

Content Creation mit KI

» Hier geht's  
direkt  
zum Buch

# **DIE LESEPROBE**

# Kapitel 5

## Audio und Musik mit KI produzieren und optimieren

*Aus Zukunftsmusik ist Realität geworden: die Revolution in der Audiobranche.*

Für das Content-Marketing für Marken, Medien und Unternehmen sind Text, Fotos und Videos drei sehr wichtige Formate. Aber aller Dynamik zum Trotz: *Video did not kill the »Audio-Star«*, ganz im Gegenteil. Laut Statista suchen und konsumieren die Deutschen immer mehr Audioinhalte: »Während sich im Jahr 2018 die Anzahl der Nutzer von Online-Audioangeboten noch auf 32,8 Millionen belief, ist diese Zahl im Jahr 2021 auf 45,3 Millionen gestiegen. Damit erreichen Audioangebote im Internet mittlerweile rund 69,5 % der Bevölkerung ab 14 Jahren.« Die rasante Entwicklung des Podcast-Genres in den letzten Jahren ist ein deutlicher Beleg für den Bedeutungsgewinn dieses Formats. So hört jeder Zweite ab 14 Jahren in Deutschland (50 %) mindestens monatlich Podcasts oder Radio über das Internet.<sup>1</sup>

Für Content Creator, die mit Marken arbeiten, ist Audio-Content also ein zunehmend wichtiger Schlüssel, um direkt ins Ohr und damit in die Herzen ihrer Communitys vorzustoßen, wann, wo und wie diese es sich wünschen. Und in derlei Hinsicht hat Audio einiges zu bieten:

- ▶ Audio-Content kann in Echtzeit konsumiert, aber auch im Voraus produziert werden. Audio-Content hatte schon immer eine hohe Eignung für Content on demand. Die Digitalisierung der Medien hat diesen Vorteil noch weiter verstärkt.
- ▶ Audio-Content bietet den Hörerinnen und Hörern ein freihändiges und bildschirmfreies Inhaltserlebnis. Damit kann er – im Gegensatz zu Texten, Bildern, Videos und Präsentationen – im Hintergrund, nebenher, während der ansonsten medienfreien Zeit und unterwegs konsumiert werden.
- ▶ Audio-Content ermöglicht barrierefreien Content-Genuss.

---

<sup>1</sup> Statista, Statistiken zum Thema Online-Audio-Inhalte, <https://de.statista.com/themen/5667/online-audio-inhalte/>, 28.02.2024 [02.04.2024]

- ▶ Audio-Content, ob als Stimme, Musik oder Soundlogo, verleiht einer Marke und einem Unternehmen eine zusätzliche Qualität und Emotionalität.
- ▶ Audio-Content hilft Marken, die eine eigene Corporate Language und einen eigenen Corporate Sound entwickelt haben, sich bei den Zuhörenden nachhaltig zu etablieren und – wortwörtlich – als einmalig im Ohr zu bleiben. Damit bekommen Unternehmen einen höheren Wiedererkennungswert und können die Bindung zu ihren Kunden stärken.

Durch neue Formate und Nutzungsweisen steigt die Konkurrenz um das Ohr des Publikums, und der Markt für Audioinhalte wächst dynamisch. Daher gehen Content Creator immer neue Wege, um immer mehr qualitativ überzeugende und immer vielfältigere Audioinhalte zu schaffen. Und der Einsatz von künstlicher Intelligenz bietet ihnen genau dafür großartige, effiziente und bisher ungeahnte Möglichkeiten – heute und in ganz naher Zukunft.

### 5.1 Audioinhalte produzieren

Schauen wir uns zunächst einige der wichtigsten und aktuellen Formate an, mit denen Content Creator Audio für Marken, Unternehmen und Medienhäuser produzieren und einsetzen – und für deren Produktion und Konzeption der Einsatz künstlicher Intelligenz neue Wege öffnet.

- ▶ *Podcast*: Der Begriff Podcast setzt sich aus *Pod* (Play on demand) und *Cast* (Broadcast) zusammen. In mehreren meist wöchentlich erscheinenden Episoden liefern Podcasts Unterhaltung, Nachrichten, Informationen und Wissen – auf Knopfdruck. Das heißt, die Hörerschaft entscheidet sich bewusst dafür, Ihnen und Ihrer Marke zuzuhören – unabhängig von Zeit und Ort. Das Audioformat erlebt einen Boom. In Deutschland erreichte die Nutzung im Jahr 2022 einen neuen Rekordwert: So hörten sich annähernd zwei von fünf Deutschen (43 %) zumindest selten Podcasts an. Im Jahr 2016 waren es laut Statista nur 14 %. Und 5,8 % hören Podcasts fast täglich – aus Interesse an Themen, zum Stressabbau und vor allem, um Langeweile abzubauen.<sup>2</sup>
- ▶ *Hörbuch*: Das Hörbuch ist ein in sich abgeschlossenes, inhaltlich klar strukturiertes Werk, das ein Thema umfassend behandelt. Auf dem Weg zur Arbeit, beim Joggen, im Auto oder auf Reisen – ein Hörbuch kann man hören, egal wo und wann. Das ist eine gute Chance für Marken, Audioinhalte zu produzieren, die eine gewisse Tiefe, Stringenz und Expertise haben. Eine spannende Unternehmensgeschichte, eine interessant erzählte Jahresbilanz oder auch ein gut erzählter Nachhaltigkeitsbericht sind

---

<sup>2</sup> Statista, Statistiken zum Thema Podcasts, <https://de.statista.com/themen/4343/podcasts/#topicOverview>, 28.02.2024 [02.04.2024]

Audioinhalte, die als Hörbuch veröffentlicht werden können. Doch davon gibt es noch viel zu wenig. Denken Sie deshalb bei Ihrer nächsten Audio-Content-Produktion auch an das Medium Hörbuch und überraschen Sie Ihre Interessenvertreter.

- ▶ *Musik*: Musik ist von jeher mit der Menschheit verbunden. Sie verändert Herzschlag, schafft Nähe nicht nur zwischen Menschen, sondern auch zwischen Marken und ihrer Community. Und sie weckt tiefe Emotionen: So bekam der Netflix-Blockbuster »Im Westen nichts Neues« 2023 den Oscar für die beste Filmmusik, und kaum ein Reel auf Instagram oder TikTok kommt mehr ohne Musik-Snippets aus. Auch das Streamen von Musik wird in Deutschland immer beliebter. Laut dem Online-Audio-Monitor 2023 hören 16,8 % der befragten Nutzerinnen und Nutzer ein- bis mehrmals in der Woche Musik im Streaming.<sup>3</sup>
- ▶ *Radio*: Der Audioklassiker ist das Radio, ein ausschließlich auditives Medium mit hoher Reichweite. Fast ein Viertel (23,4 %) der Teilnehmenden lassen sich täglich oder fast täglich vom Radio berieseln.<sup>4</sup> Radio zeigt, wie man mit einer guten Programmstrategie und einer variablen Programmplanung Hörerinnen und Hörer mit Audio-Content binden kann – und das rund um die Uhr. Dieses Wissen kann man für die eigene Content-Produktion nutzen: Die Supermarktkette Penny beispielsweise betreibt für ihre Märkte das eigene Instore-Radio »PennyLive«. Seit 2015 läuft dort in den Filialen von 6 bis 20 Uhr nicht nur Musik mit eingestreuter Werbung, sondern auch ein live moderiertes Radioprogramm. Betrieben wird der Sender vom konzern-eigenen Instore-Radiodienstleister Maxfive<sup>5</sup>. Ziel ist es, einerseits gezielt Kaufimpulse zu setzen und andererseits die Kundschaft mit Themen wie Food-Trends und Ernährung zu entschleunigen, damit sie sich im Markt wohlfühlt, länger bleibt und – logischerweise – mehr einkauft. Penny nutzt das Programm auch vor Ladenöffnung für die interne Kommunikation mit den Beschäftigten – Communitybuilding in Echtzeit.

Mit dem Aufkommen von KI-gestützten Technologien für Spracherkennung, Sprachsynthese, Musikgenerierung und Sounddesign können die Produktion der vorgestellten Formate, die damit verbundenen Arbeitsabläufe und sogar der Konsum völlig neu gedacht werden. Sei es, um Audioaufnahmen qualitativ aufzuwerten, Texte vorlesen zu lassen oder sie sogar in ein fesselndes Hörspiel, ein mitreißendes Musikstück oder eine Live-Radiosendung umzuwandeln (*Text zu Audio*). Oder um bereits produzierte Audioaufnahmen lesbar und damit barrierefrei als Text zur Verfügung zu stellen (*Audio zu Text*).

3 Statista, Umfrage zur Nutzungshäufigkeit von Online-Audio-Angeboten in Deutschland 2023, <https://de.statista.com/statistik/daten/studie/636767/umfrage/nutzungshaeufigkeit-von-radio-und-musikformaten-im-internet-in-deutschland/>, 02.01.2024, [02.04.2024]

4 Ebenda

5 <https://maxfive.com/hear/>

### Podcast produzieren

*Podcasts* sind informativ, unterhaltsam oder beides. Es gibt sie inzwischen für fast jedes Thema und jede Nische. Und sie eignen sich hervorragend für gutes Storytelling. Die Moderatorinnen und Moderatoren sind auf Augenhöhe mit ihrem Publikum, und es entsteht eine emotionale Bindung zum Publikum. Alles gute Gründe, einen Podcast zu produzieren!

Doch hinter jedem erfolgreichen Podcast steckt auch eine detaillierte Planung:<sup>6</sup>

- ▶ In der Konzeptionsphase gilt es, die Ziele der Produktion festzulegen. Was möchten Sie mit dem Podcast für sich oder Ihre Marke erreichen? Wer soll angesprochen werden? Welches Thema bewegt Ihre Zielgruppe? Entsprechend eingängig und einzigartig sollten der Name und das Cover des Podcasts gestaltet werden.
- ▶ Ein Podcast braucht auch Musik für Intro und Outro, Zwischentitel und Jingle. Sie macht das Angebot einmalig und unverwechselbar.
- ▶ Zum Podcast-Konzept gehört ebenfalls die Entscheidung, welche Tonalität der Podcast haben soll: informativ seriös oder doch unterhaltsam spaßig?
- ▶ Bei der Umsetzung hilft ein Podcast-Skript, die richtige Dramaturgie einzuhalten und die Hörer von den ersten Sekunden einer Episode an zu fesseln. Für einen stimmigen Ablauf sollte jede Podcast-Episode entsprechend geplant werden.
- ▶ Die Aufnahme sollte möglichst professionell mit Mikrofon und ohne Hintergrundgeräusche erfolgen.
- ▶ Der anschließende Schnitt des aufgenommenen Materials macht den Podcast erst hörtauglich: Versprecher und redundante Inhalte werden herausgeschnitten, Audiopegel und Lautstärke optimiert.
- ▶ Shownotes machen jede Podcast-Episode suchmaschinentauglich, indem Sie die entsprechenden Keywords einfügen. Und sie erhöhen den Mehrwert Ihres Audio-Contents mit entsprechenden Servicelinks.

#### **Workflow: Podcast-Produktion mit KI, BZ-Podcast »200 Sekunden Baden«**

Der Podcast der Badischen Zeitung »200 Sekunden Baden«<sup>7</sup> soll werktags einen schnellen regionalen Nachrichtenüberblick bieten. Dabei werden KI-Stimmen für die Sprachausgabe genutzt. Der Einsatz von KI bietet dabei sowohl Effizienzsteigerung als auch

6 Andreas Berens, Carsten Bolk, Einen Podcast erstellen. In: Create Content! Konzeption, Kreation, Content-Management, Rheinwerk Verlag, Bonn, 2021, S. 331 ff.

7 200 Sekunden Baden, <https://open.spotify.com/show/OP6jPGkBAFyBMAcIq2ZizZ?si=be23669a733740b6>

Lernmöglichkeiten hinsichtlich ihrer Anwendbarkeit und Grenzen im Journalismus. Der Workflow dazu sieht in etwa so aus:<sup>8</sup>

- ▶ *Auswahl der Nachrichten:* Am Abend wählt das Redaktionsteam die fünf wichtigsten Nachrichten aus, die bearbeitet und für das Audioformat vorbereitet werden.
- ▶ *Bearbeitung durch freie Mitarbeitende:* Ab 22 Uhr übernimmt an jedem Wochentag eine andere Person die Aufgabe, die Nachrichten für den Podcast zu kürzen, umzuformulieren und für die Veröffentlichung vorzubereiten.
- ▶ *Erstellung des Podcasts:* Die bearbeiteten Nachrichten werden in das Tool Storyflash eingespeist, von zwei KI-Stimmen gesprochen und zwischen einem Intro und Outro – gesprochen von einer realen Person – in den Podcast integriert.
- ▶ *Überprüfung und Anpassung:* Trotz der Automatisierung durch KI müssen die Beiträge vor der Veröffentlichung gegebenenfalls auf Aussprachefehler überprüft und angepasst werden.
- ▶ *Veröffentlichung:* Der fertige Podcast wird hochgeladen und ist täglich um 6:30 Uhr verfügbar.

Wer nun einen Podcast aufnehmen möchte und nicht auf ein professionelles Tonstudio zurückgreifen kann, z. B. bei Außenaufnahmen oder einfach mangels Budgets, hat in der Regel ein Problem mit Hintergrundgeräuschen oder Hall. Selten sind die Rahmenbedingungen optimal: Der Raum, in dem die Aufnahme stattfindet, war entweder zu groß oder zu leer, der Gesprächspartner saß während des Gesprächs plötzlich zu weit weg vom Mikrofon, oder Letzteres war einfach nicht optimal platziert.

Der dazu passende Spruch »Fix it in post«, also »Das bügeln wir in der Nachbereitung aus«, ist zwar alt, aber er wird zunehmend an Relevanz gewinnen: Denn mithilfe neuester KI-Audio-Enhancer können Content Creator beispielsweise Hintergrundgeräusche und *Raumhall* im Nachhinein entfernen, Effekte ihrer Wahl hinzufügen oder voreingestellte Optionen verwenden, um die Ausgabe für verschiedene Geräte wie Lautsprecher und Kopfhörer anzupassen und die Aufnahme professioneller klingen zu lassen. Und das, ohne Änderungen an der vorhandenen Hardware wie Mikrofon oder Aufnahmegerät vorzunehmen.

*KI-Audio-Enhancer*, deren Algorithmen im Nachhinein für ein klares Hörerlebnis sorgen, revolutionieren die Audiobranche. Sie können sie je nach Ihren Bedürfnissen nutzen. Daher hier nur eine kleine Auswahl aus einer weitaus größeren Palette von Tools.<sup>9</sup>

8 Synthetische Stimmen, echte Geschichten: Der Podcast-Plan der »Badischen Zeitung«. In: turi2, <https://www.turi2.de/community/badischen-zeitung-podcast-tipps-journalismus-kuenstliche-intelligenz>, 22.03.2024 [04.03.2024]

9 <https://www.futurepedia.io/ai-tools/audio-editing>

- ▶ *Adobe Podcast Enhance*: Adobe Podcast Enhance<sup>10</sup> entfernt störende Hintergrundgeräusche oder Raumklang. Um eine Audiodatei zu verbessern, zieht man sie einfach in ein Dateifenster, und die KI beginnt, die Aufnahme zu analysieren und zu optimieren. Das Tool ist kostenlos und lässt sich einfach bedienen – erfordert allerdings einen Adobe-Account.
- ▶ *Audo*: Audo<sup>11</sup> ist ein AI-Audio-Enhancer mit einer einfachen und intuitiven Benutzeroberfläche. Mit der App kann man seine Audiodateien unkompliziert hochladen oder direkt aufnehmen, um sie im Anschluss zu bearbeiten. Audo kann im Starterpaket mit einem Zeitlimit von 20 Minuten pro Monat kostenlos genutzt werden.
- ▶ *CleanVoice*<sup>12</sup> ist ein KI-gestütztes Audibearbeitungstool, mit dem sich automatisch Füllwörter, unangenehme Pausen oder Hintergrundgeräusche aus Podcast-Episoden entfernen lassen.

Aber KI kann nicht nur helfen, den Sound für einen Podcast zu optimieren, sondern ihn auch in verschiedenen Stimmen moderieren oder Skripte schreiben, wie das folgende Beispiel aus der Praxis zeigt:

### Der erste KI-generierte Podcast – AI4Democracy

*AI4Democracy*<sup>13</sup> war einer der ersten Versuche, einen komplett KI-generierten Podcast zu produzieren. Produziert wurde er von der Berliner Kommunikationsberatung Cosmonauts & Kings. Dabei war seine Form zugleich Programm: Es wurden Themen rund um künstliche Intelligenz im Zusammenhang mit politischer Kommunikation und Politik besprochen. Auf LinkedIn<sup>14</sup> und X (ehemals Twitter)<sup>15</sup> hat Berater Juri Schnöller die Erstellung des Podcasts beschrieben. Die Skripte wurden von GPT-4 geschrieben: »Die Prompts haben wir mit aktuellen akademischen Arbeiten und Argumenten gefüttert, um KI-Experten zu simulieren, die kontroverse Themen und aktuelle Positionen highlighten. Sie alle haben ihre eigene Persönlichkeit, und die Qualität der Prompts ist ziemlich beeindruckend«, beurteilt Schnöller das Ergebnis (siehe Abbildung 5.1). Host des englischsprachigen Podcast-Projekts ist der agentureigene Avatar Neil King, der über ein eigenes LinkedIn-Profil verfügt.<sup>16</sup> Er wird von anderen KI-basierten Moderatoren unterstützt. Die

<sup>10</sup> <https://podcast.adobe.com/enhance>

<sup>11</sup> <https://audo.ai>

<sup>12</sup> <https://cleanvoice.ai>

<sup>13</sup> <https://open.spotify.com/show/3iUcsyHxKxUXXKJcA1KNlo?si=5d0fd80088cf4465>

<sup>14</sup> <https://www.linkedin.com/pulse/first-ai-generated-political-podcast-here-juri-schn-25C3%25B6ller/?trackingId=CLntx4GPSR6z4ALt4jQorA%3D%3D>

<sup>15</sup> <https://x.com/jurischnoeller/status/1648256047222599687?s=20>

<sup>16</sup> <https://www.linkedin.com/in/neil-king-481589271/>

entsprechenden Stimmen liefern die Voice-Generatoren von ElevenLabs (siehe weiter hinten in diesem Kapitel) und play.ht.<sup>17</sup> Hören Sie am besten einmal rein, und machen Sie sich selbst ein Bild von der Wirkung dieses bemerkenswerten Projekts auf Ihre Ohren.

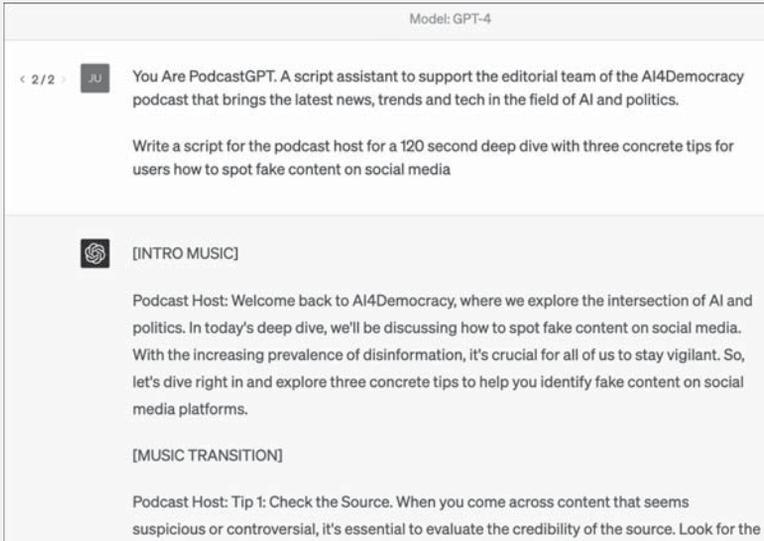


Abbildung 5.1 Ein Beispiel für den Scripting-Prozess mit GPT-4 (Quelle: LinkedIn)

### Content Repurposing: Audio-Content mit KI transkribieren und mehrfach verwenden – von Audio zu Text

Wenn Sie nun einen Podcast zu einem spannenden Thema mit einem interessanten Gegenüber produziert haben, um daraus Leads für den Vertrieb oder neue Hörerinnen und Hörer zu gewinnen, dann möchten Sie wahrscheinlich nicht, dass der Beitrag nach dem Live-Gang als »Einweg-Content« auf einer Playlist oder irgendwo auf der Website vor sich hinvegetiert. Doch vielen Content Creatorm im Unternehmensumfeld dürfte es ähnlich ergehen. Fest steht: Effizienter Umgang mit aufwendig produziertem Content geht auch anders.

Gerade produktionsintensive Audio- und Videoinhalte verdienen ein längeres Leben: *Content Repurposing* heißt das Stichwort. Unter Repurposing (auch Recycling oder Up-cycling genannt) versteht man die Wiederverwendung bereits bestehender Inhalte, indem diese in neuen Formaten aufbereitet und auf anderen Kanälen ausgespielt werden.

<sup>17</sup> <https://play.ht>

Gerade bei knappen Ressourcen in der Content-Redaktion ist die Weiterverwendung von Audioinhalten in Textform ein guter Schachzug: als Skript für einen Podcast oder ein Video zum Nachhören, als eigener Artikel auf der Website, als Transkript für die redaktionelle Recherche oder in Auszügen für eine Post-Serie in Social Media mit entsprechenden Zitaten.

Die wesentlichen Vorteile des Repurposing von Audio-Content sind:

- ▶ Sie erleichtern die Verteilung Ihrer Inhalte in verschiedenen Formaten und auf weiteren Kanälen.
- ▶ Suchmaschinen können Ihre Inhalte leichter finden.
- ▶ Sie erreichen ein breiteres Publikum mit unterschiedlichen Content-Vorlieben.
- ▶ Sie erhöhen die Zugänglichkeit Ihrer Inhalte für Menschen mit Hörproblemen.

Audiodateien, die im Rahmen der Content Creation für Vlogging, Podcasting, Bildung, journalistische Recherche oder andere redaktionelle Zwecke entstehen, von Hand in Text zu transkribieren, erfordert nicht nur ein hohes Maß an Konzentration, sondern ist auch sehr zeitaufwendig – und damit teuer. KI-basierte Transkriptionssoftware, die Audio in Text umwandeln, sind daher zu einer großen und produktiven Erleichterung für diesen Arbeitsschritt geworden.

Hier eine kleine Auswahl aus vielen Transkriptionstools:

- ▶ *Microsoft 365*<sup>18</sup>, nur für Word for the Web, Audio via Upload oder direkt in ein Word-Dokument aufgenommen und gespeichert, alle Sprecher werden einzeln getrennt aufgeführt, nach der Besprechung können Teile der Aufnahme durch Abspielen mit einem Zeitstempel versehen und bearbeitet werden.
- ▶ *GoSpeech*<sup>19</sup>, intuitiver Workflow, inklusive Onlineeditor nach Aufnahme. Vorteil: fokussiert auf die deutsche Sprache, läuft DSGVO-konform, deutscher Server
- ▶ *Whisper*<sup>20</sup> von OpenAI, funktioniert in knapp 100 Sprachen, filtert Füllwörter, auf GitHub kostenlos oder auf Server von OpenAI kostenpflichtig für 0,006 Dollar pro Minute
- ▶ *MacWhisper*<sup>21</sup> für Mac, basiert auf Whisper, als Download, verarbeitet Audiodateien lokal, mehrere Transkriptionsmodelle (kostenlos und kostenpflichtig)

---

18 <https://support.microsoft.com/en-us/office/transcribe-your-recordings-7fc2efec-245e-45f0-b053-2a97531ecf57>

19 <https://www.gospeech.com/>

20 <https://openai.com/research/whisper>

21 <https://goodsnooze.gumroad.com/l/macwhisper>

### Workflow: Podcast transkribieren und mehrsprachig übersetzen – mit KI

Mithilfe von ChatGPT und dem DeepL Übersetzer können Sie aufwendig produzierte Audioinhalte wie Podcasts nicht nur barrierefrei, sondern auch über Sprachgrenzen hinweg einem breiteren Publikum zugänglich machen.

- ▶ *Schritt 1:* Erstellen Sie ein Transkript Ihrer Podcast-Episode. Dies können Sie entweder manuell durch einen menschlichen Schreibservice oder eben digital mit der Hilfe eines der beschriebenen KI-basierten Transkriptionstools erledigen.
- ▶ *Schritt 2:* Kontrollieren Sie das Skript, weisen Sie beispielsweise Sprecher zu, korrigieren Sie falsche Wörter und Sätze, eliminieren Sie Füllwörter und Redundanzen.
- ▶ *Schritt 3:* Lassen Sie den Text Abschnitt für Abschnitt vom DeepL Übersetzer übersetzen – und vielleicht von ChatGPT zusätzlich als Bulletpoints zusammenfassen.
- ▶ *Schritt 4:* Lassen Sie die Übersetzung auf Richtigkeit prüfen.
- ▶ *Schritt 5:* Stellen Sie das übersetzte Transkript und das Originalskript auf Ihrer Website und die Zusammenfassung in den Shownotes neben dem Audioplayer zur Verfügung, zusätzlich auch jeweils als PDF zum Download. Damit haben Sie zwar keine übersetzte Audioversion, aber Sie erreichen ein größeres Publikum, das Ihnen für diesen Service dankbar sein wird. So erschließen Sie sich vielleicht auch einen Teil der Gruppe der Nicht-Podcast-Hörer, die lieber lesen statt hören.<sup>22</sup>

### Redaktionelle Texte und Manuskripte mit KI zum Leben erwecken – von Text zu Sprache (TTS)

Als Spiegel-Online-Leser haben Sie sicherlich schon festgestellt, dass Artikel des Magazins inzwischen nicht mehr nur online zu lesen sind, sondern auf Wunsch auch von freundlichen Stimmen vorgelesen werden. Auch wer den Handelsblatt-Newsletter abonniert hat, kann sich die morgendlichen Nachrichten auf dem Weg zur Arbeit vorlesen lassen, wenn man den Blick statt aufs Smartphone auf die Straße richten muss.

Um geschriebene Texte in vorgelesene Audiotexte umzuwandeln, werden jedoch immer weniger professionelle Sprecherinnen und Sprecher eingesetzt. Gerade in Anwendungsbereichen, in denen es nicht auf eine perfekte Sprachdramaturgie ankommt, sondern auf das flüssige Vorlesen vieler und langer Inhalte in einer bestimmten Tonlage, helfen zunehmend KI-Werkzeuge, die geschriebene Inhalte in Sprache umwandeln und damit einem breiteren Publikum in einem anderen Nutzungskontext zugänglich machen. Dies wird als *Text zu Sprache (TTS)* bezeichnet.

<sup>22</sup> Kristina Altfator, The Podcast Landscape: Das sind die Top Learnings aus der Studie. In: Podstars, <https://podstars.de/blog/podcast-landscape-studie/>, 06.09.2023 [02.04.2024]

Diese Anwendungen verarbeiten Text und lesen ihn dann wie ein Mensch vor, ohne dafür ein professionelles Tonstudio zu nutzen. Diese Tools sind einfach zu bedienen, vielseitig einsetzbar und machen Redaktionsarbeit damit um einen Arbeitsschritt reicher: für die Erstellung von Audio- und Video-Tutorials, Businesspräsentationen, das Vorlesen von Teilen des Geschäftsberichts, für die Erstellung von Event-Trailern, Gaming, Animationen, die Entwicklung von Sprachassistenten oder ganzen Hörbuchproduktionen. Oder für Moderationen, der moderierende AI-DJ des Musikstreamingdienstes Spotify, der für seine zahlenden Abonnenten einen auf künstlicher Intelligenz basierenden DJ-Modus eingeführt hat, der das Musikhören noch persönlicher machen soll, indem die Vorschläge von einem KI-Assistenten live moderiert werden.<sup>23</sup>

Mit dem Boom der künstlichen Intelligenz kommt eine Vielzahl von TTS-Generatoren und -Tools. Bei der Suche im Internet stößt man auf entsprechend viele Angebote. Viele nutzen dabei den gleichen TTS-Dienst, etwa den der Microsoft Azure Cloud.<sup>24</sup> Andere wiederum entwickeln eigene Algorithmen und trainieren diese aufwendig. Im Folgenden werden ein paar der herausragenden KI-basierten Tools im Detail vorgestellt:

- ▶ *ElevenLabs*: Die Sprach-KI des Start-ups ElevenLabs<sup>25</sup> ist sicherlich das Flagship der TTS-Tools. ElevenLabs ist darauf spezialisiert, Texteingaben in eine natürlich und menschlich klingende Sprachausgabe umzusetzen. User können aus mehreren Stimmvorlagen die passende auswählen, die den eingegebenen Text sprechen soll. Dazu wurde die Voice-KI mit Hunderttausenden Stunden Audiomaterial trainiert,<sup>26</sup> was sie zu einem besonders beeindruckenden Tool macht: Sie kann menschliche Stimmen erstaunlich realistisch simulieren, inklusive Stimmlage, Sprechrhythmus und korrekter Betonung. Sogar Emotionen kann sie überbringen, lässt Stimmen glücklich, wütend oder traurig klingen – und das über ganze Textabschnitte hinweg. Und im Gegensatz zu anderen Generatoren kann sie Worte ihrem Kontext entsprechend und Abkürzungen korrekt aussprechen. Auch Voice Cloning, die Nachahmung der eigenen oder sogar fremder Stimmen, ist mit diesem Tool möglich. Was allerdings bereits zu einigen Deepfakes geführt hat, bei denen Trolle Promistimmen hochproblematische Dinge haben sagen lassen.<sup>27</sup> Solche Deepfakes stellen sicherlich eine der größten Gefahren im Umgang mit KI und Audio dar, die schleunigst gebannt werden muss.

23 <https://newsroom.spotify.com/2023-02-22/spotify-debuts-a-new-ai-dj-right-in-your-pocket/>

24 <https://azure.microsoft.com/en-us/products/cognitive-services/speech-to-text/>

25 <http://blog.elevenlabs.io/>

26 Vgl. Thomas Richter, Essays: Text zu Sprache per KI: Ersetzen synthetische Stimmen bald professionelle Sprecher? In: slashCAM, <https://www.slashcam.de/artikel/Essays/Text-zu-Sprache-per-KI-Ersetzen-synthetische-Stimmen-bald-professionelle-Sprecher-.html>, 02.02.2023 [20.03.2023]

27 Markus Böhm, Online-Trolle lassen Emma Watson »Mein Kampf« vorlesen. In: Spiegel Netzwelt (paid), <https://www.spiegel.de/netzwelt/web/elevenlabs-stimmengenerator-online-trolle-lassen-emma-watson-mein-kampf-vorlesen-a-780f1457-5a1c-40e0-b909-57835f89125d>, 31.01.2023 [03.02.2023]

### Tipp: Filme synchronisieren mit ElevenLabs AI Dubbing

ElevenLabs bietet viele überraschende und disruptive Features, von denen hier eines gesondert herausgestellt werden soll: *AI Dubbing*. Bei der automatischen Synchronisation oder Videoübersetzung wird der Originalton eines Videos übersetzt und durch eine neue Sprache ersetzt, wobei die einzigartigen Merkmale der Stimmen der Originalsprecher tatsächlich erhalten bleiben.

Die Nutzung ist einfach: Entweder die Videodatei oder den Link zum YouTube-, Vimeo- oder TikTok-Video einsetzen, TARGET LANGUAGE auswählen und auf DUB IT! klicken. Es dauert nur Minuten, und das Video wird in die gewünschte Sprache mit den entsprechenden Stimmen synchronisiert.

Diese Funktion kann für alles genutzt werden, von der Übersetzung von Videoinhalten, um ein breiteres Publikum zu erreichen, über die Bereitstellung von Lernressourcen für Nichtmuttersprachler bis hin zur Synchronisierung von Filmen und Fernsehsendungen für internationale Zielgruppen. Unternehmen können auch mühelos mehrsprachige Versionen ihrer Werbevideos oder Schulungsmaterialien erstellen und so Kommunikationslücken innerhalb einer vielfältigen Kundschaft oder Belegschaft überbrücken.

### ElevenLabs im Überblick

- ▶ *Features und Anwendungen:* Text zu Sprache, Voice Changer, Dubbing, gesprochene Langform-Audios, Voice-Cloning, große KI-Stimmen-Bibliothek, dazu kommen Text zu Sound und Text zu Musik.
  - ▶ *Use Cases:* Vertonen und Übersetzen von Videos mit Voiceovers durch KI-Stimmen, Hörbucherzählung, Games mit KI-Stimmen produzieren, KI-Agenten und Chatbots mit realistischen KI-Stimmen im Unternehmensumfeld, Filme mit Soundeffekten und Musik vertonen, Radio machen (siehe Beispiel bigGPT in Abschnitt 5.3, »Radio machen mit KI – Best Practice: bigGPT«).
  - ▶ *Gebührenmodell:* ElevenLabs gibt es in sechs Preisvarianten. Neben einer kostenlosen Free-Version mit reduziertem Funktionsumfang kostet das Tool zwischen 5 und 330 US-Dollar im Monat.
- ▶ *Murf:* Die KI-gestützte TTS-App<sup>28</sup> wandelt Text in menschlich klingende Stimmen um und unterlegt sie auf Wunsch mit Musik, Bild und Video-Content aus dem eigenen oder einem Stock-Archiv. Die Stimmen sind von denen echter Menschen kaum zu unterscheiden. In der Pro-Version kann man aus einer Vielzahl solcher Stimmen die passende auswählen – sogar mit verschiedenen Akzenten, Tonalitäten und Stilen.

<sup>28</sup> <https://murf.ai>

### Workflow: Eine Podcast-Produktion mit ElevenLabs und Albert Einstein – von Sascha Lobo

In seinem Podcast *Zukunft verstehen. Wie Technik die Welt verändert* diskutiert Sascha Lobo mit sachkundigen und prominenten Gästen, wie Digitalisierung die Welt verändert oder uns vor neue Herausforderungen stellt. »Eine manchmal nerdige, aber immer unterhaltsame Reise in Deutschlands digitale Zukunft«, beschreiben die Macher selbst das Format.<sup>29</sup>

In der Dezember-Ausgabe 2023<sup>30</sup> (siehe Abbildung 5.2) plaudert überraschenderweise Albert Einstein mit Sascha Lobo über unsere digitale Zukunft. Möglich gemacht hat das launige Gespräch generative KI. Sascha Lobo beschreibt den Workflow zur Erstellung dieser speziellen Podcast-Folge wie folgt:<sup>31</sup>

1. »Wir haben eine ChatGPT-Variante (genauer: GPT) gebaut, der wir mit Lerndialogen und hochgeladenen Informationen beigebracht haben, so zu antworten wie Albert Einstein – das EinsteinGPT. [Details zu GPT siehe Kapitel 3, »Texte mit KI schreiben, optimieren und zusammenfassen«.]
2. Dann habe ich Einstein bzw. dem EinsteinGPT Fragen über die digitale Zukunft gestellt, z. B. welche fünf Entwicklungen er für die wichtigsten der nächsten Jahre hält. Da war die Antwort: Quantum Computing, künstliche Intelligenz, digitaler Zwilling, Metaverse und Datensicherheit plus Datenschutz.
3. Für jeden einzelnen Bereich habe ich dann mit Einstein besprochen, worin er genau die Relevanz sieht und was uns konkret bevorsteht. Er hat sehr unterhaltsam und – natürlich – einsteinhaft sachkundig geantwortet. (... An manchen, wenigen Stellen haben wir Einsteins Antworten leicht modifiziert, manchmal durch ›regenerate‹ und manchmal mit zusätzlichen Inhalten.)
4. Und dann haben wir (rechtfreie) Tondokumente von Albert Einstein benutzt, um seine Stimme zu klonen, mithilfe der Plattform ElevenLabs – und schließlich die Antworten von EinsteinGPT mit Einsteins geklonter Stimme vertont.«

Den Zeitaufwand für die Produktion rechnet Lobo wie folgt vor:

- ▶ Training der EinsteinGPT: zwei Stunden, inklusive Startdialog, Fragen/Antworten und Uploads
- ▶ Produktion Inhalte mit EinsteinGPT: acht Stunden

<sup>29</sup> <https://zukunft-verstehen.podigee.io>

<sup>30</sup> <https://zukunft-verstehen.podigee.io/38-albert-einstein-zur-digitalen-zukunft-von-ki-bis-metaverse>, 20.12.2023 [22.12.2023]

<sup>31</sup> Sascha Lobo. In: LinkedIn, [https://www.linkedin.com/posts/saschalobo\\_es-ist-vielleicht-keine-riesensensation-activity-7143321353830543360-Nh\\_c](https://www.linkedin.com/posts/saschalobo_es-ist-vielleicht-keine-riesensensation-activity-7143321353830543360-Nh_c), [08.01.2023]

- ▶ Voice Cloning: sehr lange und zu dritt, um die Einstein-Stimme zu reinigen, zu beschleunigen, zu filtern und zu modulieren
- ▶ Schnitt/Studio: »unendlich lange Arbeit«, um die zu langen Sprechpausen von Einstein einzeln rauszuschneiden



Abbildung 5.2 Sascha Lobo im Gespräch mit Albert Einstein (Screenshot: Apple Podcast)

Neben der Faszination für solche innovativen Einsatzmöglichkeiten von KI wirft das Experiment aber auch ethische und moralische Fragen auf: Wie steht es beispielsweise um den Umgang mit Persönlichkeiten und deren Rechten bei der Erstellung von Inhalten für kommerzielle Zwecke? Auf LinkedIn wurde zu diesem Beispiel diskutiert, inwiefern die Nutzung der Identität Verstorbener für die Simulation KI-gestützter Gespräche das Andenken und die Persönlichkeitsrechte der Person verletzen könnte. Offen zur Diskussion steht die Frage, welche Auswirkungen es haben könnte, wenn fiktive, mit KI generierte Interviewfetzen aus dem Zusammenhang gerissen als »Originalzitat« in anderem Kontext auftauchen.

Hören Sie selbst in dieses interessante Experiment hinein: Da Lobo die Tondokumente von Einstein aus dessen Reden genommen hat, klingt dessen Stimme auf eine unterhaltsame Art und Weise »grammophonig«. Dafür entschuldigt sich Einstein gleich zu Beginn des Gesprächs: »Das lässt sich bei einem Podcast mit Zeitreisenden kaum ändern.« Stimmt.

- ▶ *Listnr*: Listnr<sup>32</sup> wandelt Text in Sprache nach Genre- und Akzentauswahl in verschiedene Formate um. Das Ergebnis kann dann in einen Artikel oder einen Blog in Form eines Audioplayers eingebettet werden, dessen Design und Sharing-Optionen individualisiert anpassbar sind. Listnr kann auch für die Verbreitung und Konvertierung von Audiodateien auf Streamingplattformen wie Spotify und Apple verwendet werden.

Weitere Text-zu-Sprache-Tools, bei denen es sich lohnt, sie im Auge zu behalten, sind Lyrebird<sup>33</sup> und IBM-Watson<sup>34</sup> sowie Voice Engine von OpenAI:

32 <https://www.listnr.tech/text-to-speech/germany-german>

- ▶ *Voice Engine*:<sup>35</sup> OpenAI hat Ende März 2024 sein eigenes Stimmenmodell namens Voice Engine vorgestellt. Damit soll jeder seine Stimme sehr schnell klonen können: Dafür reichen dem Modell 15 Sekunden Originalstimme. Einer der ersten Anwender von Voice Engine ist HeyGen (siehe Kapitel 6, »Video mit KI konzipieren und produzieren«), eine KI-Plattform für visuelles Storytelling, die es Unternehmen und Marken ermöglicht, benutzerdefinierte, menschenähnliche »Humatare« für Videos im Produktmarketing und Verkaufsdemos zu nutzen. HeyGen nutzt Voice Engine für die Videoübersetzung. Damit wird die Stimme eines Sprechers in mehrere Sprachen übersetzt, um mit einem Video ein globales, mehrsprachiges Publikum zu erreichen. Bei der Übersetzung behält Voice Engine sogar den muttersprachlichen Akzent des ursprünglichen Sprechers bei: Wenn z. B. Englisch mit einem Audiobeispiel eines französischen Sprechers generiert wird, entsteht Sprache mit französischem Akzent. Zum Redaktionsschluss war es noch nicht möglich, Voice Engine selbst zu testen – wohl mit Blick auf die Wahlen in den USA und der Verbreitung von Deepfakes hielt sich OpenAI mit der Veröffentlichung noch zurück. Clone-Beispiele zu Voice Engine kann man sich auf dem OpenAI-Blog anhören.<sup>36</sup> Dabei zeigt OpenAI weitere konstruktive Beispiele für die Nutzung des Klon-Tools: zum Erreichen globaler Communities, um Patienten zu helfen, ihre Stimme wiederzuerlangen oder für die Unterstützung von Menschen, die sich nicht verbal ausdrücken können.

Konkurrent ElevenLabs, der schon seit Längerem das Klonen von Stimmen anbietet, hat das Thema Sicherheit bereits adressiert: Bei einer sogenannten *Professional Voice* muss man während des Cloning-Prozesses einen vorgegebenen Text live vorlesen. Dadurch wird sichergestellt, dass man nur die eigene Stimme klonen kann.

#### Workflow: Ein Hörspiel als kreatives KI-Experiment – mit Text zu Sprache<sup>37</sup>

Ein weiteres Fallbeispiel für den Einsatz von ElevenLabs in einem kreativen, experimentellen, aber wohl auch anschaulichen Workflow:

Digital Artist CoffeeVectors hat auf X (ehemals Twitter) in einem sehr hörenswerten Audio-Snippet schon früh gezeigt, wie TTS eingesetzt werden kann, um beispielsweise ein Hörspiel zu produzieren. In der als animierter Tweet präsentierten Hörspielszene

33 <https://www.descript.com/lyrebird>

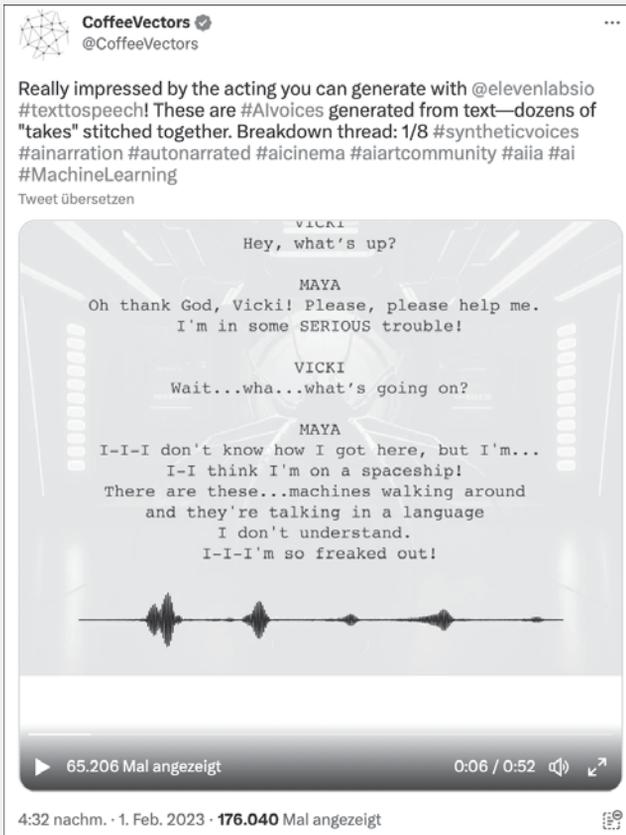
34 <https://www.ibm.com/de-de/cloud/watson-text-to-speech>

35 OpenAI, Navigating the Challenges and Opportunities of Synthetic Voices. In: Blog, <https://openai.com/blog/navigating-the-challenges-and-opportunities-of-synthetic-voices>, 29.03.2024

36 OpenAI, Navigating the Challenges and Opportunities of Synthetic Voices. In: OpenAI Blog, <https://openai.com/blog/navigating-the-challenges-and-opportunities-of-synthetic-voices#OpenAI>, 29.03.2024 [03.04.2024]

37 <https://x.com/CoffeeVectors/status/1620807387073642496?s=20>

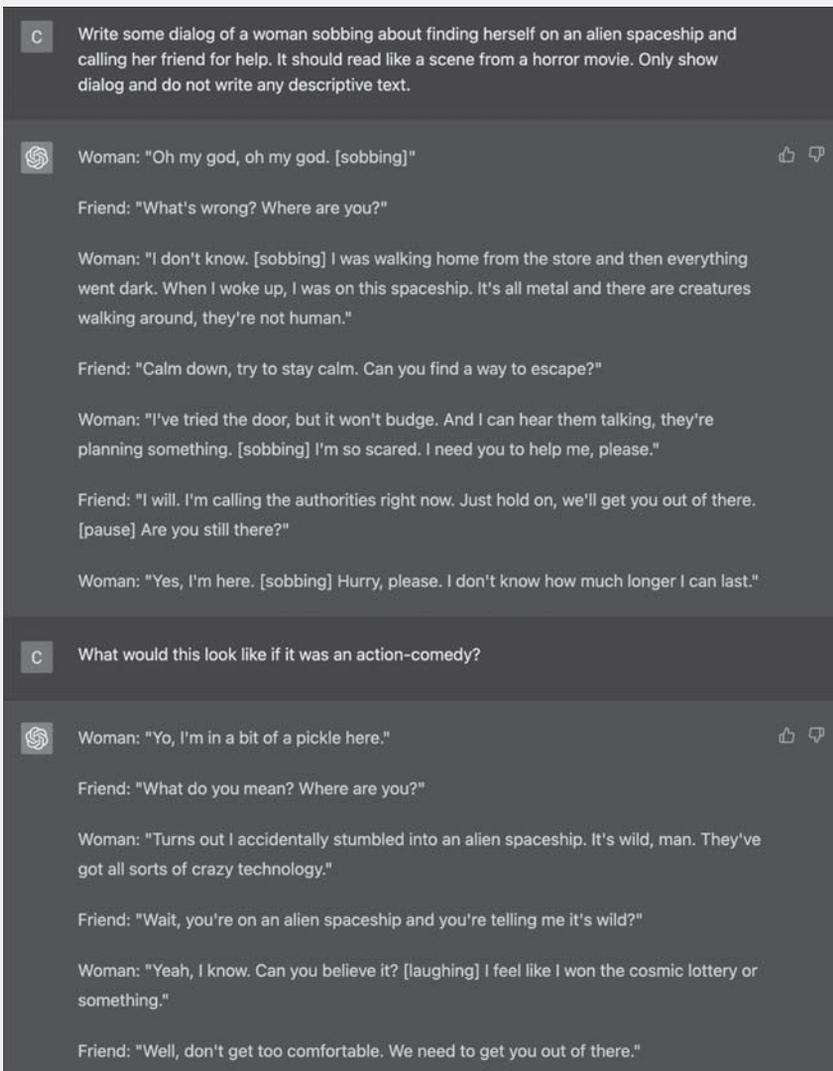
ruft eine *hörbar* in Panik geratene Frau ihre *hörbar* ratlose Freundin aus einem Raumschiff heraus an, um sie um Hilfe zu bitten. Die äußerst lebendigen Stimmen auf den zahlreichen Takes sind KI-Stimmen – generiert aus einem Textskript. Diese Takes wurden anschließend aufwendig bearbeitet und zusammengeschnitten.



**Abbildung 5.3** Text zu Sprache – ein kleines KI-Sci-Fi-Drama zum Anhören (Quelle: X [ehemals Twitter], <https://x.com/CoffeeVectors/status/1620807387073642496?s=20>)

Den entsprechenden Workflow hat die Künstlerin in ihrem Thread wie folgt beschrieben (der besseren Nachvollziehbarkeit im Original belassen):

- »I started by having #chatGPT write a few rough drafts of a scene involving a panicked character calling her friend for help from a spaceship. I was going for something that would involve heightened emotions but not be too serious. 2/8« (siehe Abbildung 5.4).



**Abbildung 5.4** »Schreibe ein paar grobe Entwürfe für eine Szene, in der eine Figur in Panik ihre Freundin aus einem Raumschiff heraus um Hilfe bittet ...« (Quelle: X [ehemals Twitter], ebenda)

- »Then I wrote a short script using some of those ideas plus my own and put the whole thing into @elevenlabsio. I generated a few takes using low Stability (1–2 %) and high Clarity (90–99 %). Each take usually had parts I liked, or at least gave me ideas for direction. 3/8«

- ▶ »I stuck to one voice I liked for simplicity. Changing voices can sometimes dramatically alter the sound to where it almost feels like diff mics were used. I decided I'd just change the pitch of the voices in post to differentiate them more. 4/8«
- ▶ »After doing a few takes of the whole script, I generated individual lines. There I'd experiment with the ›prompt‹ to see if I could direct the acting more by adding ellipses, diff punctuation, line breaks, and misspellings ... 5/8«
- ▶ »Then I laid everything out in #premierepro. I cut up the audio into sections with different takes and methodically edited down to my favorites, trying to choose parts that blended well together. 6/8«
- ▶ »When parts wouldn't blend well together, I'd just rewrite the lines and generate a few more takes in @elevenlabsio. It's almost like instantaneous ADR. Then I used #adobeaudition for shifting the pitch in the voices and adding reverb. 7/8«
- ▶ »Last step was using the script as rolling credits and put it on an image I made in #midjourney. I added the audio wave in After Effects. 8/8«

Wie kreativ *Text zu Sprache* eingesetzt werden kann, zeigt auch das folgende Beispiel, in dem DJ David Guetta für Stimmung in vielerlei Hinsicht sorgte:

#### Workflow: Musik komponieren mit einem »Text-zu-Text-zu-Sample« – Guetta klont Eminem

»Ich habe das aus Spaß gemacht. Und es hat so gut funktioniert, dass ich es kaum glauben konnte«, erzählt der DJ und Musikproduzent David Guetta in einem Tweet,<sup>38</sup> der einen kleinen Videoausschnitt von einem seiner Konzerte Anfang 2023 zeigt (siehe Abbildung 5.5).

In dem Ausschnitt ist zu hören, wie der DJ die markante Stimme des Rappers Eminem live als Sampling in einen seiner Songs einbaut und mit den Worten ankündigt: »Let me introduce you to ... Emin-Al-em.« Beim Sampling wird ein Teil einer bereits fertigen Ton- oder Musikaufnahme in einen anderen Song eingebaut. Diese gesampelte Stimme stammte jedoch nicht original von Eminem, sondern wurde von einer KI im unverkennbaren Stil des Rappers gesprochen. Guetta erklärt seinen kleinen Workflow wie folgt: »Ich habe diese Websites über KI entdeckt. Im Grunde kann man dort Texte im Stil eines beliebigen Künstlers schreiben. Also tippte ich ›Schreibe eine Strophe im Stil von Eminem über Future Rave‹. Dann bin ich zu einer anderen KI-Website gegangen, die die Stimme imitieren kann. Ich gab den Text dort ein, spielte die Platte ab, und die Leute flippten aus!«

38 <https://x.com/davidguetta/status/1621605376733872129?s=20>



**Abbildung 5.5** DJ Guetta experimentiert mit KI und sampelt Eminems Stimme via Voice-Cloning für einen Song. (Quelle: X [ehemals Twitter])<sup>39</sup>

In den Kommentaren zum Video entbrannte daraufhin eine Debatte über die Rechtmäßigkeit der Nutzung solcher geklonten Inhalte. Guetta stellte selbst klar, dass er den Song nicht kommerziell vertreiben werde.

Es gibt inzwischen viele Beispiele, in denen die Stimmen von Künstlerinnen und Künstlern mittels KI imitiert werden. Viele davon finden sich auf YouTube oder TikTok. Ein vermeintlicher Song von Drake und The Weeknd beispielsweise ging bis zu seiner Löschung in sehr kurzer Zeit viral.<sup>40</sup>

Wie man mit künstlicher Intelligenz aber nicht nur Sprache, Fakes und Samples, sondern sogar ganze Musikstücke und faszinierende Soundeffekte für eine Filmproduktion oder andere Formate erstellen kann, darum geht es im nächsten Abschnitt.

<sup>39</sup> <https://x.com/davidguetta/status/1621605376733872129?s=20>

<sup>40</sup> Benjamin Fischer, Ein KI-Duett von Drake und The Weeknd – und die Folgen. In: FAZ, <https://www.faz.net/aktuell/wirtschaft/digitec/ki-generiert-fake-song-von-drake-und-the-weeknd-18829126.html>, 18.04.2023 [25.04.2023]

## 5.2 Musik und Sound kreieren – Text zu Musik und Text zu Sound

Content Creation braucht anspruchsvolle, einzigartige Musik und professionelle Soundeffekte – mit den entsprechenden Nutzungsrechten. Diese sorgen erst für die wichtige hörbare Differenzierung im kommunikativen Auftritt einer Marke. Dieser Qualitätsanspruch gilt nicht nur für visuelle Medien wie TV, Online und Social Media, sondern auch für die Produktion von Audioinhalten wie beispielsweise dem Intro und Outro eines Podcasts, einer Warteschleife, dem Jingle einer YouTube-Serie oder für Games und Hörbücher.

Für aufwendige Markenproduktionen kommen normalerweise Komponisten und Sounddesignerinnen zum Einsatz, die die entsprechende Musik und die Soundeffekte (SFX) individuell produzieren. Im kleineren Maßstab kann man auch auf bestehende Lieder und vorproduzierte SFX zurückgreifen. Dann heißt es, auf die Lizenzanforderungen zu achten.

An dieser Stelle gibt es aber neue Unterstützung für Content Creator, denn neben der Einführung von ChatGPT im Text- inklusive DALL-E 3 und Co. im Bildbereich wird der Einstand von KI sicherlich auch ein Meilenstein für den Einsatz von generativer künstlicher Intelligenz in der Musik- und Tonproduktion werden. Ob Soundoptimierung, Audio- oder Musikerstellung mit Text-Prompt oder per freie Textsuche zum perfekten Song – die Musikproduktion bekommt neue, mächtige Unterstützung. In diesem Abschnitt des Kapitels geht es darum, KI-basierte Tools vorzustellen, die Ihnen helfen können, jetzt und in naher Zukunft Ihren Content mit passender Musik und beeindruckenden Soundeffekten aufzuwerten.

Entsprechende KI-basierte *Text-zu-Musik*-Generatoren (*TTM*) sind auf dem Vormarsch.

### Eigene Songs komponieren – von Text zu Musik

Auch wenn viele Apps und Tools zum Zeitpunkt der Drucklegung dieses Buches noch in der Entwicklung stehen, ist eines sicher: Im Bereich *Text zu Musik* zeichnet sich eine rasante Entwicklung ab, die die *Musikproduktion* grundlegend verändern wird. Bestes Beispiel:

- ▶ *suno.com*: Suno generiert aus Text-Prompts beeindruckend klingende Songs, einschließlich Lyrics, Instrumenten und Gesang, für jedes Genre, von deutschem Schlager bis Rap. Das Komponieren mit Suno ist dabei sehr einfach: Nach der Registrierung unter *suno.com*<sup>41</sup> und dem Klick auf CREATE beschreibt man die Stilrichtung und den Inhalt des gewünschten Musikstücks in Form eines kurzen Prompts. Die App erstellt

---

<sup>41</sup> <https://suno.com>

dann nach einem weiteren Klick auf einen weiteren CREATE-Button zwei kurze Audiostücke mit dem dazugehörigen Text. Um genau zu sein, ist der Song die Arbeit von zwei kooperierenden KI-Modellen: Sunos Modell erstellt die Musik, während OpenAIs ChatGPT den Text und einen Titel beisteuert. Seinen Song kann man dann sogar im gleichen Stil mit einfachem Klick auf EXTEND verlängern lassen und ihn im eigenen Profil mit der ganzen KI-musikinteressierten Welt teilen.

Auch Rolling Stone Redakteur Brian Hiatt hat auf diese Weise seinen ersten eigenen Wunschtitel gepromptet:

*Beispiel-Prompt: »Solo acoustic Mississippi Delta blues about a sad AI.«*

Das Ergebnis seines Stückes mit dem Namen ›Soul of the Machine‹, das Suno nach diesem Prompt geschrieben hat, beschreibt er – offensichtlich beeindruckt – so: »The voice singing those lyrics is raw and plaintive, dipping into blue notes. A lone acoustic guitar chugs behind it, punctuating the vocal phrases with tasteful runs ... ›Soul of the Machine‹ feels like something different – the most powerful and unsettling AI creation I've encountered in any medium.«<sup>42</sup>

Möchte man mehr Kontrolle über die Komposition, kann man im CUSTOM-Modus zunächst angeben, in welchem Musikstil das Stück komponiert werden und welchen Namen es tragen soll. Zusätzlich lassen sich auch LYRICS eingeben. Das ergibt insbesondere dann Sinn, wenn man einen eigenen Songtext geschrieben oder einen solchen im Pingpong mit ChatGPT eigenständig erarbeitet hat. Denn das individuelle Zusammenspiel zwischen Creator und KI ist beim Texten mit Suno (noch) nicht vorgesehen:

#### Workflow: Song-Komposition mit ChatGPT und Suno

- ▶ *Schritt 1:* Zunächst erstellen Sie einen Songtext mit ChatGPT:  
*Beispiel-Prompt ChatGPT:* »Rap über Second Hand Emotion mit KI.«
- ▶ *Schritt 2:* Der so erstellte Text wird kopiert und in Suno weiterverarbeitet:  
*Folge-Eingabe in Suno:* Menüpunkt CREATE • CUSTOM MODE • LYRICS [Copy-Paste Text aus ChatGPT] • STYLE OF MUSIC [Rap] • CREATE
- ▶ *Schritt 3:* Die so komponierten Songs lassen sich dann herunterladen oder direkt per Link teilen. Reicht einem die Länge des Tracks nicht, da Suno die Komposition nach

<sup>42</sup> Brian Hiatt, A ChatGPT for Music Is Here. Inside Suno, the Startup Changing Everything. In: Rolling Stone, <https://www.rollingstone.com/music/music-features/suno-ai-chatgpt-for-music-1234982307/>, 17.03.2024 [04.04.2024]

den ersten 80 Sekunden unterbricht, lässt sich das Stück im Menü unter CONTINUE THIS SONG ganz einfach ein gutes Stück verlängern.

Investor Antonio Rodriguez beschreibt Suno als ein radikal leistungsfähiges und einfach zu bedienendes Musikinstrument. Er glaubt, dass Suno Musik für jedermann zugänglich machen könnte, so wie Fotohandys und Instagram die Fotografie demokratisiert haben. Die Idee, so Rodriguez, sei es, »mehr Menschen die Möglichkeit zu geben, im Internet etwas zu schaffen, anstatt nur Dinge im Internet zu konsumieren«. <sup>43</sup> Also dann: Probieren Sie es aus. Denn das Ganze kostet zunächst nichts, es sei denn, Sie möchten die kostenpflichtige Version für noch mehr Features und mehr Credits nutzen. Die Reaktionen Ihres Publikums auf die so entstandenen Stücke werden von Erstaunen bis Schock reichen. Versprochen. Denn plötzlich können Sie auch ohne Noten und Instrumente beeindruckende Songs komponieren (lassen).

Und Kritik? Die gibt es natürlich reichlich. Die Gründer von Suno lehnen es beispielsweise ab, Details darüber zu verraten, mit welchen Daten sie ihr eigenes Modell trainieren. In der Musikindustrie stößt das auf großes Unverständnis: Etwa 250 US-Künstlerinnen und Songwriter haben einen Online-Aufruf gegen generative KI-Tools unterzeichnet, die ihre Arbeit ohne Genehmigung nutzen, um Modelle zu bilden, die mit ihnen konkurrieren. Die Liste der Unterzeichner umfasst Billie Eilish, Nicki Minaj, Elvis Costello, R.E.M., Mumford & Sons, Katy Perry, Pearl Jam und viele andere. <sup>44</sup>

### Gebührenmodell

Suno AI bietet ein gestaffeltes Preismodell, einschließlich eines kostenlosen Plans von 50 Credits, der sich täglich erneuert, damit die Benutzer die grundlegenden Funktionen erkunden können. Microsoft hat Suno außerdem via Plug-in in sein KI-Tool Copilot integriert.

Weitere KI-Musiktools, die hörensenswert sind:

- ▶ Größter potenzieller Konkurrent von Suno ist eine junge Musik-KI namens *Udio*. <sup>45</sup> Sie funktioniert ähnlich wie Suno: Aus einer kurzen Beschreibung des Songs, z. B. Folk oder Heavy Metal, und der Möglichkeit, einen Songtext generieren zu lassen oder eigene Texte einzugeben, generiert das Tool Songs, die man sich wirklich gut

<sup>43</sup> Ebenda

<sup>44</sup> Da Billie Eilish a Elvis Costello, 250 artisti americani contro l'IA. In: La Repubblica, [https://www.repubblica.it/spettacoli/musica/2024/04/02/news/da\\_billie\\_eilish\\_a\\_elvis\\_costello\\_250\\_artisti\\_americani\\_contro\\_ia\\_generativa-422411629/](https://www.repubblica.it/spettacoli/musica/2024/04/02/news/da_billie_eilish_a_elvis_costello_250_artisti_americani_contro_ia_generativa-422411629/), 02.04.2024 [04.04.2024]

<sup>45</sup> <https://www.udio.com>

anhören kann. Jeder Prompt wird zunächst mit zwei ca. 30 Sekunden langen Song-Varianten in Stereo-Qualität beantwortet. Jeder Song kann stilistisch dann Stück für Stück stimmig auf bis zu drei Minuten Playtime verlängert werden – nach vorn mit einem Intro, in der Mitte mit zusätzlichen Abschnitten oder mit einem Outro. Das geht wirklich ganz einfach. Mit Markierungen in den Lyrics wie [Verse], [Chorus] oder [Outro] lässt sich das Stück eindrucksvoll strukturieren. Ist das nun nur eine nette Spielerei oder vielleicht auch eine willkommene Inspiration für das eine oder andere eingängige Riff einer Band aus Fleisch und Blut – oder am Ende sogar die disruptive Konkurrenz für eine ganze Musikindustrie? Das ist die gute Frage. Für eine Antwort ist es wohl am besten, Sie komponieren einfach selbst einmal ein Lied mit Suno oder Udio – und staunen selbst über »Ihr« Meisterwerk!

- ▶ Googles *DreamTrack* erlaubt YouTubern, eigene Songs von berühmten Sängerinnen und Sängern, die dafür ihr Einverständnis gegeben haben, singen zu lassen. Aber schon dieses Feature wirkt im Vergleich zu Suno stark limitierend auf den kreativen Freiraum.
- ▶ *Jukebox*<sup>46</sup>: Jukebox ist ein Projekt von OpenAI, ein neuronales Netz, das Musik, einschließlich rudimentärem Gesang, als Rohdaten in einer Vielzahl von Genres und Künstlerstilen erzeugt. Einfach mal Reinhören.
- ▶ *Boomy*<sup>47</sup>: Man kann eigene Songs in Sekunden erstellen – spielerisch.
- ▶ *AIVA*: 2016 gegründet, gilt AIVA<sup>48</sup> als erster anerkannter virtueller Komponist, spezialisiert auf die Komposition klassischer und symphonischer Musik. Mit wenigen Klicks lassen sich KI-generierte Songs erstellen. Seit Januar 2019 bietet AIVA ein kommerzielles Produkt namens *Music Engine* an, mit dem man bis zu dreiminütige Kompositionen in verschiedenen Stilrichtungen erzeugen kann. Die Auswahl an verfügbaren Musikgenres ist riesig. Die generierten Songs kann man in einem Editor öffnen und die Instrumentalspuren noch einmal einzeln nachbearbeiten. Außerdem lassen sich eigene Songs oder MIDI-Noten hochladen, um die KI anhand eigener Werke anzupassen.

### Das passende Audio für Filme, Hörbücher und Spiele komponieren – von Text zu Musik

Eine künstliche Intelligenz anhand einer Textvorgabe kommerziell nutzbare Musikstücke komponieren lassen? Das ist keine Zukunftsmusik mehr.

<sup>46</sup> <https://openai.com/research/jukebox>

<sup>47</sup> <https://boomy.com>

<sup>48</sup> <https://www.aiva.ai>

- ▶ *Stable Audio 2*: Stable Audio<sup>49</sup> von Stability AI generiert auf Knopfdruck Audioinhalte, die sich auch kommerziell nutzen lassen. Mit dem entsprechenden Prompt entstehen Musikstücke, Soundkulissen oder einzelne Instrument-Spuren. Perfekt, um Videos nachträglich mit Musik und Sound Effects zu untermalen.

Mit dem neuen Stable Audio 2 lassen sich bis zu drei Minuten lange Musikstücke mit einer zusammenhängenden Struktur – Intro, Hauptteil und Outro – in Studioqualität mit 44,1 kHz Stereo produzieren. Dazu bietet die KI Funktionen wie Audio-zu-Audio, Geräusch- und Effektgenerierung sowie Stilanpassung. Stable Audio 2 soll angeblich ausschließlich mit Daten aus der vollständig lizenzierten Musikbibliothek AudioSparx trainiert worden sein. Laut Stability AI hatten die Künstlerinnen und Künstler die Möglichkeit, dem Training zu widersprechen. Ganz unumstritten ist diese Sachlage allerdings nicht.

Mit der neuen *Audio-zu-Audio*-Funktion können Nutzer auch eigene Audiodateien hochladen, um daraus komplett produzierte Samples zu erstellen. Außerdem lassen sich mit Stable Audio 2 Sound- und Audioeffekte erzeugen.

Stable Audio 2 kann man auf der Website kostenlos ausprobieren.<sup>50</sup> Auf dem Stable-Audio-YouTube-Kanal können Sie sich zudem den 24/7-Livestream »Stable Radio« anhören.<sup>51</sup> Chillig.

### Gebührenmodell

Verfügbar sind eine kostenlose Version, beschränkt auf 20 generierte Stücke pro Monat und eine Dauer von 20 Sekunden, und eine Bezahlversion für 12 US-Dollar pro Monat.

### Wertvolle Prompt-Tipps für Stable Audio<sup>52</sup>:

Geben Sie klare und präzise Details im Prompt an:

- ▶ **Genre und Atmosphäre:** Definieren Sie genau, welches musikalische Genre und welche Atmosphäre Sie generieren möchten, indem Sie genrespezifische Begriffe und beschreibende Adjektive verwenden, wie z. B. »Cinematic Soundtrack für ein Wildwest-Duell bei High Noon, unterlegt mit Percussion und Pfeifklängen«.
- ▶ **Gewünschte Stimmung:** Kombinieren Sie musikalische Eigenschaften (z. B. »groovy« oder »rhythmisch«) mit emotionalen Ausdrücken (wie »traurig« oder »schön«), um die gewünschte Stimmung präzise zu beschreiben.

49 <https://www.stableaudio.com/>

50 Ebenda

51 <https://www.youtube.com/watch?v=yvOXZ6SV2Rk>

52 <https://stableaudio.com/user-guide/text-to-audio>

- ▶ **Instrumentenauswahl:** Ergänzen Sie Instrumentennamen mit Adjektiven, um ihre Klangfarbe zu verdeutlichen, wie z. B. »hallige Gitarre«, »kraftvoller Chor« oder »anschwellende Streicher«.
  - ▶ **Tempo festlegen:** Geben Sie die gewünschten Beats pro Minute (BPM) an, um das Tempo Ihrer Komposition zu steuern. Achten Sie darauf, dass die BPM-Zahl zum gewählten Genre passt, wie beispielsweise 170 BPM für Drum-and-Bass.
- ▶ *Google MusicLM:* Google hat bereits Anfang 2023 eine Website veröffentlicht, auf der die Ergebnisse des eigenen MusicLM-Projekts vorgestellt werden. Es zeigte erstmals, wie man mit der Eingabe von Textvorgaben neue Musikstücke erzeugen kann. MusicLM<sup>53</sup> ist ein KI-gestütztes Tool, das mit einem Datensatz von 280.000 Stunden Musik trainiert wurde, um Musikstücke nach komplexen Beschreibungen zu generieren.<sup>54</sup> Zwar ist es noch nicht möglich, die Technologie wie bei DALL-E oder ChatGPT selbst auszuprobieren, aber Sie können und sollten sich Googles eigene Tests anhören. Diese zeigen, wie eingängige Sound- und Musikbeispiele aus Text-Prompts generiert werden – beispielsweise der Soundtrack eines Arcade-Spiels.

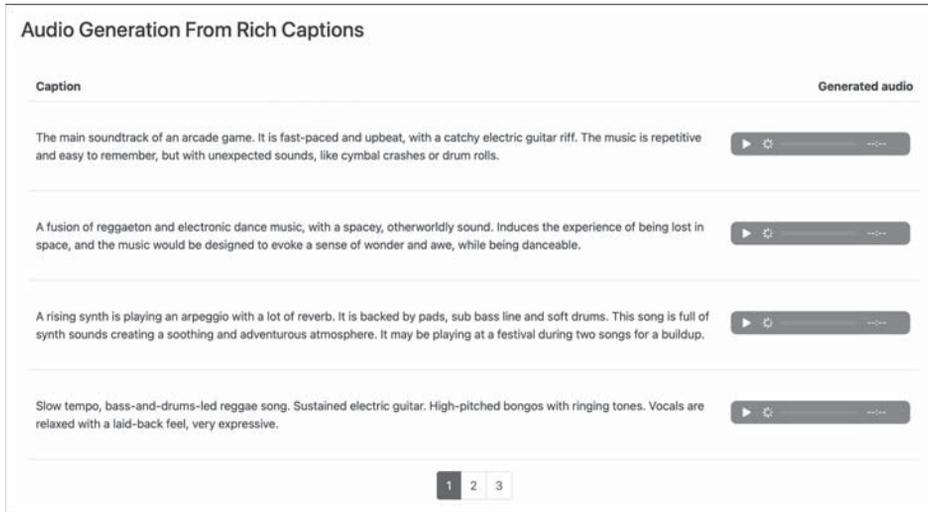
*Beispiel-Prompt (siehe Abbildung 5.6):* »The main soundtrack of an arcade game. It is fast-paced and upbeat, with a catchy electric guitar riff. The music is repetitive and easy to remember, but with unexpected sounds, like cymbal crashes or drum rolls.«

Das entsprechend generierte Musikbeispiel klingt genauso. Hören Sie sich die Snippets auf der Website selbst an. Die Bandbreite der mit MusicLM erzeugten Musik ist erstaunlich. Sie reicht von Volksmusik und Klassik über Jazz, Pop, Rap und Reggae bis hin zu Techno, 8-Bit-Computermusik und Death Metal.

Besonders interessant für Content Creator ist die Funktion *Story Mode*. Sie ermöglicht es, genaue Zeitstempel in den Prompt einzufügen und mit entsprechenden Anforderungen zu versehen, um während des Abspielens der Musik sowohl den Stil als auch den Rhythmus zu ändern. So kann das gewünschte Lied z. B. in den ersten 15 Sekunden meditativ sein, in den nächsten 15 Sekunden anregend, gefolgt von einer schnellen Phase, um sich dann wieder im Schrittempo einzupendeln:

<sup>53</sup> <https://google-research.github.io/seanet/musiclm/examples/>

<sup>54</sup> Thomas Richter, Erst Bilder, dann Sounds: Neue Google-KI generiert beliebige Musik nach Textbeschreibung. In: slashCAM, <https://www.slashcam.de/news/single/Erst-Bilder--dann-Sounds--Neue-Google-KI-generiert-17701.html>, 30.01.2023 [24.02.2023]



**Abbildung 5.6** Text-to-Music: Prompt-Beispiele, veröffentlicht für MusicLM von Google (Screenshot GitHub)<sup>55</sup>

*Beispiel-Prompt:*<sup>56</sup> »time to meditate (0:00–0:15), time to wake up (0:15–0:30), time to run (0:30–0:45), time to give 100 % (0:45–0:60)«

Darüber hinaus nennt Google Beispiele, in denen eine Melodie als Audio-Prompt gesummt oder gepfiffen wird, in einem Fall die bekannte Melodie aus der Netflix-Serie »Haus des Geldes«. Diese Vorlage wird dann auf Wunsch des Nutzers von der KI in verschiedene Instrumentierungen umgewandelt: von A-cappella-Chor über elektronische Soli und Jazz mit Saxofon bis hin zu Operngesang. Letzteres, wie übrigens jede Form des erzeugten Gesangs, ist in diesem Modell noch gruselig und zeugt von Verbesserungspotenzial der KI. An diesem Beispiel sieht man auch, wie schnell sich KI in den letzten Monaten entwickelt hat. Stable Audio, Suno und Udio sind inzwischen wesentlich weiter fortgeschritten und überraschen geradezu mit ihren stimmlichen Qualitäten.

Dass Google dieses Projekt nicht der Öffentlichkeit zugänglich gemacht hat, hat mehrere Gründe: In Tests waren etwa 1 % der generierten Musikstücke einfache Kopien des Trainingsmaterials.<sup>57</sup> Außerdem könnte das Tool bei der Musikgenerierung aufgrund unzureichender Trainingsdaten eine mögliche Unterrepräsentation bestimm-

<sup>55</sup> <https://google-research.github.io/seanet/musiclm/examples>

<sup>56</sup> Ebenda

<sup>57</sup> Jörn Brien, MusicLM: Neue Google-KI erzeugt ganze Musikstücke auf Zuruf. In: t3n, <https://t3n.de/news/musiclm-google-musik-ki-erste-einblicke-1530956/>, 01.02.2023 [17.03.2023]

ter Stile oder Kulturen widerspiegeln, so die Projektverantwortlichen. Hier stellen sich auch ethische Fragen wie die nach einer möglichen kulturellen Aneignung.

Generative KI wird zukünftig nicht nur bestimmte Musikstile, sondern auch die Stimmungen für Songs auf Befehl übernehmen. Auch das Komponieren von Liedern auf Basis einer Bildvorgabe (*Bild zu Musik*) wird möglich sein. Mit *MuseNet*<sup>58</sup> arbeitet OpenAI, das Unternehmen hinter ChatGPT, übrigens schon an einer ähnlichen Lösung.

Auch Video-Tools bieten mittlerweile musikalische Zusatzservices. Eine Funktion des Video-Generators *RunwayML* (siehe Kapitel 6, »Video mit KI konzipieren und produzieren«) ist beispielsweise die automatische Anpassung des KI-generierten Videos an den Beat der Musik. Dadurch wird das Video an Tempo und Rhythmus des Soundtracks synchronisiert.

Musikproduzent Thomas Foster schätzt die zukünftigen möglichen Auswirkungen von KI auf die Musikproduktion als beträchtlich ein. Wie das Internet ein integraler Bestandteil des täglichen Lebens geworden sei, so werde KI unweigerlich die Landschaft der Musikproduktion verändern.<sup>59</sup> Die zahlreichen Audiobeispiele machen die zukünftigen Einsatzmöglichkeiten dieser KI in der Content Creation konkret absehbar: Filmemacher könnten beispielsweise beschreiben, welche Art von Musik sie in einem Videoprojekt hören möchten. Jemand, der einen Film in Premiere oder Final Touch schneidet, könne direkt in der Timeline seines Videoprogramms entsprechende Marker mit Prompts setzen, die die Musik beschreiben, und die KI könnte die Musik in Echtzeit produzieren.

Generative KI bedeutet, dass sich die Rollen in der Musikproduktion ändern werden: Sängerinnen und Sänger werden vielleicht keine Produzenten mehr beauftragen, ihre Playbacks zu erstellen. Und – weniger erfreulich – Streamingplattformen könnten damit beginnen, ihre Playlists um Titel zu erweitern, die vollständig von KI erstellt wurden. Dennoch gebe es keinen Zweifel daran, dass KI neue Möglichkeiten für Kreativität und Zugänglichkeit eröffnen werde, so Foster.

### Workflow: Lizenzfreie Musikproduktion und Mastering mit KI

Es gibt bereits KI-Tools, die den gesamten Musikproduktionsprozess entlang – von der Konzeption eines Songs bis zum Mastering – eingesetzt werden können. Diese Tools verarbeiten kleine Dateien anstelle ganzer MP3-Dateien. *Audiocipher*<sup>60</sup> und *Soundraw*<sup>61</sup>

58 <https://openai.com/research/musenet>

59 Thomas Foster, Google Music LM – die AI kreiert Musik aus Text, YouTube, [https://www.youtube.com/watch?v=sEL-X5v\\_dDQ](https://www.youtube.com/watch?v=sEL-X5v_dDQ), 05.02.2023 [17.03.2023]

60 <https://www.audiocipher.com/>

61 <https://soundraw.io/>

sind KI-Tools, die die dafür notwendigen MIDI-Dateien erstellen (*MIDI* bedeutet *Musical Instrument Digital Interface*). Solche MIDI-Dateien bestehen aus Befehlen an den Computer, wann er welche Note in welcher Höhe wie lang und mit welcher Intensität ausspielen soll. Diese Dateien sind ideal geeignet für die Verarbeitung durch eine KI.<sup>62</sup>

Wird diese KI mit lizenzfreiem Material trainiert (Beispiel: klassische Musik), sind die damit neu komponierten bzw. errechneten synthetischen Songs sogar lizenzfrei und lassen sich beispielsweise als »Fahrstuhlmusik« an Orten einsetzen, an denen nur ein Soundteppich gewünscht ist. Solche Tools dürften sogar Portale für den Kauf lizenzfreier Musik überflüssig machen.

Auch das Mastering, also die Abstimmung der einzelnen Instrumente für die Endproduktion, verändert sich durch den Einzug der KI. Bisher erforderte das Mastering nicht nur gutes menschliches Gehör, sondern auch aufwendiges Equipment. Diese Rollen können nun KI-Tools wie *Emastered*<sup>63</sup> oder *Songmastr*<sup>64</sup> übernehmen.

### Für jede Stimmung das passende Musikstück mit KI – von Text zu Musikstück

Sicher haben auch Sie Szenen aus Filmen, Serien oder Dokumentationen im Kopf, die Sie so schnell nicht vergessen werden: vielleicht der Fernsehkoch Tim Mälzer mit seiner »Kitchen Impossible« an einem verregneten Steg irgendwo in den nebligen Weiten Skandinaviens, und passend zur Szene läuft ein völlig unbekanntes, aber fantastisch zur Stimmung passendes mystisches Musikstück, das die Atmosphäre wiedergibt oder sogar verstärkt. Szenen wie diese erzeugen regelmäßig Gänsehautmomente und bringen die Fan-Community regelmäßig zum Schwärmen.

Ein solches Musikstück zu finden, ist kreativer Anspruch, aber auch eine große Herausforderung für jeden Content Creator und Produzenten. Denn das setzt voraus, dass entweder in der Musikredaktion ein wandelndes menschliches Songarchiv sitzt (eine aussterbende Spezies) oder der Song in einer Bibliothek mit entsprechenden Tags versehen wurde, die der Redakteur oder die Produzentin kennen und nach denen man suchen kann. Nicht einfach.

Aber was wäre, wenn der Video Creator selbst mit szenischen Bildbeschreibungen eine Suche nach der passenden Musik erfolgreich gestalten könnte – also mit Prompts wie »Ein sonniger Tag im Wald« oder »Eine belebte Stadt in Nordamerika, überfüllt mit

62 Vgl. Stefan Mey, Wie künstliche Intelligenz heute schon bei der Musikproduktion hilft. In: Der Standard, <https://www.derstandard.de/story/2000143439620/wie-kuenstliche-intelligenz-heute-schon-bei-der-musikproduktion-hilft?ref=rss>, 12.02.2023 [19.03.2023]

63 <https://emastered.com/en>

64 <https://www.songmastr.com/>

Menschen in düsterer Stimmung, umgeben von Wildnis und Verfall, eine präapokalyptische Kulisse, die bald vom Meer eingenommen wird«.

Eine komplette Szenenbeschreibung, ein Synchro-Briefing oder auch nur eine Idee für einen Sound – eine entsprechende Freitextsuche, die eine Liste passender Tracks liefert, ist keine Zukunftsmusik mehr: KI-basierte Technologie wird es jedem Content Creator ermöglichen, unabhängig von musikalischen Vorkenntnissen in große Musikbibliotheken einzutauchen, um die perfekte Musik für eine bestimmte Anwendung wie einen Filmtrailer oder ein YouTube-Video zu finden. Dies reduziert die Komplexität der Musiksuche erheblich und erhöht die kreative Wirkung von Inhalten um ein Vielfaches.

Die obigen Beispiele für übersetzte Prompts stammen von Cyanite, einem Technologieunternehmen, das Softwarelösungen für KI-gestütztes Tagging und die Musiksuche anbietet.

- ▶ *Cyanite*<sup>65</sup>, ein Start-up aus Mannheim und Berlin, hat eine neue KI-basierte Suchmaschine entwickelt, die komplexe Texteingaben sofort in das nächstliegende musikalische Äquivalent übersetzen kann. Menschen können nun die passende Musik finden, indem sie der KI sagen, was sie benötigen – in ihrer eigenen Sprache und ohne an eine vorgegebene Reihe von Schlüsselwörtern gebunden zu sein – vorausgesetzt, sie verfügen über eine eigene oder angeschlossene große Musikbibliothek, mit der das Tool gefüttert wird.<sup>66</sup>

Es gibt noch weitere Einsatzbereiche solcher TTM-Tools. Streamingdiensten erlauben sie, das Musikerlebnis eines jeden Hörers mit entsprechenden Vorgaben zu personalisieren: Dabei berücksichtigen sie bei der Auswahl der Songs aus ihren Songbibliotheken sowohl Klangfaktoren als auch persönliche Vorlieben und Stimmungen, um gezielte Songempfehlungen zu geben – und binden auf diese Weise ihr Publikum an die Plattform.<sup>67</sup>

- ▶ *beatoven*<sup>68</sup> ist eine Musikplattform, auf der man mithilfe von KI-Technologie lizenzfreie, einfache Background-Musik für Videos erstellen kann. Die Plattform ist einfach aufgebaut. Man gibt seinem Projekt einen Namen und legt die Länge und das Tempo des Titels fest. Dann wählt man ein Genre oder einen Stil (Pop, Cinematic usw.) und die Stimmung, die man vermitteln möchte. Im nächsten Schritt wird der lizenzfrei zu

<sup>65</sup> <https://cyanite.ai>

<sup>66</sup> Vgl. auch: Silke Hahn, 2023, das Jahr der Musik-KI: Text-to-Audio und per Textsuche zum perfekten Song, <https://www.heise.de/news/Cyanite-finde-den-perfekten-Song-Musik-Suchmaschine-erschliesst-Archive-mit-KI-7481092.html>, 04.02.2023 [01.03.2023]

<sup>67</sup> <https://cyanite.ai/2023/02/01/pr-cyanite-launches-technology-that-can-find-music-based-on-full-text-for-the-first-time/>

<sup>68</sup> <https://www.beatoven.ai>

nutzende Track komponiert. Im anschließenden Editor lassen sich Stimmung, Volume und Instrumente weiter modifizieren und auf das hochgeladene Videomaterial anpassen.

### Für jede Szene die passende Geräuschkulisse – von Text zu Sound

Im Januar 2023 stellte ein Forscherteam der University of Surrey und des Imperial College ein weiteres richtungsweisendes Projekt vor, das vor allem für Creator, die auf Film-Content setzen, in naher Zukunft wichtig werden könnte.<sup>69</sup>

- ▶ *AudioLDM*<sup>70</sup> synthetisiert nicht nur – ähnlich wie MusicLM – Musikstücke mit Instrumenten per Text-Prompt, sondern auch Soundeffekte. So kann ein Creator mithilfe von AudioLDM komplette Geräuschkulissen für die Vertonung von Filmen produzieren. Die beim Hosting-Dienst GitHub vorgestellten Funktionen sind:
  - *Sampling*: Samples für Musik oder Sound gibt es in kurzer oder langer Form, generiert auf Basis von entsprechenden Text-Prompts.

*Beispiel-Prompt*: »Cat gently meowing and purring in long intervals.«

- *InPainting*: Ein Teil einer Audioaufnahme kann per Text-Prompt durch einen anderen Sound ersetzt werden.
- *Styletransfer*: Eine Melodie kann von mehreren Instrumenten gespielt werden.
- *Super Resolution*: Bei einer Aufnahme mit geringer Auflösung wird die Audioqualität durch Upsampling erhöht.
- *Sound-Parameter*: Töne, die generiert werden, lassen sich durch ergänzende Soundparameter variieren, wie z. B. die Art der akustischen Umgebung (Hall), welche Art Objekte die Töne machen sollen (etwa das Schneiden einer Tomate auf Holz, auf Blech usw.) sowie die zeitlichen Reihenfolgen.

*Beispiel-Prompt*: »A female is speaking followed by footstep sound.«

Man kann sich vorstellen, wie TV-Dokumentationen, Heimvideos, YouTube-Produktionen, Social-Media-Clips und sogar Blockbuster mit solchen Prompts auf der Tonspur im Videoschnittprogramm nachträglich optimiert werden. Ein entsprechender Marker mit Prompt wird auf die Tonspur gesetzt, und die KI errechnet dann

<sup>69</sup> Thomas Richter, Erst Bilder, dann Sounds: Neue Google-KI generiert beliebige Musik nach Textbeschreibung. In: slashCAM, <https://www.slashcam.de/news/single/Erst-Bilder--dann-Sounds--Neue-Google-KI-generiert-17701.html>, 30.01.2023 [01.03.2023]

<sup>70</sup> <https://audioldm.github.io>

in Echtzeit den entsprechenden Sound, abgestimmt auf den Verlauf des Inhalts, beispielsweise für rasende Verfolgungsjagden, die raunende Hintergrundkulisse einer Bar, das ländliche Hundebellen oder Vogelgezwitscher ...

### Workflow: Der unbekannte Sound des Weltalls – ChatGPT × AudioLDM

Die Forscher zeigen auch, wie man ChatGPT Prompts für AudioLDM schreiben lassen kann, um komplexere und sogar noch nie gehörte Sounds zu kreieren:

*Beispiel-Prompts:*

- ▶ Text-Prompt für ChatGPT: »Describe the sound of the outer space.«
- ▶ Text-Prompt von ChatGPT für AudioLDM: »Radio emissions from stars, planets, galaxies and other celestial bodies, high fidelity, as well as the sounds of solar winds and cosmic rays.«

Es lohnt sich, sich das Ergebnis dieser Prompt- und Soundexperimente auf der Website anzuhören. Sie eröffnen eine faszinierende Welt KI-generierter Soundeffekte, mit denen Content Creator ihre Workflows für Videoproduktionen sicherlich verändern und bereichern können.<sup>71</sup>

Das anfänglich noch innovative Forschungsprojekt hat allerdings inzwischen erste und vor allem anwendbare Konkurrenz gefunden:

- ▶ *Pika* ist eine Video-KI, mit der sich aus einfachen Prompts bzw. Bildern Video-Snippets kreieren lassen (vgl. Kapitel 6, »Video mit KI konzipieren und produzieren«). Auch diese KI bietet inzwischen eine Funktion, mit der sich KI-generierten Videos per Prompt ein Soundeffekt hinzufügen lässt. Dazu muss der User nur den entsprechenden Schalter aktivieren und einen den Sound beschreibenden Prompt eingeben.
- ▶ Auch das TTS-Start-up *ElevenLabs* möchte sich nicht mehr nur auf die Generation von Sprache beschränken, sondern mischt in diesem Segment der direkt aus Text generierten Soundeffekte kräftig mit. Darüber hinaus sollen mit dem Tool auch einfache Musikstücke generiert werden können.

Mit allen in diesem Kapitel vorgestellten Audio-KIs können Content Creator ihre Inhalte immer lebendiger gestalten. Das Innovationsrennen im KI-Audio-Markt zeigt eindrücklich: KI hat das Potenzial, die ganze Medienlandschaft grundlegend und rasant zu verändern. Wie sehr, das zeigt das folgende Beispiel von RadioGPT.

---

<sup>71</sup> <https://audioldm.github.io>

### 5.3 Radio machen mit KI – Best Practice: bigGPT<sup>72</sup>

Wie sehr generative KI die Audiolandschaft revolutionieren wird, zeigt nicht zuletzt die Tatsache, dass auch der Klassiker unter den Audioformaten, das Radio, durch den kombinierten Einsatz von den in diesem Kapitel vorgestellten KI-Technologien neue Facetten der Produktionsmöglichkeit bekommt (siehe Kapitel 10, »Neue Kreativberufe durch generative KI«).

Am 8. August 2023 um 8 Uhr ging in Mannheim erstmals *Radio bigGPT* auf Sendung. Audiotainment Südwest, ein Mediaunternehmen, bekannt für seine Senderfamilie bestehend aus Radio Regenbogen, RPR1 und bigFM, erweiterte damit sein Programm mit Inhalten, bei dem Moderation und Musikauswahl erstmals ausschließlich von KI übernommen werden. Zu empfangen ist bigGPT seitdem im Stream<sup>73</sup> und über die mobile App.<sup>74</sup>

Die Macher haben bigGPT nicht gestartet, um klassisches *Radioprogramm* durch KI-Inhalte zu ersetzen. Vielmehr bildet die Idee, bigGPT als *Lernlabor* zu betrachten, den Kern der Projekt-DNA: Kontinuierliches Lernen und Anpassen in Sachen Content Kreation mit KI stehen im Vordergrund. Daher wählte das Projektteam schon in der Planungs- und Konzeptionsphase einen pragmatischen Ansatz für das als explorativ und ergebnisoffen beschriebene Experiment:

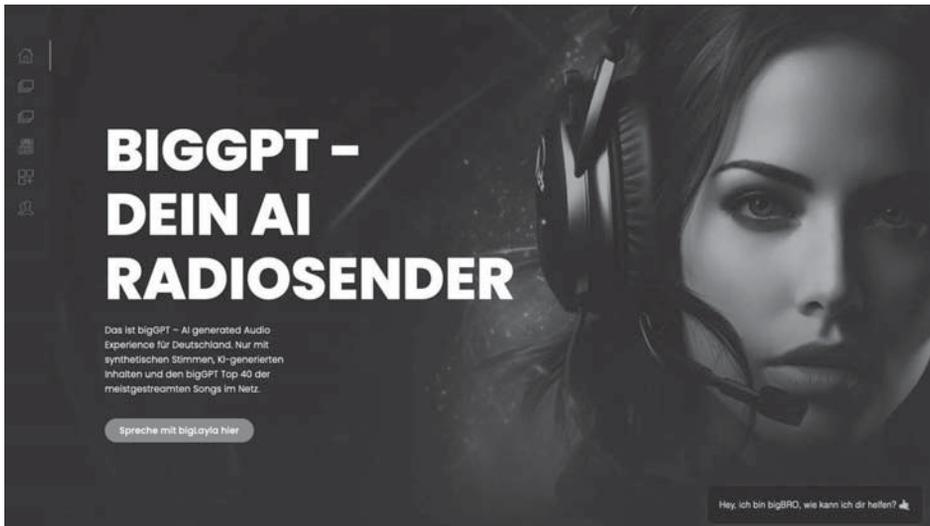
- ▶ Bei der Definition der Zielgruppe verzichtete das Team auf eine ausführliche Marktforschung. Stattdessen orientierte man sich an bekannten Nutzungsmustern und näherte sich sachbezogen: junge, technikaffine Männer zwischen 15 und 35 Jahren. (Also die Early Adopter, die vor knapp 20 Jahren auch YouTube als erste für sich entdeckten.)
- ▶ Im Mittelpunkt des Programms steht die auf der Website in Abbildung 5.7 vorgestellte Moderatorin bigLayla. Sie führt mit ihrer KI-generierten Stimme durch das zweistündige Programm im Dauerloop. Ihr Porträt ist auf der Website zu bewundern und wurde vom Team mit der Bild-KI mit *Midjourney* selbst erstellt: So stellt sich KI also eine KI-Radiomoderatorin vor. Ihr Name stammt nicht etwa aus einem bekannten Schlagertext und ist auch nicht das Ergebnis von Umfragen, sondern wurde kurzerhand von der ausgewählten Stimmvariante einer Sprach-KI übernommen. Als Sidekick von bigLayla kümmert sich der noch gesichtslose bigBen um die Präsen-

<sup>72</sup> <https://listen.streamon.fm/radiogpt>

<sup>73</sup> <https://biggpt.de>

<sup>74</sup> Apple App Store: <https://apps.apple.com/de/app/bigfm-radio/id447766328>, Google Play Store: [https://play.google.com/store/apps/details?id=de.radiocom.bigfm&pcampaignid=web\\_share](https://play.google.com/store/apps/details?id=de.radiocom.bigfm&pcampaignid=web_share)

tion der Technik-News, während der im Hipster-Look in Abbildung 5.8 visualisierte Chatbot bigBro mit den Usern chattet.



**Abbildung 5.7** KI-Moderatorin bigLayla führt durchs KI-generierte Programm von bigGPT. (Quelle: Screenshot Website)

- Ein Team, bestehend aus sieben Personen, entwickelt das Projekt von Beginn an und folgt einem disruptiven Ansatz, den Alexander Heine, Programmchef von Radio Regenbogen und Projektleiter Content von bigGPT, uns erklärt:

»Wir haben uns zu Beginn des Projektes vorgenommen, alles, was aus dem klassischen Radio kommt, zu vergessen und das Projekt bigGPT neu zu denken, auch strukturell.«

Obwohl ursprünglich aus dem klassischen Rundfunk kommend, verfolgt das Team mit der Einführung von KI-Technologien in der Content-Erstellung also den Ansatz, herkömmliche Strukturen und Arbeitsweisen des »Radiomachens« zu hinterfragen und nicht nur neu zu gestalten. Mit dem Ziel, die Potenziale der KI-Technologie voll auszuschöpfen, sollen alle traditionellen Workflows grundlegend nicht nur verändert, sondern von vornherein ersetzt werden. Schauen wir uns die Ansatzpunkte im Einzelnen an.

### Die Orchestrierung von Software und KI-Tools mit Futuri AudioAI

Im Zentrum der Content-Erstellung steht das Tool Futuri AudioAI<sup>75</sup>, ehemals RadioGPT genannt.

<sup>75</sup> <https://futuraimedia.com/solutions/audioai/>

- ▶ *Futuri AudioAI*: Das *End-to-End Content System* wurde vom US-Unternehmen Futuri-Media<sup>76</sup> entwickelt, um es innovativen Fernsehsendern, digitalen Verlagen und Streamingstationen als White Label zur Verfügung zu stellen. Das Programm, das eine Vielzahl von Technologien unter einem Dach vereint, ermöglicht es den Kunden, verschiedene Softwareprogramme und KI-Tools miteinander zu kombinieren, um beispielsweise ihre Radioprogramme mit aktuellen KI-generierten Inhalten und synthetischen Stimmen anzureichern.

Das Tool integriert neben GPT-4 mehrere Sprachmodelle, um Inhalte zu entwickeln, die auf aktuellen Nachrichten und Trends basieren. Es ermöglicht Text-zu-Sprache als auch Sprache-zu-Text-Konvertierungen, die die Grundlage für die KI-Moderation bilden.

Eine Schlüsselkomponente ist dabei die integrierte Sprachsynthese von leistungsstarken Sprach-KIs, darunter *ElevenLabs*, *PlayHT* und *Resemble AI*, um neben Futuris eigener Sprach-KI möglichst viele Stimmen verfügbar zu machen. Dadurch können die Nutzer auf eine große Auswahl an hochwertigen Moderationsstimmen zurückgreifen, um ein dynamisches und abwechslungsreiches Hörerlebnis zu schaffen.

- ▶ *Nahtlose Integration und permanenter Datenfluss*: Das Content-System ist damit das technische Herzstück der Produktion von bigGPT. Es orchestriert verschiedene KI-Tools und überführt die Ergebnisse in einen nahtlosen, automatisierten Prozess: Es arbeitet Hand in Hand mit der klassischen *Sendeablaufsteuerung* DABis von Audio-tainment Südwest, die auch für die klassischen Programme genutzt wird. Damit werden die von der KI produzierten MP3-Dateien nahtlos in das Sendesystem integriert und an die dafür vorgesehene Stelle im Programm platziert. Eine eigens entwickelte Schnittstelle zwischen der sendereigenen Sendeablaufsteuerung und dem Futuri-Tool ermöglicht einen permanenten Datenfluss und Informationsaustausch, was die Grundlage für die Erstellung der KI-generierten Moderationen bildet.

### Die Neuerung der redaktioneller Workflows

Der traditionell zeit- und personalaufwendige Redaktionsprozess wird bei bigGPT durch einen zunehmend effizienteren Workflow ersetzt:

- ▶ *Schnellere Redaktionszeiten*: Traditionelle Redaktionskonferenzen, in denen Themen in großer Runde diskutiert werden, entfallen bei bigGPT. Stattdessen identifiziert ein KI-Tool namens *TopicPulse*<sup>77</sup>, ebenfalls von Futuri Media, aktuelle Themen und Trends anhand aktueller Engagements im Netz. Daraus generiert es mehrere 30- bis

<sup>76</sup> <https://futurimedia.com>

<sup>77</sup> <https://futurimedia.com/solutions/topicpulse/>

60-sekündige Beiträge als Vorschlag. Aus diesen wählt ein Mitglied der Redaktion fokussiert relevante Themen aus und prüft diese ausgewählten Vorschläge auf ihren Wahrheitsgehalt, um Halluzinationen durch die KI auszuschließen. Dieses Vorgehen reduziert den Zeitaufwand erheblich.

- ▶ *Text-to-Speech-on-Air*: Die von der KI erstellten Skripte werden anschließend von den *Sprachsynthesemodellen* wie *ElevenLabs* zu Moderationen verarbeitet, die dann automatisch in die Sendeschleife integriert werden. Dazu werden klassische MP3-Dateien erstellt, die über ein Cloud-System in das Sendesystem übertragen werden.
- ▶ *Musik aus Social Media Charts*: Die Musikauswahl von bigGPT wird durch KI-Algorithmen unterstützt, die von der Musikredaktion gepflegt werden. Sie basiert auf der Analyse digitaler Nutzungsmuster, um Musikstücke, die auf YouTube, Shazam, Spotify, TikTok und Co. gerade angesagt sind, zu identifizieren und in eine nie da gewesene, durchaus auch gewöhnungsbedürftige Rotation zu bringen. »Damit verstoßen wir bewusst gegen alle Regeln des klassischen Musikformats. Auch eine emotionale Achterbahnfahrt von Helene Fischer zu AC/DC kann dabei passieren. Aber nur so entsteht ein neues Hörerlebnis«, erklärt uns Alexander Heine das disruptive Prinzip.

### Die Herausforderungen der KI-gestützten Content-Erstellung

Der Einsatz generativer KI-Modelle wie ChatGPT zur Content-Erstellung in Medienprojekten wie bigGPT birgt allerdings klassische wie auch spezifische Herausforderungen, wie das Projektteam berichtet:

- ▶ *Halluzinationen*: Generative KI-Modelle neigen bekanntlich dazu, Informationen zu »halluzinieren« oder zu erfinden, insbesondere wenn ihnen nicht genügend relevante Basisdaten zur Verfügung stehen. Das kann dazu führen, dass Inhalte generiert werden, die sehr plausibel erscheinen, aber nicht der Realität entsprechen. »Generative KI erfindet ein Taylor-Swift-Konzert so überzeugend, dass man eben wirklich zwei oder drei Mal hingucken muss. Weil die Wahrscheinlichkeit, dass das stattfinden könnte, vielleicht bei 80 % liegt, aber diese fehlenden 20 % sind halt genau der entscheidende Punkt«, begründet Heine die Herausforderung.
- ▶ *Anpassung des Kreativitätslevels*: Über den »Temperaturregler« kann die Balance zwischen lockerer Kreativität und dem Festhalten an Trainingsdaten gesteuert werden. Die Herausforderung besteht auch beim Radio darin, die richtige Ausgewogenheit zu finden, um realitätsnahe, aber gleichzeitig auch spannende Inhalte zu generieren.
- ▶ *Sprachliche Genauigkeit und Übersetzung*: Die Arbeit mit KI-Modellen, die primär auf Englisch basieren, führt bei Projekten in anderen Sprachen, wie bei bigGPT eben Deutsch, zu Übersetzungs- und Ausspracheproblemen. Insbesondere die korrekte

Wiedergabe von Namen, Zahlen und ins Deutsche adaptierten englischen Begriffen stellt eine Herausforderung für Redaktion und Technik dar. Die Notwendigkeit, der KI bestimmte Aussprachen oder Begriffe explizit beizubringen, zeigt, dass trotz fortgeschrittener Technologie eine individuelle Anpassung und Betreuung solcher Tools notwendig ist.

- ▶ *Rekonstruktion von Inhalten:* Generative KI hat grundsätzlich den Hang, nichts noch einmal genauso wiederholen zu wollen, wie schon einmal generiert. Das macht es schwer, einen einmal entwickelten Text zu rekonstruieren.

### Der konstruktive Umgang mit Herausforderungen

Den geschilderten Herausforderungen gilt es zu begegnen. Praktizierte Ansatzpunkte bei bigGPT sind:

- ▶ *Klassisches Vier-Augen-Prinzip:* Die bigGPT-Redaktion stellt die Einhaltung journalistischer Standards und die Relevanz der Informationen sicher, indem sie die KI-generierten Inhalte kontinuierlich von Menschen überprüfen und anpassen lässt. Alexander Heine erklärt uns dazu eine wichtige Regel: »Wir achten darauf, dass die Inhalte grundsätzlich journalistischen Standards entsprechen. Das fängt bei der Frage an: Ist das, worüber gesprochen wird, richtig? Gibt es das? Stimmt das? Deshalb ist der Faktor Mensch nach wie vor unverzichtbar. Wir lassen unsere KI nicht einfach allein laufen, sondern die redaktionelle Kontrolle bleibt.« Der Mensch bleibt somit ein entscheidender Faktor im Qualitätsmanagement und bei der Feinabstimmung der KI-generierten Inhalte, um die Integrität und Authentizität des Endprodukts zu gewährleisten.
- ▶ *Spezifische Dateneingabe:* Gibt man eigenes validiertes, recherchiertes, qualitätsgeprüftes Datenmaterial in ChatGPT ein und verbindet es mit der folgenden Aufforderung, so erhöht sich die Wahrscheinlichkeit, valide und belastbare Ergebnisse zu erhalten:

*Beispiel-Prompt:* »Arbeite genau mit diesem Material!«

- ▶ *Agilität und Pragmatismus im Projektmanagement:* Die ersten Erfahrungen mit bigGPT zeigen, dass ein flexibler, adaptiver Workflow notwendig ist, um auf unerwartete Ergebnisse oder Fehler mit dem neuen Partner KI reagieren zu können. Das erfordert eine enge Zusammenarbeit zwischen Redaktion und Entwicklern für eine kontinuierliche Optimierung der Arbeitsabläufe. »Das Hin und Her bei eigentlich verbindlichen Meilensteinen ist zwar nervtötend für alle Entwickler, aber ständiges Hinterfragen ist bei einem solchen Projekt unerlässlich«, schildert Alexander Heine. Die

Idee, das Projekt als Lernlabor zu betrachten, bei dem kontinuierliches Lernen und Anpassen im Vordergrund steht, ermutigt schließlich zum Ausprobieren, Scheitern und Weiterentwickeln.

### Mensch vs. KI in der Projektarbeit? Persönliche Einblicke und Philosophie

In der heutigen Arbeitswelt ist der Umgang mit KI eine der Schlüsselfragen für Teams und Einzelpersonen. Hier sind einige wichtige Punkte zur Philosophie und geliebten Praxis bei bigGPT:

- ▶ *Freiwillige Teilnahme und Engagement:* Die Grundlage für den Erfolg von KI-Projekten wie bigGPT ist das freiwillige Engagement der Teammitglieder. Beteiligung aus echtem Interesse und Neugier, nicht aus Zwang, soll bei bigGPT eine produktive Atmosphäre schaffen und die für explorative und innovative Projekte notwendige Belastbarkeit fördern.
- ▶ *Die Bedeutung von Dialog und Offenheit:* Ein offener Dialog innerhalb des Teams ist unerlässlich, um sich mit den grundlegenden Fragen und Anliegen auseinanderzusetzen, die KI aufwirft. Dazu gehören sowohl praktische Überlegungen zur Rolle von KI im Produktionsprozess als auch persönliche Fragen zur Bedeutung und Zukunft der eigenen Arbeit.
- ▶ *Wechsel von Angst zu Verständnis:* Die Auseinandersetzung mit KI birgt das Potenzial, Ängste und Vorbehalte in Wissen und Verständnis umzuwandeln. Durch die direkte Auseinandersetzung mit der Technologie im Rahmen eines Projekts wie bigGPT wird deutlich, dass hinter der scheinbaren »Intelligenz« strukturierte und formale Prozesse stehen, die zwar zu beeindruckenden Ergebnissen führen, aber keineswegs ein eigenes Bewusstsein oder eine eigene Willkür implizieren.
- ▶ *KI als Werkzeug:* KI wird bei bigGPT als ein Werkzeug verstanden, das beherrscht und sinnvoll eingesetzt werden kann. Diese Perspektive soll den Beteiligten helfen, Ängste abzubauen und die Technologie als unterstützendes Element im Arbeitsalltag zu sehen, das zur Effizienzsteigerung und Entlastung von Routineaufgaben beitragen kann, ohne die Rolle des Menschen zu ersetzen.
- ▶ *Fokus auf menschliche Stärken:* »Die zentrale Frage ist nicht, inwieweit KI den Menschen ersetzen kann, sondern wie sie den Arbeitsalltag unterstützen und verbessern kann. Es geht um Empowerment«, erklärt uns Heine die grundsätzliche Haltung. Das öffne den Blick dafür, wie KI als Assistent fungieren kann, der es dem Menschen ermöglicht, sich auf jene Aufgaben zu konzentrieren, die kreatives Denken, Empathie und zwischenmenschliche Interaktion erfordern – Bereiche, in denen der Mensch unersetzlich sei.

## Die Hörer-Reaktionen – Erwartungen und Realität

Aber wie reagiert die Hörerschaft auf das neue Programm der KI? Wie kommen bigLayla und bigBen an?

- ▶ *Feedback der bigGPT-Hörerschaft:* »Die Reaktionen der Hörer reichen von absoluter Faszination bis hin zu konstruktiver Kritik, dass es monoton sei, dass es im Zweifel technisch zu lange braucht, bis eine KI antwortet, dass es manchmal auch anstrengend sei, dem Programm zuzuhören. Und es stimmt: Man braucht wohl schon die Neugier an dieser Technologie, um einem bigGPT-Projekt dauerhaft zu folgen«, berichtet der Projektleiter. Viele Nutzer erwähnten vor allem den Unterschied im emotionalen Gehalt der Beiträge: Trotz technologischer Fortschritte bleibt die emotionale Komponente, die menschliche Moderatorinnen in ihre Sendungen einbringen, eine große Herausforderung für KI.

- ▶ *Emotionale Intelligenz vs. technische Empathie:* Aber kann KI überhaupt emotional intelligent sein? Diese Frage beantwortet uns Alexander Heine mit einer plakativen Erkenntnis:

*»Nein, denn in der KI brennt kein Feuer. Aber es gibt offenbar so etwas wie ›technische Empathie‹, die sie entwickeln kann.«*

Die zeigt sich z. B., wenn BigLayla mit einer fünfjährigen ZuhörerIn vermutlich geduldiger als jeder Elternteil über das Malen eines ganz bestimmten Bildmotivs spricht, ohne die ständigen Fragen des Kindes ungeduldig abzubrechen. Auch, wenn sie ihrem menschlichen Gegenüber ruhig zuhört, ohne dessen Sätze zu unterbrechen und dann selbst zu Ende zu führen, wirkt sie empathischer als so mancher Zeitgenosse aus Fleisch und Blut. Vor allem, wenn sie am Ende eines Gesprächs eine unbequeme Empfehlung ausspricht, etwa eine professionelle Beratung zu besuchen, für die uns Menschen untereinander oft der Mut fehlt, zeigt BigLayla ihre technische Empathie auf bemerkenswerte Art und nützliche Weise. Dabei kann die KI allerdings keinen emotionalen Bezug zum Gesagten herstellen, wie es eine menschliche Radiomoderatorin kann. Dieser Unterschied ist hörbar. Diese Lücke könnte sich aber in Zukunft schließen, wenn die Fähigkeit der KI zur emotionalen Interaktion weiterentwickelt wird.

- ▶ *Zukünftige Anwendungen von KI im Radio:* Bei den Prognosen für die Integration von KI im Radio orientiert sich bigGPT stark an der Reaktion, der Akzeptanz und den konkreten Bedürfnissen der Hörerschaft. Bereiche wie Nachrichtensendungen, die primär informative Inhalte ohne großen emotionalen Wert transportieren, bieten sich für KI-Anwendungen an. Daher ist das Sprachprofil von bigBen auch eher sachlich angelehnt. Auch andere Medienunternehmen experimentieren bereits mit KI-generierten Verkehrsmeldungen in Randzeiten.

### Die Vision eines synthetischen Audiokommunikationsraums

Es geht den Machern von bigGPT nicht darum, Radio mit KI zu imitieren, sondern sie möchten einen völlig neuen Erlebnisraum schaffen, in dem Menschen und KI auf eine neue Art und Weise interagieren. Anstatt KI lediglich als Werkzeug zu nutzen, um traditionelle Radioinhalte zu replizieren, wird ein Umfeld angestrebt, in dem KI und menschliche Kreativität zusammenfließen und so eine neue Form der Audiokommunikation entsteht.

- ▶ *User Experience im Fokus:* Ein Schlüsselaspekt dieses Konzepts ist die Fokussierung auf die User Experience. Indem man den traditionellen Radiobegriff hinter sich lässt, öffnet man sich für ein Erlebnis, das über das bloße Zuhören hinausgeht. Der synthetische Audiokommunikationsraum zielt darauf ab, eine tiefere emotionale Verbindung zum Hörer aufzubauen, ähnlich dem »Bauchmedium« Radio, das nicht nur gehört, sondern auch gefühlt wird.
- ▶ *Neue Möglichkeiten der Zuhörerbindung:* Durch den Einsatz von KI-Technologie eröffnen sich für bigGPT neue Wege, eine direkte Beziehung zu den Nutzern aufzubauen. Anstatt einer anonymen Masse gegenüberzustehen, ermöglicht der synthetische Audiokommunikationsraum eine 1-zu-1-Beziehung mit jedem Hörer. Dies kann ansatzweise durch intelligente, lernende Chatbots wie bigBro realisiert werden. Diese gehen weit über die eher primitiven Möglichkeiten herkömmlicher, eher mechanisch reagierender Chatbots hinaus und bieten mit KI eine echte, interaktive Kommunikation. Das führt auch gleich zum nächsten Punkt:
- ▶ *One to One – Balance zwischen Individualität und Gemeinschaftserlebnis:* Das Konzept eines individualisierten »One-to-One-Radios« ist eine durchaus faszinierende Vision für die Zukunft der Medien. Sie beruht auf der Idee, dass jeder Nutzer ein maßgeschneidertes Audioerlebnis erhält, das seinen persönlichen Präferenzen und Bedürfnissen entspricht. KI könnte diese Personalisierung ermöglichen, indem sie aus den Interaktionen und dem Feedback der Nutzer lernt und Inhalte entsprechend anpasst. Wir kennen das von den Play- und Shoppinglists auf Spotify, Netflix, Amazon und Co. Inwieweit eine vollständige Individualisierung wünschenswert oder möglich ist, bleibt noch offen. Trotz des Potenzials einer weitgehenden Personalisierung betont Alexander Heine die Bedeutung gemeinsamer Erfahrungen und die Rolle der Medien als Bindeglied in der Gesellschaft. Die menschliche Natur und die Sehnsucht nach Zugehörigkeit zu einer Gemeinschaft könnten der Idee eines vollständig individualisierten Radios entgegenstehen. Während individuell zugeschnittene Verkehrsinformationen oder Nachrichten sinnvoll und hilfreich sind, besteht die Gefahr, dass einzigartige und gemeinschaftlich teilbare Erlebnisse verloren gehen. Geschichten, die das Potenzial haben, Menschen zusammenzubringen und Gespräche anzuregen,

könnten in einem streng individualisierten Format untergehen. Die Identifikation und das Gefühl der Zugehörigkeit zu einer Gemeinschaft bleiben damit wichtige Aspekte, die auch in einem zunehmend digitalisierten und individualisierten Medioumfeld ihre Gültigkeit behalten. Die Fähigkeit der Medien, Menschen zu verbinden und gemeinsame Erlebnisse zu schaffen, sollte also auch im KI-Zeitalter nicht unterschätzt werden.



**Abbildung 5.8** Hipster bigBro chattet parallel zum Audio-Programm mit jedem Nutzer, one-to-one. (Quelle: Screenshot Website)

- *Praktikabilität und Nutzerwunsch:* Die Umsetzung eines One-to-One-Radios stößt auch auf praktische Herausforderungen. Frühere Versuche, die Hörerschaft zu einer aktiven Personalisierung ihres Erlebnisses zu bewegen, haben gezeigt, dass viele die Bequemlichkeit eines *Lean-Back*-Erlebnisses bevorzugen, bei dem Inhalte vorgegeben und nicht selbst gestaltet werden müssen. Ein intelligenter, lernender Algorithmus, der Nutzerpräferenzen erkennt und anpasst, könnte eine Lösung bieten, ohne die Hörerinnen und Hörer zu überfordern.

## 5.4 Disruption und Innovationsboom für Audio-Content durch KI

Wie kann man all diese Erkenntnisse und technischen Entwicklungen von KI im Audio-bereich zusammenfassen? Vielleicht hilft dazu ein kurzer Blick in den Rückspiegel: In kaum mehr als einem Jahrzehnt sind wir im Musikbereich vom physischen Tonträger zum Download und dann zum Streaming übergegangen. Der erneute Wandel, den KI nun im Audio-Segment anstößt, fühlt sich gerade aber nochmals anders an. »Das ist etwas, das noch stärker stört als der Untergang des physischen Mediums«, bringt der

britische Filmkomponist Julian Nott, bekannt für seine Arbeit für die Wallace-and-Gro-mit-Filme, es auf den Punkt.<sup>78</sup> »Es geht nicht nur um die Bereitstellung von Musik für unsere Kundschaft. KI wirft viel grundlegendere Fragen darüber auf, was wir unter Musikkreativität an sich verstehen, und darüber, welche Musik einen Wert hat und welche nicht – sowohl kreativ als auch finanziell gerechnet«, sagt Julian Nott und wünscht sich, dass die Musikschaaffenden schnell einen Konsens darüber finden, wie sie an KI herangehen.

Ein Beispiel, wie man sich offen dem disruptiven Einzug von KI in traditionelle Medienumfelder nähern kann, ist das Lernlabor bigGPT. Die Erfahrungen und Erkenntnisse, die das Projekt vermittelt, verdeutlichen das disruptive Potenzial, das KI für die gesamte Medienlandschaft birgt. Die geschilderte Herangehensweise bietet zugleich wertvolle Einblicke für andere Projekte und Organisationen, die den Einsatz von KI in kreativen und dynamischen Umgebungen erforschen möchten: Die Offenheit für grundsätzliche Veränderung, das kontinuierliche Lernen aus direkten Erfahrungen und die Bereitschaft, Annahmen zu hinterfragen, sind Schlüsselemente, um in einem sich schnell entwickelnden technologischen Umfeld erfolgreich zu sein und Raum für Innovation und kreative Lösungen zu schaffen.

Gleichzeitig werfen die Erkenntnisse wichtige Fragen zur Rolle der KI in unserer Gesellschaft auf. Die Herausforderung besteht darin, diese Technologie so einzusetzen, dass sie den menschlichen Erfahrungen und dem Bedürfnis nach sozialer Interaktion und Gemeinschaft gerecht wird. Der Schlüssel zum Erfolg liegt in einem nutzerzentrierten Ansatz, der die Stärken der KI nutzt, um die Medienwelt zu bereichern, ohne dabei die wesentliche menschliche Komponente zu vernachlässigen. bigGPT ist ein Beispiel dafür, wie Kreativität, Technologie und menschliche Intuition zusammenkommen können, um die Grenzen des Möglichen neu zu definieren und gleichzeitig den Kern dessen, was uns menschlich macht, zu bewahren.

---

<sup>78</sup> Stuart Dredge, AI music: disruption, opportunities, and the importance of human musicians. In: *musically*, <https://musically.com/2023/09/26/ai-music-disruption-opportunities-and-the-importance-of-human-musicians/>, 26.09.2023 [15.03.2024]