und macht es dem Nutzer einfach eine Stimme zu konvertieren. Hierzu stehen verschiedene Zielsänger und deren ausführliche Beschreibung zur Verfügung.

Abbildung 13: Sängerbeschreibung Audimee

## Convert vocals



**Watch the intro and get started**

Let MHA explain which types of input work best for conversion. Also, learn how to use the Pitch shift, Conversion strength, and Isolate vocals settings.

### Mark

Meet Mark, the AI singer whose electrifying voice effortlessly navigates the realms of Pop and EDM, captivating audiences with a unique blend of airy vocals that soar across a wide range. With a seamless fusion of genres, Mark delivers a mesmerizing auditory experience, pushing the boundaries of artificial intelligence in the realm of music.

Pop Dance Singer Airy Wide Range

Mark >

Quelle: audimee.com (2024).

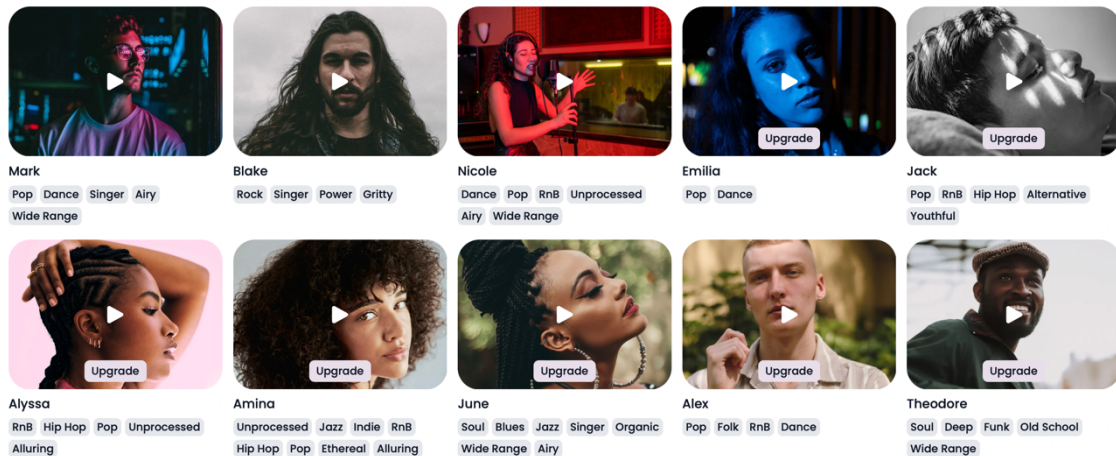
Neben der Beschreibung gibt es einen Audioclip des jeweiligen Sängers mit einem Audiobeispiel. Erfüllt dieser nicht die gewünschten Kriterien kann man zwischen 74 weiteren Sängern entscheiden und diese probieren.

---

<sup>128</sup> Vgl. Maria. (2024).

Abbildung 14: Sängerauswahl Audimee

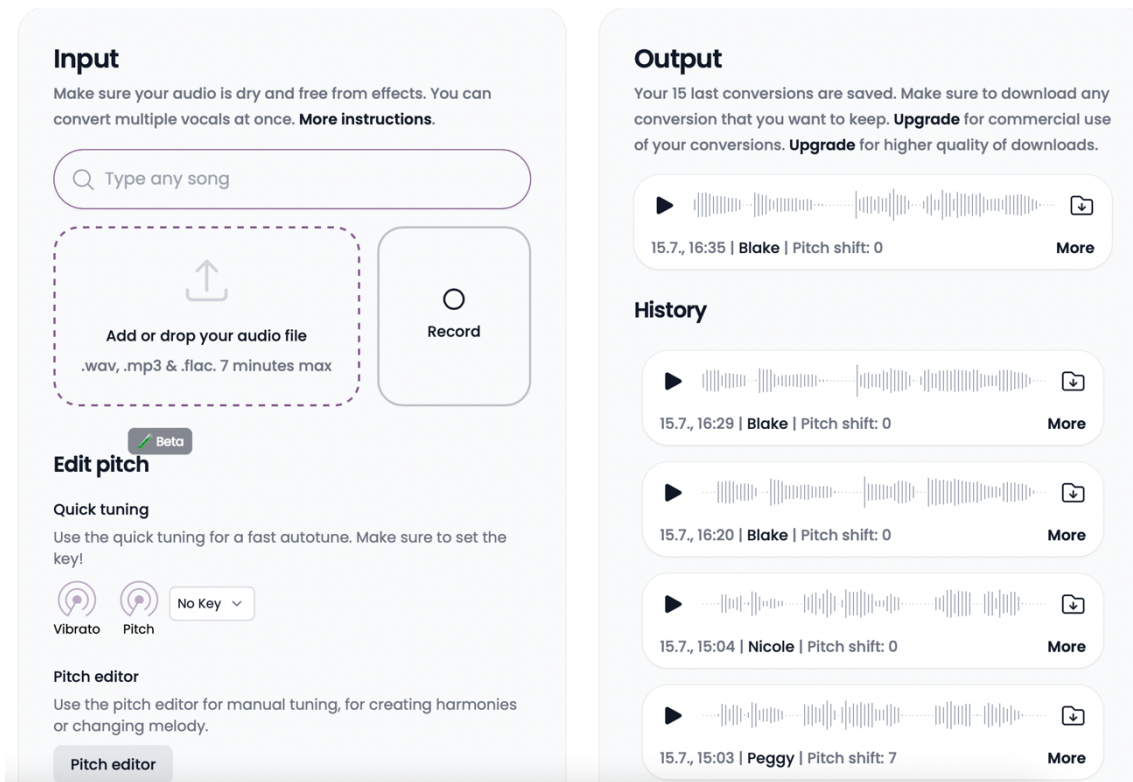
Featured royalty-free voices



Quelle: audimee.com (2024).

Die Sänger sind mit Tags markiert, dementsprechend ist es möglich nach verschiedenen Attributen wie Genres oder Stimmqualität zu filtern. Wurde eine Zielstimme ausgewählt lässt sich eine Audiofile aufnehmen oder hochladen. Diese kann man nun bei Bedarf manuell oder automatisch auf eine Tonart tunen und korrigieren.

Abbildung 15: GUI Audimee



Quelle: audimee.com (2024).



Neben dem Edit Pitch Fenster gibt es noch ein Setting Fenster, in dem ein Pitch-Shift-Slider (bis zu +/- 24 Halbtöne) und ein Conversion-Strength-Slider integriert sind. Letzterer steuert, wie sehr Artikulation und Charakteristika der Inputstimme verändert werden. Im Folgenden wird dieser Wert auf 30% belassen.

Weiterhin besteht die Möglichkeit Gesang aus dem hochgeladenen Track zu isolieren. Instrumente, Background Vocals, Reverb/ Echo und Rauschen kann man gesondert aus der Inputdatei isolieren, um so eine trockene Vocal-Spur zu generieren.

Im ersten Versuch mit Audimee wird nun der Clip aus Beispiel 9 hochgeladen, also die trockene Gesangspur von Slave or King. Als Zielsänger werden Mark, Blake und Nicole ausgewählt:

Beispiel 26: Audimee Mark

Beispiel 27: Audimee Blake

Beispiel 28: Audimee Nicole

[Hier anhören](#)

Marks Stimme klingt eher airy/ hauchig während Blake eine kratzige Rockstimme besitzt. Die Konversion in eine weibliche Stimme stellt für Audimee ebenfalls kein Problem dar, wobei die Aussprache beim letzten Wort hier nicht 100% korrekt ist. Abgesehen davon sind alle drei Konversionen äußerst gut gelungen und treffen gut auf die Beschreibungen der Sänger zu. Es sind keinerlei Artefakte hörbar.

Im nächsten Versuch wird eine zuvor mit ACE Studio erstellte Stimme in die KI hochgeladen und konvertiert.

Beispiel 29: Audimee ACE Input

Beispiel 30: Audimee Blake Output

[Hier anhören](#)

In diesem Fall kann der Output nicht ganz mit den vorherigen Outputs mithalten. Doch stellt das Ergebnis eine erhebliche Verbesserung zur Inputdatei dar.

Es soll nun untersucht werden, wie Audimee mit qualitativ schlechteren Inputdateien umgeht. Hierzu wird eine Stelle aus der Vokalspur der Demo von Slave or King hochgeladen, bei der einige Töne nicht in pitch sind. In Audimee wird im Bereich „Edit pitch“ die Tonart nun auf B min gestellt.

Beispiel 31: Audimee Vocal Snippet Demo Input

Beispiel 32: Audimee Vocal Snippet Demo Output

[Hier anhören](#)

Die automatische Stimmkorrektur von Audimee hat die Noten nicht auf die gewünschten Töne geschoben, allerdings muss man anmerken, dass die nicht getroffenen Noten relativ weit von der jeweiligen Wunschnote entfernt waren.

Das Ergebnis ist trotzdem äußerst ansprechend. Audimee hat die Qualität der Spur um ein Vielfaches erhöht. Die Stimme klingt präsenter, weniger dumpf und weniger nasal.

Als nächstes wird statt einer trockenen Vocal Spur ein Snippet eines vollen Songs hochgeladen. Audimee soll die Vocals isolieren und konvertieren. So entsteht ein Vocal Cover des Songs. Hierzu wird ein kurzes Snippet aus dem Song „Get You“ des Künstlers Daniel Ceasar verwendet. In den Einstellungen geben wir an, dass das Instrumental und der Reverb entfernt werden sollen.

Beispiel 33: Audimee Get You Input

Beispiel 34: Audimee Get You Blake Output

[Hier anhören](#)

Wie versprochen erhalten wir eine trockene Vocal-Spur des Songs, allerdings mit einem anderen Sänger. Die KI konnte das Instrumental erfolgreich von den Vocals trennen und diese konvertieren.

#### 4.4.2 Stem Separation

Nun wird ein kurzer Clip aus dem von uns generierten Suno Song in das Isolate Vocals Tool hochgeladen. Es handelt sich hier um eine Art Stem Separator.

Unter Stem Separation versteht man die Extrahierung einzelner Elemente wie Vocals, Bass, Drums etc. aus einem fertig produzierten Song. Auch diese Tools können KI gestützt arbeiten und reduzieren Kosten und Zeit gegenüber einer manuellen Separierung oder der Anheuerung von Session Musikern. KI Stem Separatoren werden mit einer großen Anzahl an Songs mit und ohne Vocals trainiert. So erkennt die KI die Charaktereigenschaften der menschlichen Stimme und kann diese von den restlichen Instrumenten trennen.<sup>129</sup>

---

<sup>129</sup> Wieduwilt, C. (2024).

Beispiel 35: Audimee Suno Song Input

Beispiel 36: Audimee Suno Song Isolated

[Hier anhören](#)

Das Ergebnis ist nicht zufriedenstellend. Immer wieder ist die Stimme kurz weg und man hört nun deutlich die Artefakte, die schon Suno verursacht hat. Interessant ist, dass auch bei der Konvertierung Audimee bei dem Suno Song einige Probleme hat. Es scheint so, als würde bei jeder Konversion zunächst der Stem Separator und danach das RVC-Modell arbeiten. Wie im vorherigen Beispiel ist die Stimme stellenweise weg oder weist Artefakte auf.

Beispiel 37: Audimee Suno Song Blake

[Hier anhören](#)

Das Isolierungstool wird mit dem Snippet aus „Get You“ von Daniel Caesar wiederholt.

Beispiel 38: Audimee Get You Isolated

[Hier anhören](#)

Die Isolierung der Vocals stellt hier kein Problem dar. Man kann demnach festhalten, dass die Stem Separation und die daraufhin folgende Konvertierung besser funktioniert, wenn die Input Datei hochwertiger ist und wenig bis keine Artefakte aufweist.

Hinzu kommt, dass Stem Separator oft besser arbeiten können, wenn weniger gleiche Frequenzen sich überlappen.<sup>130</sup>

Auch andere Stem Separator scheinen Probleme mit dem generierten Suno Song zu haben. Im Folgenden wird Gaudio Studio, ein kostenloser online Stem Separator, getestet.

Beispiel 39: Gaudio Suno Song Isolated

[Hier anhören](#)

Besonders im Refrain, wo mehr Frequenzen sich überlappen und es allgemein voller klingt, scheint die Stimme immer wieder etwas leiser und lauter zu werden.

---

<sup>130</sup> Vgl. Woods, B. (2024).

#### **4.4.3 Fazit Audimee und Stem Separation**

Audimee bzw. RVC und Stem Separation bieten besonders in Kombination extreme Flexibilität und Möglichkeiten. Ist die Inputdatei hochwertig und der Gesang leicht zu isolieren, kann problemlos eine Vocal-Cover-Spur erzeugt oder der Gesang in einem Song ersetzt werden. Insbesondere bei der Erstellung von Remixes könnten Stem Separation und RVC äußerst nützlich sein.

Weiterhin ist es möglich mit Audimee neben der Konversion in eine andere Stimme, die allgemeine Klangqualität einer Vocal-Spur zu verbessern. Auch dies erschließt neue Möglichkeiten und Chancen, worauf im nachfolgenden Teil näher eingegangen wird.

## 5 Chancen und Herausforderungen mit KI in der Produktion von Vocals

### 5.1 Chancen

Der Einsatz von KI in der Produktion von Vocals bietet mehrere Chancen und Möglichkeiten.

ChatGPT kann definitiv als Inspiration und Hilfstool genutzt werden. Die Versuche mithilfe der Silbenanzahl und Reimschemas Texte zu generieren, zeigt aber die Defizite von ChatGPT auf. Es besteht die Möglichkeit, dass durch besseres Prompting und Training von ChatGPT mithilfe von Beispielen es möglich ist, diese Probleme zu umgehen. Allerdings wurde ChatGPT nicht speziell für diesen Zweck entwickelt. Hier besteht eine Chance für Softwarehersteller und KI-Entwickler einen Algorithmus zu erstellen, der sowohl Reimschemas, Betonungen einzelner Wörter und Sinnhaftigkeit des gesamten Textes erkennen kann bzw. entsprechende Songtexte erzeugen kann.

Dies würde es möglich machen Texte zu bestehenden Melodien zu erzeugen.

Trotzdem muss man nochmal betonen, dass gute Texte ohne die Vorgabe von Silben etc. schon jetzt mit ChatGPT erstellbar sind.

Tools wie Suno und Udio besitzen das Potential die Musikproduktion zu revolutionieren und dieses Gebiet einer breiten Masse möglich zu machen. Momentan können diese KIs sehr gut als Inspiration dienen oder Dateien generieren, die als Demos definitiv nutzbar sind. Mit etwas mehr Aufwand und den richtigen Prompts ist es schon möglich, Songs zu produzieren, die ein Hörer nicht als KI-Song erkennen würde.

Die Entwicklung dieser Tools schreitet äußerst schnell voran. Während der Bearbeitung dieser Arbeit wurde erneut eine neue Version von Udio veröffentlicht.<sup>131</sup>

Es ist davon auszugehen, dass Artefakte in den Stimmen in den kommenden Versionen immer weniger zu hören sind, da dies schon jetzt teils der Fall ist.

In Kombination mit Stem Separatoren und RVC-Modellen erschließen sich in diesem Feld völlig neue Möglichkeiten. Man könnte so in Zukunft jede Stimme in jedem Song, auch in den generierten Songs von Suno, isolieren, ersetzen und neu mit dem Instrumental zusammenführen. Mit Audimee funktioniert dies schon relativ gut, sofern das Inputsignal gut isolierbar ist.

---

<sup>131</sup> Vgl. [udio.com](https://www.udio.com) (2024, aufgerufen am 1. August).

Stem Separation generell ist trotzdem noch ein herausforderndes Problem im Audioprozessing und bedarf weiterer Forschung, um das volle Potential zu erreichen.<sup>132</sup>

RVC ist neben Stem Separation auch in Kombination mit SVS möglich und führt zu besseren Ergebnissen als SVS allein. Dies ermöglicht beispielsweise eine Produktion von Musik, ohne einen Sänger engagieren zu müssen.

SVS weist jedoch Defizite auf, wenn es darum geht mit Emotion und nicht monoton zu singen.

RVC allein bietet die Chance jede Stimme durch eine Stimme aus einem scheinbar unendlichen Pool zu ersetzen. Demos können dadurch massiv aufgewertet werden. Für Sänger oder Artists könnte RVC ebenfalls sehr interessant sein, besonders wenn sie ein eigenes Stimmmodell damit trainieren. Entscheidend hierbei ist, dass das Trainingsmaterial sehr gut sein muss.

Ein neuer Workflow könnte darin bestehen, dass der Sänger einmal ins Studio geht, um etwas Material in sehr guter Aufnahmeumgebung und mit einem guten Mikro zu sammeln und damit dann seine eigene KI-Stimme zu trainieren.

In Zukunft kann er schließlich von zu Hause aus in schlechterer akustischer Umgebung mit schlechterem Mikrofon z.B. eine Demo oder einen Song aufnehmen. Nun füttert er das Modell mit dem Input und bekommt als Output eine Version die qualitativ eher der einer professionellen Studioaufnahme entspricht. Diese Methode könnte dem Sänger demnach Studiokosten und Zeit sparen. Beispiel 32 zeigt, dass dieses Vorgehen mit einem guten Modell jetzt schon möglich ist.

## 5.2 Herausforderungen

Wie in der Einleitung schon erwähnt, ist eine große Herausforderung der verantwortungsvolle Umgang mit KI. Die Sänger und Künstler, die in dem offenen Brief ihre Bedenken über die Zukunft der Musikindustrie und ihre Angst vor der Abwertung ihrer Arbeit äußern, sind sich gleichzeitig darüber einig, dass KI auch kreativ eingesetzt werden kann und viel Potenzial bietet. In dem Brief werden Rechtsgrundlagen gefordert, wie beispielsweise die Auszahlung von Tantiemen für den Fall, dass KI mittels ihrer Musik trainiert wird.<sup>133</sup>

---

<sup>132</sup> Vgl. Woods, B. (2024).

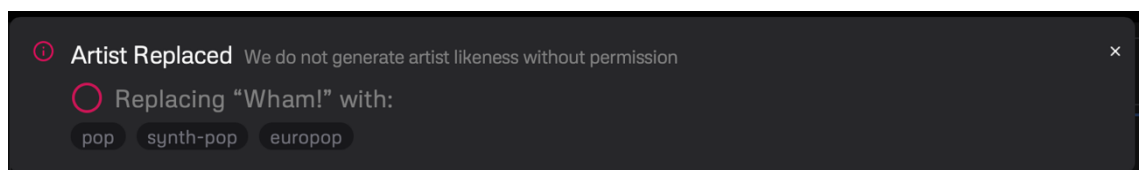
<sup>133</sup> Vgl. Jörg, D. (2024).

Auch Suno und Udio stehen unter Verdacht, dass sie massenhaft Musik als Trainingsmaterial für ihre KIs verwendet haben, ohne die Zustimmung der Rechteinhaber eingeholt zu haben. Die Recording Industry Association of America hat deswegen im Juni 2024 im Namen der größten Musiklabels, Universal, Warner und Sony Music dazu Klage eingereicht. Es wird in der Klageschrift darauf hingewiesen, dass die Vielzahl an KI generierten Songs, durch die Ähnlichkeit zu den Originalen mit diesen in Konkurrenz stehen und so den Künstlern und Labels Tantiemen streitig machen. Weiterhin sind in der Klageschrift beispielhaft verschiedene Songs aufgelistet, die sehr große Ähnlichkeiten zu Songs von Green Day, ABBA oder Mariah Carey haben. Ed Newton-Rex, ehemaliger Verantwortlicher der Audioabteilung des Startups Stability AI schlussfolgerte aus seinen Versuchen, in denen es ihm gelang mit den richtigen Prompts sehr ähnliche Songs zu generieren, dass sowohl Suno als auch Udio ihre KI mit entsprechenden Originalen gefüttert haben muss. Suno sagt dazu, dass das Tool transformativ sei und es nicht darauf ausgerichtet sei, das Material, mit dem die KI trainiert wurde, einfach wieder als Output zu liefern. Daher ist es auch nicht möglich als Prompt beispielsweise Künstlernamen oder Bands einzugeben.<sup>134</sup>

Bei der Nutzung von Suno wird nach der ersten Anmeldung folgender Text angezeigt: „Suno is designed for creating original music. Please confirm you will only submit AI-generated lyrics, original lyrics or lyrics to which you otherwise hold rights to continue.“<sup>135</sup>

Auch Udio sorgt entsprechend vor: In folgender Abbildung wurde „Wham“ und „Christmas“ als Prompt eingegeben.

**Abbildung 16: Anzeige bei Prompting mit Künstlernamen**



Quelle: udio.com (2024)

Bei dem Versuch mit dem Songtext des Weihnachtssongs ein KI-Cover zu erzeugen, wird die Meldung „copyrighted lyrics detected“ angezeigt und dementsprechend kein KI-Song generiert. In einem weiteren Versuch wird folgender überarbeiteter Text

---

<sup>134</sup> Vgl. Fischer, B. (2024).

<sup>135</sup> Vgl. O.V. suno.com (2024).

eingegeben und tatsächlich generiert die KI daraufhin zwei 32-sekündige Clips, wovon einer auch die Melodie des bekannten Wham-Songs zumindest am Anfang enthält:

*[Chorus]*

*Last X-mas, I gave you my hard*

*But the very nekst day, you gave idd away*

*That year, you safe me from tears*

*And give idd do someone special*

Beispiel 22: Last Christmas Text Prompt

[Hier anhören](#)

Arrangement, Tonlage und Text stimmen nicht mit dem Original überein, aber das Beispiel zeigt, dass die KI den Songtext erkennt und entsprechende Melodie daraus generiert. Auch lässt sich sagen, dass man die KI letztendlich durch geschicktes Prompting dazu zwingen kann, Clips zu generieren, die dem Trainingsmaterial nahekommen.

Auf die Frage, ob man Kreationen von Udio für die Musikdistribution, Werbung oder sonstiges monetarisieren kann, schreibt Udio:

“Yes, as long as the content does not contain copyrighted material that you do not own or have explicit permission to use, and as long as you properly indicate that the content was generated using Udio.”<sup>136</sup>

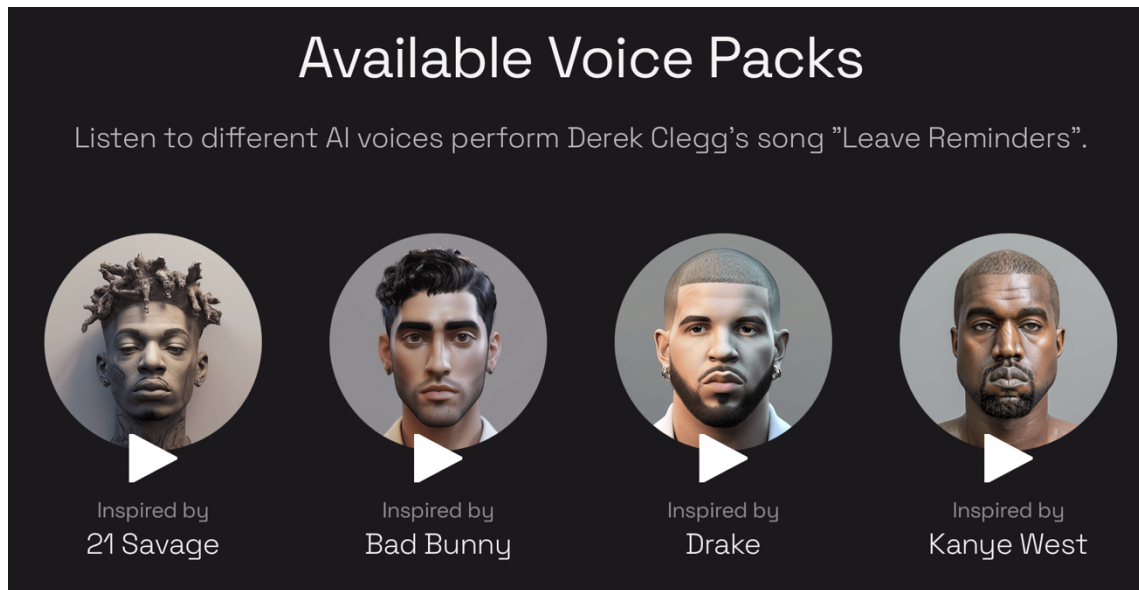
Ob sich Nutzer an derartige Abkommen halten oder nicht bleibt offen. Bezüglich RVC gibt es ähnliche kritische Bedenken. Websites werben zum Teil aktiv damit die eigene Stimme in die von bekannten Superstars oder Promis umwandeln zu können.<sup>137</sup>

---

<sup>136</sup> O.V. *How do I make music with Udio?* udio.com (2024).

<sup>137</sup> Vgl. lalal.ai (2024).



**Abbildung 17: Anzeige bei Prompting mit Künstlernamen**

Quelle: lalal.ai (2024)

Bei der Website [jammable.com](https://jammable.com) gibt es neben einigen Sängern sogar Modelle von Angela Merkel, dem Drachenlord oder SpongeBob.<sup>138</sup> Die RVC-Community ist riesig und es scheint von fast jeder bekannten Stimme ein Modell zu geben. 85,778 KI-Stimmen gibt es bereits auf der Seite [voice-models.com](https://voice-models.com).<sup>139</sup>

Dass Songs mit gestohlenen Stimmen viral gehen können, zeigt der etwas über zwei-minütige Song „Heart on My Sleeve“, welcher unter dem Pseudonym „Ghostwriter“ veröffentlicht wurde. Mit den Stimmen von Drake und The Weekend erreichte der Track auf YouTube 197000 Clicks in nur zwei Tagen. Universal Music bewirkte zwar nach kurzer Zeit, dass der Song von den Streamingdiensten und YouTube wieder runtergenommen wurde, doch es stellt sich dennoch die folgende Frage: Stehen die Stakeholder der Musikindustrie auf der Seite der Artists, Fans und der menschlichen Kreativität oder auf der der Deep Fakes und des Betrugs?<sup>140</sup>

KI in Musik ist ein kontroverses Thema und neben den oben genannten Problemen treten auch weitere Bedenken auf. Der englische Musikproducer Beardyman gab zu dem Teaser des neuen Eleven Labs-Musikgenerators folgendes Statement auf der Plattform X ab:<sup>141</sup> "It will be almost impossible to make a living in commercial music within 3 years. The only ways will be touring, patreon, bespoke composition, but even these will be

---

<sup>138</sup> Vgl. [jammable.com](https://jammable.com) (2024).

<sup>139</sup> Vgl. [voice-models.com](https://voice-models.com) (2024).

<sup>140</sup> Vgl. Willman, C. (2023).

<sup>141</sup> Vgl. Morrison, R. (2024).

greatly reduced and are already hard. Music is not just a job category, it is a way of expressing our spirit, harnessing our soul.. when it is this easy, what's the point in the artistic struggle? What does this mean for art? I'm so depressed about this... AI was supposed to come for the menial jobs first... but it's come for art first. We are not as special as we thought... art is a process... self expression is a therapeutic activity... but none of that matters now...<sup>142</sup>

---

<sup>142</sup> Vgl. Beardyman (2024).

## 6 Ausblick und Fazit

KI bietet Producern und Musikern die Möglichkeit ihren bisherigen Workflow von Grund auf zu überdenken. Von der Text-Ideengenerierung bis zur Erstellung eines kompletten Songs gibt es ausgesprochen nützliche Anwendungen.

ChatGPT kann als Inspirationsquelle und Hilfsmittel genutzt werden, auch wenn es beim Generieren von Texten anhand von Silbenanzahl und Reimschemas noch Defizite aufweist.

Musikgeneratoren wie Suno und Udio haben das Potenzial, die Musikproduktion zu revolutionieren. Teils sind die generierten Songs von diesen KIs von menschlichen Produktionen kaum unterscheidbar. Sowohl bei ChatGPT als auch bei diesen Tools ist gutes Prompting jedoch unabdingbar. Oft bekommt man einen zufriedenstellenden Output erst nach mehrmaligen Versuchen als Ergebnis eines Trail-and-Error Prozesses.

Die Kombination von KI-Tools wie Stem Separatoren und RVC-Modellen eröffnen kreative Möglichkeiten, wie die Isolation und Ersetzung von Stimmen in sowohl bestehenden als auch generierten Songs. Mit selbsttrainierten Modellen wäre es möglich die eigene Stimme in diese Songs einzusetzen. Ebenso kann man mit guten Trainingsdaten der Stimme schlecht klingende Demo-Vocals klanglich um ein Vielfaches aufwerten und sich Studiozeit sparen.

Trotzdem birgt die Nutzung von KI in der Musikproduktion erhebliche ethische als auch rechtliche Herausforderungen. Künstler äußern ihre Bedenken bezüglich der unkontrollierten Nutzung von KI und fordern klare Regelungen, insbesondere in Bezug auf die Nutzung von Trainingsmaterial und der Urheberrechte. Die Entwicklung schreitet äußerst schnell voran und es ist davon auszugehen, dass Musikgeneratoren in wenigen Jahren mit dem Eingeben eines einzelnen Prompts vollständige, radiotaugliche Songs generieren können. Es liegt an Entwicklern, Künstlern und der gesamten Musikindustrie, Wege zu finden, diese Technologien verantwortungsvoll und innovativ zu nutzen, ohne die menschliche Kreativität zu untergraben.

## Literaturverzeichnis

- Beardyman. (10. Mai 2024). Abgerufen am 1. August 2024 von x.com:  
[https://x.com/beardyman/status/1788827519812055075?ref\\_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1788827519812055075%7Ctwgr%5E701141c7f84a7a9d111da5b40db00893f0a89fba%7Ctwcon%5Es1\\_&ref\\_url=https%3A%2F%2Fwww.tomsguide.com%2Fai%2Felevenlabs-is-launch](https://x.com/beardyman/status/1788827519812055075?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1788827519812055075%7Ctwgr%5E701141c7f84a7a9d111da5b40db00893f0a89fba%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fwww.tomsguide.com%2Fai%2Felevenlabs-is-launch)
- Berens, A., & Bolk, C. (2023). *Content Creation mit KI*. Bonn: Rheinwerk Computing.
- Beshko. (9. November 2023). *AI Songs 2023 - Vocal Production mit künstlicher Intelligenz*. Abgerufen am 30. Juni 2024 von musikproduzentwerden.de:  
<https://musikproduzentwerden.de/blogs/news-angebote/ai-songs-covers-2023-kuenstliche-intelligenz-vocal-production>
- Bricio, C. (11. November 2023). *A Beginners Guide to Vocal Production*. Abgerufen am 12. Juli 2024 von abbeyroadinstitute.co.uk:  
<https://abbeyroadinstitute.co.uk/blog/a-beginners-guide-to-vocal-production/>
- Brown, O. (2021). Sociocultural and Design Perspectives on AI-Based Music Production: Why Do We Make Music and What Changes if AI Makes It For Us? In E. R. Miranda (Hrsg.), *Handbook of Artificial Intelligence for Music* (S. 1-20). Plymouth: Springer.
- Butler, G. (27. Juni 2023). *The Creative Process of Ed Sheeran: Authenticity and Innovation*. Abgerufen am 9. Juli 2024 von ownerartist.com:  
<https://www.ownerartist.com/journal/the-creative-process-of-ed-sheeran/>
- Clarke, L. (2023. Juni 2023). *Artificial Intelligence: The Future of the Music Industry?* Abgerufen am 19. Juni 2024 von recordingarts.com:  
<https://recordingarts.com/artificial-intelligence-the-future-of-the-music-industry/>
- Danesi, M. (23. März 2024). *AI-Generated Music*. Abgerufen am 20. Juni 2024 von link.springer.com: [https://link.springer.com: https://link.springer.com/chapter/10.1007/978-3-031-54752-2\\_4#citeas](https://link.springer.com/chapter/10.1007/978-3-031-54752-2_4#citeas)
- Demetriou, A., Jansson, A., Kumar, A., & Bittner, R. M. (September 2018). Vocals in Music Matter: The Relevance of Vocals in the Minds of Listeners. *ISMIR*, 514-520.
- Drapkin, A. (13. März 2024). *Google Gemini vs ChatGPT 2024: AI Chatbot Head-to-Head Test*. Abgerufen am 12. Juli 2024 von tech.co:  
<https://tech.co/news/google-bard-vs-chatgpt>
- Dredge, S. (14. Juni 2024). *AI music startups Suno and Udio both add audio uploads feature*. Abgerufen am 24. Juni 2024 von musically.com:  
<https://musically.com/2024/06/14/ai-music-startups-suno-and-udio-both-add-audio-uploads-feature/>
- Duden. (o.D). *Intelligenz*. Abgerufen am 9. Juli 2024 von Duden.de:  
<https://www.duden.de/rechtschreibung/Intelligenz>
- Duden. (o.D.). *künstlich*. Abgerufen am 9. Juli 2024 von Duden.de:  
<https://www.duden.de/rechtschreibung/kuenstlich>

- Dumont, L. (o.D.). *Ablauf einer Musikproduktion* Quelle: [www.delamar.de](http://www.delamar.de). Abgerufen am 9. Juli 2024 von [delamar.de](http://delamar.de):  
<https://www.delamar.de/musikproduktion/ablauf-musikproduktion-51401/>
- Edwards, B. (10. April 2024). *New AI music generator Udio synthesizes realistic music on demand*. Abgerufen am 30. Juli 2024 von [arstechnica.com](http://arstechnica.com):  
<https://arstechnica.com/information-technology/2024/04/new-ai-music-generator-udio-synthesizes-realistic-music-on-demand/>
- Emmerich, N. (10. Juli 2023). *Das solltest du über Prompting wissen*. Abgerufen am 21. Juli 2024 von [gew.de](http://gew.de): <https://www.gew.de/aktuelles/detailseite/das-solltest-du-ueber-prompting-wissen>
- Europäisches Parlament. (14. September 2020). *Was ist künstliche Intelligenz und wie wird sie genutzt?* Abgerufen am 14. Juli 2024 von [europarl.europa.eu](http://europarl.europa.eu):  
<https://www.europarl.europa.eu/topics/de/article/20200827STO85804/was-ist-kuenstliche-intelligenz-und-wie-wird-sie-genutzt>
- Fischer, B. (25. Juni 2024). *Musikriesen verklagen KI-Start-ups Udio und Suno*. Abgerufen am 24. Juni 2024 von [faz.net](http://faz.net):  
<https://www.faz.net/aktuell/wirtschaft/unternehmen/ki-start-ups-universal-warner-und-sony-verklagen-udio-und-suno-19812529.html>
- Fraunhofer Institut. (o.D.). *Künstliche Intelligenz (KI) und maschinelles Lernen*. Abgerufen am 14. Juli 2024 von [iks.fraunhofer.de](http://iks.fraunhofer.de):  
<https://www.iks.fraunhofer.de/de/themen/kuenstliche-intelligenz.html>
- Fresh, D. (18. Juni 2017). *Songwriting: Aufbau & Themenfindung*. Abgerufen am 11. Juli 2024 von [soundandrecording.de](http://soundandrecording.de):  
<https://www.soundandrecording.de/tutorials/songwriting-aufbau-themenfindung/>
- Goodman, E. (27. September 2018). *Matt Tuck: "I don't give a f\*ck what people want to hear, it's what I want to write"*. Abgerufen am 12. Juli 2024 von [loudersound.com](http://loudersound.com): <https://www.loudersound.com/features/matt-tuck-i-dont-give-a-fck-what-people-want-to-hear-its-what-i-want-to-write>
- Guo, R. (Juli 2023). *AI Pop Music Composition with Different Levels of Control: Theory and Application*. Sussex: University of Sussex.
- Hahn, M. (10. Juni 2024). *Vocal Production: The Ultimate Guide for Beginners*. Abgerufen am 11. Juli 2024 von [blog.landr.com](http://blog.landr.com): <https://blog.landr.com/vocal-production-guide/>
- Hughes, A. (25. September 2023). *ChatGPT: Everything you need to know about OpenAI's GPT-4 tool*. Abgerufen am 19. Juli 2024 von [sciencefocus.com](http://sciencefocus.com):  
<https://www.sciencefocus.com/future-technology/gpt-3>
- Jörg, D. (3. April 2024). *Offener Brief gegen Verwendung von KI in der Musikindustrie*. Abgerufen am 20. Juni 2024 von [SWR.de](http://SWR.de): <https://www.swr.de/swr1/rp/offener-brief-gegen-ki-musikstars-warnen-vor-missbrauch-100.html#:~:text=Die%20Warnung%2C%20die%20jetzt%20von,und%20letztlich%20auch%20ihre%20Identität.>
- Jones, A. (19. Oktober 2022). *Best pitch correction plugins and apps: Software to improve or enhance your vocal recordings*. Abgerufen am 15. Juli 2024 von

- musicradar.com: <https://www.musicradar.com/news/best-pitch-correction-plugins-and-apps>
- Kabir, S. (7. Juni 2024). *You Can Now Generate Music from Your Own Audio Uploads – Thanks to Udio AI*. Abgerufen am 28. Juni 2024 von nomusica.com: <https://nomusica.com/generate-music-from-your-audio/>
- Kaufhold, B. (11. April 2024). *KI-Musikgeneratoren: Die neuen Werkzeuge für Musiker?* Abgerufen am 30. Juli 2024 von ki-im-alltag.de: <https://www.ki-im-alltag.de/ki-musik-generieren/>
- Klawonn, T. (2023). *Künstliche Intelligenz, Musik und das Urheberrecht*. Tübingen: Mohr Siebeck.
- Kurz, P. (26. November 2021). *Yesterday – die Melodie kam im Traum*. Abgerufen am 7. Juli 2024 von wz.de: [https://www.wz.de/nrw/paul-mccartneys-song-biografie-lyrics\\_aid-64291979](https://www.wz.de/nrw/paul-mccartneys-song-biografie-lyrics_aid-64291979)
- Landschoot, C. (22. Januar 2024). *Audio Diffusion: Generative Music's Secret Sauce*. Abgerufen am 30. Juli 2024 von towardsdatascience.com: <https://towardsdatascience.com/audio-diffusion-generative-musics-secret-sauce-f625d0aca800>
- Maria. (19. Juli 2024). *RVC AI Voice: Was ist das und wie benutzt man es*. Abgerufen am 30. Juli 2024 von multimedia.easeus.com: <https://multimedia.easeus.com/de/voice-changer-tipps/rvc-ai-voice.html>
- Mockenhaupt, A. (2021). *Digitalisierung und Künstliche Intelligenz*. Sigmaringen: Springer.
- Morrison, R. (10. Mai 2024). *ElevenLabs is launching a new AI music generator — and you have to hear these clips to appreciate it*. Abgerufen am 28. Juni 2024 von tomguide.com: <https://www.tomsguide.com/ai/elevenlabs-is-launching-a-new-ai-music-generator-and-you-have-to-hear-these-clips-to-appreciate-it>
- o.D. (2024). *Audimee*. Abgerufen am 30. Juli 2024 von audimee.com: <https://audimee.com>
- O.V. (26. Juli 2024). Abgerufen am 26. Juli 2024 von officialcharts.com: <https://www.officialcharts.com>
- O.V. (17. Juni 2024). *About*. Abgerufen am 27. Juni 2024 von udio.com: <https://www.udio.com/about-us>
- O.V. (5. Juli 2024). *acestudio.ai*. Abgerufen am 2. Juli 2024 von Pro AI Vocalist For All Your Vocal Creations: <https://www.cestudio.ai>
- O.V. (2024). *Create AI Covers with your Favorite Voices*. Abgerufen am 1. August 2024 von jammable.com: <https://www.jammable.com>
- O.V. (10. Juli 2024). *Disclosure for Voice Talent*. Abgerufen am 4. Juli 2024 von support.cestudio.ai: <https://support.cestudio.ai/article/124-disclosure-for-voice-talent>
- O.V. (28. Juni 2024). *How do I make music with Udio?* Abgerufen am 25. Juni 28 von udio.com: <https://www.udio.com/guide#prompting>
- O.V. (2024). *List of AI Voice Models (85,778)*. Abgerufen am 1. August 2024 von voice-models.com: <https://voice-models.com>

- O.V. (2024). *Meet LALAL.AI Voice Changer*. Abgerufen am 1. August 2024 von lalal.ai:  
<https://www.lalal.ai/voice-changer/>
- O.V. (2024). *Musikproduktion*. Abgerufen am 10. Juli 2024 von peak-studios.de:  
<https://www.peak-studios.de/musikproduktion/>
- O.V. (28. Juni 2024). *ud.io.com*. Abgerufen am 20. Juni 2024 von ud.io.com:  
<https://suno.com/create>
- O.V. (o.D.). *AI Prompt Marketplace*. Abgerufen am 10. Juli 2024 von chatx.ai:  
<https://chatx.ai/marketplace/>
- O.V. (o.D.). *Chat GPT*. Abgerufen am 27. Juli 2024 von chatgpt.com:  
<https://chatgpt.com/c/3b2571b0-e10c-4f82-9d42-3c9719eacaf8>
- Reinartz, L. S. (o.D.). *Songwriting*. Abgerufen am 18. Juli 2024 von gitarrebass.de:  
<https://www.gitarrebass.de/thema/songwriting/>
- Rember, R. (20. April 2024). *Suno AI: Revolutionizing Music Creation for Everyone*. Abgerufen am 30. Juli 2024 von elblog.pl: <https://elblog.pl/2024/04/20/suno-ai-revolutionizing-music-creation-for-everyone/>
- Richter, B. (2016). Stimme. In A.-C. Mecke, M. Pfeleiderer, B. Richter, & T. Seedorf (Hrsg.), *Lexikon der Gesangsstimme* (S. 584-586). Laaber: Laaber.
- Rouard, S. (o.D.). *CRASH: Raw Audio Score-based Generative Modeling for Controllable High-resolution Drum Sound Synthesis*. Abgerufen am 30. Juli 2024 von github.com: <https://github.com/simonrouard/CRASH>
- Scheuer, D. (2020). *Akzeptanz von Künstlicher Intelligenz*. Hamburg: Springer.
- Seedorf, T. (2023). Stimme und Gesang. In H. von Loesch, R. Wolf, & T. Ertelt, *Geschichte der musikalischen Interpretation im 19. und 20. Jahrhundert* (Bd. 3). Stuttgart: J.B. Metzler.
- Singh, S. K., Kumar, S., & Mehra, P. S. (2023). Chat GPT & Google Bard AI: A Review. *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*. Delhi: IEEE.
- Tencer, D. (10. April 2024). *New AI-Powered 'instant' Music-Making App Udio Raises \$10M; Launches with backing from WILL.I.AM, Common, Unitedmasters, A16Z*. Abgerufen am 10. Juli 2024 von musicbusinessworldwide.com:  
<https://www.musicbusinessworldwide.com/new-ai-powered-instant-music-making-app-udio-raises-10m-launches-with-backing-from-will-i-am-common-unitedmasters-a16z/>
- Tiedemann, M. (29. Mai 2018). *KI, künstliche neuronale Netze, Machine Learning, Deep Learning: Wir bringen Licht in die Begriffe rund um das Thema „Künstliche Intelligenz“*. Abgerufen am 15. Juli 2024 von alexanderthamm.com:  
[https://www.alexanderthamm.com/de/blog/ki\\_artificial-intelligence-ai-kuenstliche-neuronale-netze-machine-learning-deep-learning/](https://www.alexanderthamm.com/de/blog/ki_artificial-intelligence-ai-kuenstliche-neuronale-netze-machine-learning-deep-learning/)
- Warburton, T. (21. März 2024). *How to make Your BAD Vocals GREAT with ACE Studio*. Abgerufen am 3. Juli 2024 von youtube.com:  
<https://www.youtube.com/watch?v=pKsIHlw7DDs&t=196s>
- Warburton, T. (5. Mai 2024). *Make YOUR voice sound AWESOME: Trying out ACE Custom Voice*. Abgerufen am 3. Juli 2024 von Youtube.com:  
<https://www.youtube.com/watch?v=mJPwLyvQMpl>

- Weitzman, C. (4. Oktober 2023). *RVC Vocal Models: Revolutionizing Voice Technology*. Abgerufen am 30. Juli 2024 von [speechify.com](https://speechify.com/blog/rvc-vocal-models/):  
<https://speechify.com/blog/rvc-vocal-models/>
- Wieduwilt, C. (1. März 2024). *The best 5 AI stem separation tools in 2024*. Abgerufen am 31. Juli 2024 von [aimusicpreneur.com](https://www.aimusicpreneur.com/ai-tools-news/the-best-ai-stem-separation-tools-ai-stem-splitter/): <https://www.aimusicpreneur.com/ai-tools-news/the-best-ai-stem-separation-tools-ai-stem-splitter/>
- Willman, C. (17. April 2023). *AI-Generated Fake 'Drake'/'Weeknd' Collaboration, 'Heart on My Sleeve,' Delights Fans and Sets Off Industry Alarm Bells*. Abgerufen am 1. August 2024 von [variety.com](https://variety.com/2023/music/news/fake-ai-generated-drake-weeknd-collaboration-heart-on-my-sleeve-1235585451/): <https://variety.com/2023/music/news/fake-ai-generated-drake-weeknd-collaboration-heart-on-my-sleeve-1235585451/>
- Woods, B. (26. Februar 2024). *Why Stem Separation is so hard*. Abgerufen am 1. August 2024 von [audiostrip.com](https://www.audiostrip.com/blog/why-stem-separation-is-hard): <https://www.audiostrip.com/blog/why-stem-separation-is-hard>



## Anhangsverzeichnis

Anhang 1: Songtexte Prompt 1 ChatGPT .....	73
Anhang 2: I like the way you kiss me .....	78
Anhang 3: R U Mine.....	80
Anhang 4: Pumped Up Kicks .....	82
Anhang 5: Don't .....	84
Anhang 6: Notizen Song .....	87
Anhang 7: Output ChatGPT fehlender Vers .....	88
Anhang 8: Songtexteingabe in Suno .....	89

## Anhang 1: Songtexte Prompt 1 ChatGPT

### Text 1: Title: "Summer by the Shore"

**Chorus:** Under the stars, where the waves kiss the sand,  
We're chasing the night with our hearts in our hands.

In the glow of the fire, with the moon in our eyes,  
We're living the dream, where the freedom lies.

Summer by the shore, where our spirits are free,  
I'm right where I belong, and it's all meant to be.

Laughing with you as the sun starts to fade,  
We're dancing in the light of a memory we made.

**Verse 1:** Remember that night, we were young and so bright,  
Bonfire blazing, with the sky as our sight.

Every word was a melody, every laugh was a song,  
In the warmth of that evening, we felt we belonged.

You said, "Life's just a journey, and we're on the right track,"  
With the sand in our toes and the world at our back.

We were dreaming of futures, of places we'd go,  
But tonight, all we needed was the waves and the glow.

**Chorus:** Under the stars, where the waves kiss the sand,  
We're chasing the night with our hearts in our hands.

In the glow of the fire, with the moon in our eyes,  
We're living the dream, where the freedom lies.

Summer by the shore, where our spirits are free,  
I'm right where I belong, and it's all meant to be.

Laughing with you as the sun starts to fade,  
We're dancing in the light of a memory we made.

**Verse 2:** Time's flown like a seagull, soaring high and wide,  
But those moments we shared are never left behind.

Every wave that crashes brings us back to that place,  
With a smile and a heartbeat, I can see your face.

We were fearless in the twilight, with stars in our eyes,  
Creating forever in the summer skies.

Now as the years roll by and we chase new dreams,  
That beach is a beacon in the soft moonbeams.

**Chorus:** Under the stars, where the waves kiss the sand,  
We're chasing the night with our hearts in our hands.

In the glow of the fire, with the moon in our eyes,

We're living the dream, where the freedom lies.  
Summer by the shore, where our spirits are free,  
I'm right where I belong, and it's all meant to be.  
Laughing with you as the sun starts to fade,  
We're dancing in the light of a memory we made.

**Bridge:** And when life gets heavy, and the days feel too long,  
I'll recall that night, and it will make me strong.

The echoes of our laughter, the waves on the sand,  
Will remind me forever, we were just where we planned.  
With every sunset and each whispering breeze,  
I'll carry that freedom wherever I please.

**Chorus:** Under the stars, where the waves kiss the sand,  
We're chasing the night with our hearts in our hands.

In the glow of the fire, with the moon in our eyes,  
We're living the dream, where the freedom lies.  
Summer by the shore, where our spirits are free,  
I'm right where I belong, and it's all meant to be.  
Laughing with you as the sun starts to fade,  
We're dancing in the light of a memory we made.

**Outro:** So here's to the nights and the summer's sweet call,  
To the moments that bind us, to the memories of all.  
With the sea as our witness and the stars shining bright,  
We'll treasure those summers and the freedom of night.

## **Text 2: Title: "Golden Hour on the Shore"**

**Chorus:** Underneath the sky so blue, with the ocean in our view,  
We're laughing till the morning light, in the magic of the night.  
In the golden hour's glow, where the waves gently flow,  
We find our freedom, feel alive, under the stars, we thrive.  
Summer's song in the air, and our hearts are laid bare,  
This is where we're meant to be, in this sweet serenity.

**Verse 1:** Remember those nights, the sand between our toes,  
With the fire burning bright and the moonlight's gentle glow.  
Every moment felt so right, with your hand holding mine,  
We were young, we were free, living life in double time.  
The breeze whispered secrets, tales of old and new,  
With every crashing wave, our dreams were coming true.

We sang songs of tomorrow, danced through the haze,  
In that summer's embrace, we found endless ways.

**Chorus:** Underneath the sky so blue, with the ocean in our view,  
We're laughing till the morning light, in the magic of the night.  
In the golden hour's glow, where the waves gently flow,  
We find our freedom, feel alive, under the stars, we thrive.  
Summer's song in the air, and our hearts are laid bare,  
This is where we're meant to be, in this sweet serenity.

**Verse 2:** Years may pass us by, like a whisper in the breeze,  
But those nights by the shore are our timeless memories.  
Every sunset brings me back to that sacred place,  
Where laughter was our language and love filled every space.  
We chased the horizon, with dreams in our sights,  
In the warmth of the sand, under countless starlit nights.  
We were more than just friends, we were family in disguise,  
And those moments we lived still light up my skies.

**Chorus:** Underneath the sky so blue, with the ocean in our view,  
We're laughing till the morning light, in the magic of the night.  
In the golden hour's glow, where the waves gently flow,  
We find our freedom, feel alive, under the stars, we thrive.  
Summer's song in the air, and our hearts are laid bare,  
This is where we're meant to be, in this sweet serenity.

**Bridge:** And when the world feels heavy, and the night is too long,  
I'll close my eyes and remember, the place where we belong.  
With the echoes of our laughter and the salt in the breeze,  
We're forever bound together, by those golden memories.  
We'll sail through every storm, with the stars as our guide,  
In the heart of summer's promise, where our spirits collide.

**Chorus:** Underneath the sky so blue, with the ocean in our view,  
We're laughing till the morning light, in the magic of the night.  
In the golden hour's glow, where the waves gently flow,  
We find our freedom, feel alive, under the stars, we thrive.  
Summer's song in the air, and our hearts are laid bare,  
This is where we're meant to be, in this sweet serenity.

**Outro:** So here's to the nights and the dreams that we share,  
To the love that we found in the summer's warm air.  
With the sea as our witness and the stars as our guide,  
We'll cherish these moments, with our hearts open wide.

In the golden hour's light, where our spirits soar,  
We'll always find our way back, to the shore we adore.

**Text 3: Title: "Echoes of the Tide"**

**Chorus:** Underneath the starlit sky, where the waves meet the shore,  
We're lost in the rhythm, wanting nothing more.

With the summer breeze in our hair, and the moonlight in our eyes,  
We're living in the moment, feeling so alive.

Echoes of laughter, memories collide,

We're right where we belong, in the warmth of the tide.

**Verse 1:** Remember those nights, the fire's gentle glow,

Our hearts intertwined, with nowhere else to go.

Every word a whisper, every smile a song,

We danced through the darkness, right where we belong.

Friendship painted in the stars, love written in the sand,

We knew we had forever, with our toes in the land.

Nostalgia in the air, freedom in our sight,

We embraced the night, holding on so tight.

**Chorus:** Underneath the starlit sky, where the waves meet the shore,  
We're lost in the rhythm, wanting nothing more.

With the summer breeze in our hair, and the moonlight in our eyes,

We're living in the moment, feeling so alive.

Echoes of laughter, memories collide,

We're right where we belong, in the warmth of the tide.

**Verse 2:** Time flows like the ocean, but those moments stay,

In the golden light of dawn, they never fade away.

Every sunset a reminder, of the love we hold dear,

In the quiet of the night, it's always crystal clear.

We were young and wild, with dreams in our gaze,

Building castles in the sand, lost in summer's haze.

With every wave that crashes, I feel your heartbeat near,

And in every star that twinkles, our story's always here.

**Chorus:** Underneath the starlit sky, where the waves meet the shore,  
We're lost in the rhythm, wanting nothing more.

With the summer breeze in our hair, and the moonlight in our eyes,

We're living in the moment, feeling so alive.

Echoes of laughter, memories collide,  
We're right where we belong, in the warmth of the tide.

**Bridge:** When the nights grow colder, and the days drift apart,  
I'll hold on to the summer, and the fire in my heart.

We'll sail through the seasons, with the stars as our guide,  
In the embrace of the ocean, where dreams never hide.

We'll cherish these moments, like a treasure so rare,  
In the echoes of the tide, we'll always find our share.

**Chorus:** Underneath the starlit sky, where the waves meet the shore,  
We're lost in the rhythm, wanting nothing more.

With the summer breeze in our hair, and the moonlight in our eyes,  
We're living in the moment, feeling so alive.

Echoes of laughter, memories collide,  
We're right where we belong, in the warmth of the tide.

**Outro:** So here's to the nights and the dreams that we weave,  
In the glow of the fire, it's hard to believe.

With the sea as our witness and the stars shining bright,  
We'll carry these memories through every night.

In the echoes of the tide, where our hearts reside,  
We're forever bound together, in the summer's pride.

## Anhang 2: I like the way you kiss me

### I like the way you kiss me – Artemas

**Chorus:** I like the way you kiss me

I like the way you –, uh

I like the way you kiss me, I can tell you miss me

I can tell it hits, hits, hits, hits

Not tryna be romantic, I'll hit it from the back

Just so you don't get attached (yes, yes, yes)

I like the way you kiss me, I can tell you miss me

I can tell it hits, hits, hits, hits

Not tryna be romantic, I'll hit it from the back

Just so you don't get attached (yes, yes, yes)

You bite my lip just for the taste

You're on your knees, I'm on the case

You take the heat and with such grace

You say, "We're done, " but here you stay

You said you're scared, I'll let you down (I'll let you down)

Just stick around and you'll find out (and you'll find out)

But don't you wanna make me proud? (So make me proud)

'Cause I'm so proud, baby, I'm so proud of you

**Chorus:** I like the way you kiss me, I can tell you miss me

I can tell it hits, hits, hits, hits

Not tryna be romantic, I'll hit it from the back

Just so you don't get attached (yes, yes, yes)

I like the way you kiss me, I can tell you miss me

I can tell it hits, hits, hits, hits

Not tryna be romantic, I'll hit it from the back

Just so you don't get attached (yes, yes, yes)

Do I stress you out? Can I help you out?

Does it turn you on when I turn you around?

Can we make a scene? (Scene)

Can you make it loud? (Loud)

'Cause I'm so proud, baby, I'm so proud of you

So proud of you

Hmm, so proud of you

**Chorus:** I like the way you kiss me, I can tell you miss me

I can tell it hits, hits, hits, hits

Not tryna be romantic, I'll hit it from the back

Just so you don't get attached (yes, yes, yes)

I like the way you kiss me, I can tell you miss me

I can tell it hits, hits, hits, hits

Not tryna be romantic, I'll hit it from the back

Just so you don't get attached (yes, yes, yes)

I like the way you kiss me

I like the way you –, uh



### Anhang 3: R U Mine

#### R U Mine – Arctic Monkeys

I'm a puppet on a string  
Tracy Island, time-travelin' diamond cutter-shaped heartaches  
Come to find you four in some velvet mornin' years too late  
She's a silver linin', lone ranger ridin' through an open space  
In my mind, when she's not right there beside me

I go crazy 'cause here isn't where I wanna be  
And satisfaction feels like a distant memory  
And I can't help myself  
All I wanna ever say is, "Are you mine?"  
Well, are you mine?  
Are you mine?  
Are you mine? Oh, ah

I guess what I'm tryin' to say is I need the deep end  
Keep imaginin' meetin', wished away entire lifetimes  
Unfair we're not somewhere misbehavin' for days  
Great escape, lost track of time and space  
She's a silver linin', climbin' on my desire

And I go crazy 'cause here isn't where I wanna be  
And satisfaction feels like a distant memory  
And I can't help myself  
All I wanna ever say is, "Are you mine?"  
Well, are you mine? (Are you mine tomorrow?)  
Are you mine? (Or just mine tonight?)  
Are you mine? (Are you mine? Mine?)

And the thrill of the chase moves in mysterious ways  
So in case I'm mistaken, I  
Just wanna hear you say, "You got me baby"  
"Are you mine?"  
She's a silver linin', lone ranger ridin' through an open space  
In my mind, when she's not right there beside me

I go crazy 'cause here isn't where I wanna be  
And satisfaction feels like a distant memory  
And I can't help myself  
All I wanna ever say is, "Are you mine?"  
Well, are you mine? (Are you mine tomorrow?)  
Are you mine? (Or just mine tonight?)  
Are you mine? (Are you mine tomorrow?)  
(Oh, ah, or just mine tonight?)

## Anhang 4: Pumped Up Kicks

### Pumped Up Kicks – Foster the People

Robert's got a quick hand  
He'll look around the room, but won't tell you his plan  
He's got a rolled cigarette  
Hanging out his mouth, he's a cowboy kid, yeah  
He found a six-shooter gun  
In his dad's closet, and with a box of fun things  
I don't even know what  
But he's coming for you, yeah, he's coming for you

All the other kids with the pumped up kicks  
You better run, better run outrun my gun  
All the other kids with the pumped up kicks  
You better run, better run faster than my bullet  
All the other kids with the pumped up kicks  
You better run, better run outrun my gun  
All the other kids with the pumped up kicks  
You better run, better run faster than my bullet

Daddy works a long day  
He'd be coming home late, yeah, he's coming home late  
And he's bringing me a surprise  
'Cause dinner's in the kitchen, and it's packed in ice  
I've waited for a long time  
Yeah, the sleight of my hand is now a quick pull trigger  
I reason with my cigarette  
Then say, "Your hair's on fire, you must have lost your wits, yeah"

All the other kids with the pumped up kicks  
You better run, better run outrun my gun  
All the other kids with the pumped up kicks  
You better run, better run faster than my bullet  
All the other kids with the pumped up kicks  
You better run, better run outrun my gun  
All the other kids with the pumped up kicks  
You better run, better run faster than my bullet

Run, run, run, run

R-r-r-run, run, run, run

R-r-r-run, run, run, run

R-r-r-run

Run, run

Run, run, run

All the other kids with the pumped up kicks

You better run, better run outrun my gun

All the other kids with the pumped up kicks

You better run, better run faster than my bullet

All the other kids with the pumped up kicks

You better run, better run outrun my gun

All the other kids with the pumped up kicks

You better run, better run faster than my bullet

All the other kids with the pumped up kicks

You better run, better run outrun my gun

All the other kids with the pumped up kicks

You better run, better run faster than my bullet

## Anhang 5: Don't

### Don't - Ed Sheeran

I met this girl late last year  
She said, "Don't you worry if I disappear"  
I told her, "I'm not really looking for another mistake"  
I called an old friend thinking that the trouble would wait  
But then I jumped right in a week later, returned  
I reckon she was only looking for a lover to burn  
But I gave her my time for two or three nights  
Then I put it on pause until the moment was right  
I went away for months until our paths crossed again  
She told me, "I was never looking for a friend  
Maybe you could swing by my room around ten  
Baby, bring the lemon and a bottle of gin  
We'll be in between the sheets 'til the late AM"  
Baby, if you wanted me then should have just said, she's singing

Don't fuck with my love  
That heart is so cold  
All over my home  
I don't wanna know that, babe  
Don't fuck with my love  
I told her she knows  
Take aim and reload  
I don't wanna know that, babe

And for a couple weeks I only wanna see her  
We drink away the days with a takeaway pizza  
Before, a text message was the only way to reach her  
Now she's staying at my place and loves the way I treat her  
Singing out Aretha, all over the track like a feature  
And never wants to sleep, I guess that I don't want to either  
But me and her, we make money the same way  
Four cities, two planes, the same day  
And those shows have never been what it's about  
But maybe we'll go together and just figure it out  
I'd rather put on a film with you and sit on a couch  
But we should get on a plane or we'll be missing it now

Wish I'd have written it down, the way that things played out  
When she was kissing him, how I was confused about  
Now she should figure it out, while I'm sat here singing

Don't fuck with my love  
That heart is so cold  
All over my home  
I don't wanna know that, babe  
Don't fuck with my love  
I told her she knows  
Take aim and reload  
I don't wanna know that, babe

On my hotel door  
I don't even know if she knows what for  
She was crying on my shoulder, I already told ya  
Trust and respect is what we do this for  
I never intended to be next (you bastard)  
But you didn't need to take him to bed, that's all  
And I never saw him as a threat (you bastard)  
Until you disappeared with him to have sex, of course  
It's not like we were both on tour  
We were staying on the same fucking hotel floor  
And I wasn't looking for a promise or commitment  
But it was never just fun, and I thought you were different  
This is not the way you realized what you wanted  
It's a bit too much, too late if I'm honest  
And all this time, God knows I'm singing

Don't fuck with my love  
That heart is so cold  
All over my home  
I don't wanna know that, babe  
Don't fuck with my love  
I told her she knows  
Take aim and reload  
I don't wanna know that, babe

Don't fuck with my love  
That heart is so cold  
All over my home  
I don't wanna know that, babe  
Don't fuck with my love  
I told her she knows  
Take aim and reload  
I don't wanna know that, babe

## Anhang 6: Notizen Song

Small feet went this way  
sang a lullaby  
Ran through a summers day  
catcher in the rye

pre-chorus 1

please don't fall, grasping to this dream /Just watch out, can't save you from yourself

Chorus

Bursting out/Run wherever/ last forever I wanna sing  
to/ and just sing to a butterfly's dance, so charming, melodies of a childhoods swing  
Hear me out and grow a wing/ laugh/ fly wherever/ together grow some wings  
Glide over this old world my darling, we can do just anything, I'm on my way now /Feel  
the warmth around your body, life can be such a pleasant thing, be with the flow now

vers a 2

Sky is violet, where's my pilot  
towards uncharted places  
Thought I'm all set/We are all set, let us forget/ make me forget lies we tell to ourselves

vers b 2

Come on, this is now or never,  
catch me in the sky  
So we can just be together,  
dancing through the rye



## Anhang 7: Output ChatGPT fehlender Vers

Output 2 nach Aufforderung mehr Versionen zu kreieren:

Certainly! Here are a few more versions of Verse A1 with a similar syllable count:

### **Version 1:**

Sun is rising, no disguising  
Hope in all our faces  
Steps we're taking, hearts are waking  
Dreams we keep to ourselves

### **Version 2:**

Dawn is breaking, light cascading  
Over open spaces  
Journey started, paths uncharted  
Stories we tell to ourselves

### **Version 3:**

Day is dawning, softly calling  
Through familiar places  
Eyes wide open, future spoken  
Whispers we tell to ourselves

### **Version 4:**

Sky is clearing, dreams appearing  
In forgotten spaces  
Heartbeats racing, life embracing  
Songs we sing to ourselves

Feel free to mix and match lines or modify them as needed to fit your melody and the overall feel of your song!

## Anhang 8: Songtexteingabe in Suno

Lyrics:

[Instrumental intro]

[Verse]

My master is pulling the strings in a wrong direction  
Have I ever had control over myself?  
Disaster can't be avoided, my life is reckless  
Have a ticking time bomb in every cell

[Pre-Chorus]

Oh I know, it's just hysteria  
Oh I'm in the front row

[Chorus]

Don't feel my feet, I'm just drifting around  
Let my body float through the masses  
Am I a slave or a king?  
Screaming blood out of my throat

[Verse]

Too tired, I feel like I am on autopilot  
Seeing everything just rushing by  
These faces are choosing loud over quiet  
They are carrying me across the sky

[Pre-Chorus]

Oh I know, is it hysteria?  
Oh I should enjoy the show

[Chorus]

Don't feel my feet, I'm just drifting around  
Let my body float through the masses  
Am I a slave or a king?  
Screaming blood out of my throat

[Instrumental]

[Instrumental Build-up]

[Bridge]

Ohh I am a ticking time bomb

Ohh ready to explode

[mellow, soft Chorus, one octave down]

Don't feel my feet, I'm just drifting around

Let my body float through the masses

Am I a slave or a king?

Screaming blood out of my throat

Don't feel my feet, I'm just drifting around

Let my body float through the masses

Am I a slave or a king?

Screaming blood out of my throat