



KI in der Audio- branche

Eine Revolution in Klangwellen

Die Audiobranche befindet sich im Wandel, und Künstliche Intelligenz spielt dabei eine entscheidende Rolle. Vor kurzem galt KI in der Branche noch als Nischenphänomen, doch aktuelle Trends zeigen, wie wichtig ihre Rolle in naher Zukunft sein wird. Von KI-gesteuerten Radiosendungen bis hin zu fortschrittlicher Audibearbeitung - die Möglichkeiten sind grenzenlos. Andrea Anders von aprile consulting GmbH, ein Ad Tech-Beratungsunternehmen mit Spezialisierung auf die Audio- und Medienbranche, gibt im folgenden Interview Einblicke in die aktuellen Entwicklungen.

Wie hat die Künstliche Intelligenz (KI) die Audio- und Podcast-Branche in den letzten Jahren verändert?

Noch vor zwei Jahren war KI nur bei 13 Prozent der deutschen Unternehmen im Einsatz (Quelle: Trendstudie Digitalisierung 2021, www.tcs.com). Der Wert ist 2023 mit 15 Prozent noch nicht viel höher, aber drei von vier Unternehmen halten den künftigen Einsatz von KI für wettbewerbsentscheidend (Quelle: rnd RedaktionsNetzwerk Deutschland, 22.06.23). Seit etwa einem Jahr beobachten wir ein steigendes Interesse auch in der Audiobranche. Zahlreiche Unternehmen experimentieren mit KI-Anwendungen, und erste Produkte sind inzwischen auf dem Markt. Lassen Sie mich zwei Beispiele nennen:

Baden.fm war bereits 2022 Vorreiter mit dem Projekt „Eva“. Eva ist eine künstliche weibliche Stimme, die auf diesem Radiosender aktuelle Wetter- und Verkehrsnachrichten in Tagesrand- und Nachzeiten spricht, wenn echte Moderatorinnen oder Moderatoren teuer sind. Wetter- und Verkehrsdaten von verschiedenen

Lieferanten werden automatisch in lesbare Texte umgewandelt, um radiotypische Audioelemente ergänzt und dann in das Playoutsystem des Senders eingebunden. (Quelle: baden.fm)

Die ersten vollständig mit Hilfe von KI erzeugten Radioprogramme sind in Deutschland seit diesem Sommer on air. Moderation, Inhalte, Musikauswahl ... alle Programmbestandteile werden mittels KI erzeugt und zu einem 24-Stunden-Radioprogramm zusammengestellt. Antenne Deutschlands Sender nennt sich „Absolut Radio AI“ und kann über DAB+ im Raum Braunschweig empfangen werden.

„BigGPT“ in Anlehnung an ChatGPT und die Privatradiomarkte BigFM ist ein KI-Sender von der Audiotainment Südwest und ging mit KI-Moderatorin „Big Layla“ am 8. August 2023 auf Sendung.

Glauben Sie, dass KI jemals die menschliche Stimme in Podcasts oder Hörbüchern ersetzen könnte?

Die Qualität von künstlich erzeugten Stimmen ist allein in den letzten Monaten deutlich angestiegen. Wir können davon ausgehen, dass diese Entwicklung weitergehen wird. Ein wichtiger Faktor ist die zunehmende Leistungsfähigkeit von KI-Sprachmodellen. Diese Modelle können mit riesigen Datensätzen an Sprachdaten trainiert werden, was ihnen ermöglicht, sehr realistische Stimmen zu generieren. Darüber hinaus werden KI-Sprachmodelle immer komplexer und raffinierter, was bedeutet, dass sie in der Lage sind, subtile Nuancen der menschlichen Sprache zu erfassen.

Ein weiterer Faktor, der zur Entwicklung künstlicher Stimmen

beitragen wird, ist die hohe Verfügbarkeit von Daten. Dank der großen Verbreitung von Smartphones und anderen mobilen Geräten zeichnen immer mehr Menschen ihre Stimmen digital auf. Diese Daten werden für die Entwicklung von KI-Sprachmodellen verwendet.

Noch gibt es einige Herausforderungen. Vor allem können KI-Sprachmodelle bisher kaum Emotionen in der Stimme vermitteln. Deshalb klingen sie oft eher „tot“. Doch ist es wahrscheinlich, dass künstliche Stimmen in den nächsten Jahren immer realistischer werden. Ich vermute, dass sie bereits in den 2030er-Jahren nicht mehr von menschlichen Stimmen zu unterscheiden sein werden.

Wie sehen Sie die Zukunft von Audioinhalten mit der fortschreitenden Entwicklung von KI-Technologien?

Rund 27 Millionen Deutsche hören Podcasts, Hörbücher und Hörspiele (Quelle: Audible Hörkompass 2021). Die Ansprüche an die Qualität der Stimmen sind gerade für Unterhaltungsmedien hoch, und Podcasts, die zum Beispiel aufgrund des intellektuellen oder humoristischen Schlagabtausches zweier Persönlichkeiten populär sind, lassen sich derzeit noch nicht durch künstliche Charaktere ersetzen.

Sobald die jetzt noch bestehenden Qualitätsprobleme gelöst werden, können künstliche Stimmen gut, zum Beispiel für Hörbücher, aber auch für jede Art von Vertonung von Inhalten eingesetzt werden. 10,9 Millionen Deutsche sind aus gesundheitlichen Gründen auf Audio angewiesen (Quelle: Statista). Gerade diese Zielgruppe ist bereits heute froh über die vielen Text-to-Speech-Angebote,



die ihnen zum Beispiel von zahlreichen Zeitungsverlagen gemacht werden. Auch wenn die Qualität der Stimmen noch nicht hundertprozentig ist, die KI übernimmt für die Publisher einen Großteil der Arbeit und ermöglicht Barrierefreiheit für große Zielgruppen.

Gibt es bereits Beispiele für KI-generierte Musik oder Geschichten in Podcasts oder Hörbüchern?

Der kreative Prozess des Texteschreibens und des Komponierens von Musik kann heute schon durch KI unterstützt oder sogar vollständig übernommen werden. Es gibt zahlreiche „AI Sound Generators“, die musikalische Elemente wie Harmonien, Melodien, Rhythmen oder Jingles bis hin zu ganzen Songs erzeugen können. Beispiele für solche Musikgeneratoren sind „AIVA“, „Soundful“, „Mubert“, „Loudly Loudly“, „Music LM“ oder „Ecrett Music“. Man kann den Tools Texte vorgeben und diese werden in Songs mit einem bestimmten Musikstil umgewandelt. Musikgeneratoren können Beats erstellen und Soundeffekte nach Wunsch liefern. Textgeneratoren helfen beim Schreiben von Lyrics für Songs, selbst RAP-Texte sind kein Problem mehr. Ein neues Tool namens

„TextFX“ entstand in Zusammenarbeit zwischen Google und dem Rapper Lupe Fiasco (Quelle: [TextFX: Warum Google ein KI-Tool für Rapper gebaut hat](#)). Kurz gesagt „spielt“ die KI mit Worten und kreiert ganze Welten um einzelne Worte herum, die den Musiker oder die Musikerin inspirieren.

Michael Turbot von Sony Computer Science Laboratories sagte auf der US-Musik- und Technologiemesse South by Southwest in diesem Jahr: „Ganz ehrlich: In zwei Jahren oder früher wird es nicht mehr zu unterscheiden sein, ob KI oder ein Mensch den Song gemacht hat.“

Mit diesen neuen Tools ergeben sich für Audioproduktionen vielfältige Einsatzmöglichkeiten, zum Beispiel:

- » Intros und Outros
- » Hintergrundmusik
- » Soundeffekte für Hörbücher, Hörspiele und Podcasts
- » Ganze Songs als Teil der Story

In welcher Weise kann KI die Produktion von Audioinhalten effizienter und qualitativ hochwertiger machen?

- » Künstliche Intelligenz hilft bei verschiedenen Aufgaben, die im Zusammenhang mit Audioinhalten anfallen:
- » Ein Interview, ein Podcast oder andere Sprachaufnahmen sollen nachträglich bearbeitet, zum Beispiel einzelne Worte oder ganze Sätze entfernt werden: KI-Tools wie „TechSmith Audiate“ (textbasierte Audiotbearbeitung), „Podcastle“ (Podcast-Bearbeitung) oder „Descript“ (All-in-One Audio- und Videobearbeitung) ermöglichen dies schnell und einfach. Grundkenntnisse in Audiotbearbeitung reichen aus.
- » Text-to-Speech-Generatoren übernehmen die Vertonung von Texten. Dabei kann auf vollständig künstliche Stimmen zurückgegriffen werden. Oder KI kann mit echten Stimmen, zum Beispiel von bekannten Moderatoren und Moderatorinnen, in recht kurzer Zeit trainiert werden. Deren Einsatz ist dann später ohne den Menschen möglich.
- » KI hilft bei der Analyse von großen Datenmengen. Algorithmen suchen und erfassen Sprache, Objekte, Text, Personen, Sprecherinnen und Sprecher, Intentionen und Szenen. Diese Meta-

daten können mit Nutzerdaten in Mediatheken kombiniert werden. Dadurch wird es möglich, individuelle Empfehlungen für Nutzer KI-basiert zu erstellen und große Mengen an Inhalten in Mediatheken besser zu erschließen.

- » Gleichzeitig gibt es inzwischen Plug-Ins für gängige Digital Audio Workstations (DAW) und Audio-Editoren für Podcasts, die es den Produzenten ermöglichen, Musik- und Sprachelemente perfekt abzumischen. Aufnahmen werden analysiert und mittels Algorithmen abgemischt. Producer sparen mit diesen Tools wertvolle Zeit und können aus einem Mix das letzte bisschen Sound herauskitzeln. Beispielanbieter für solche Plug-Ins wären Sonible (deutscher Entwickler) und iZotope (USA).

Welche ethischen und rechtlichen Bedenken gibt es in Bezug auf die Verwendung von KI in der Audioindustrie?

Die Verwendung von Künstlicher Intelligenz (KI) in der Audioindustrie wirft verschiedene ethische und rechtliche Bedenken auf. Hier sind einige der wichtigsten:

- » **Urheberrecht und geistiges Eigentum:** Die Verwendung von KI-Technologien zur Erzeugung von Musik oder Audioinhalten kann Fragen zum Urheberrecht aufwerfen. Wer ist der Urheber, wenn ein Computerprogramm den Großteil der Arbeit geleistet hat? Es müssen klare rechtliche Rahmenbedingungen geschaffen werden, um die Rechte von Künstlerinnen und Künstlern zu schützen.
- » **Plagiat und Originalität:** Mit KI kann relativ einfach Musik im Stil bekannter Künstlerinnen und Künstler oder Musikgenres erzeugt werden. Dies kann zu Plagiatvorwürfen führen und die Frage aufwerfen, ob KI-generierte Musik als Originalmusik angesehen werden sollte.
- » **Manipulation und Täuschung:** KI kann verwendet werden, um Audioinhalte zu manipulieren, einschließlich der Erstellung gefälschter Sprachaufnahmen oder der Verfälschung von Audio in Videos. Dies kann für betrügerische oder manipulative Zwecke missbraucht werden und ethische Bedenken hinsichtlich der Verbreitung von Desinformation oder Falschinformation aufwerfen.
- » **Diskriminierung und Voreingenommenheit:** Wenn KI-Modelle zur Analyse von Audioinhalten oder zur automatischen Erzeugung von Musik trainiert werden, können sie Vorurteile aufgrund von Trainingsdaten aufweisen. Dies kann zu diskriminierenden oder voreingenommenen Ergebnissen führen, die ethische Bedenken hinsichtlich der Fairness und Gerechtigkeit aufwerfen.
- » **Datenschutz:** Die Verwendung von KI zur Analyse von Audioinhalten wie Spracherkennung oder Stimmungsanalyse kann Datenschutzfragen aufwerfen. Die Erfassung und Verarbeitung von Audioaufnahmen erfordern die klare Zustimmung und den Schutz der Privatsphäre der Benutzerinnen und Benutzer.
- » **Arbeitsplatzverlust:** Die Automatisierung von Aufgaben in der Audioindustrie durch KI könnte zu Arbeitsplatzverlusten führen, was soziale und ethische Fragen zur Arbeitsplatzsicherheit aufwirft.
- » **Qualitätskontrolle:** Die Verwendung von KI zur Erzeugung von Audioinhalten kann die Qualität und Originalität von Musik

Use Case: OMD und Penny setzen bei der Produktion von Hörfunkspots erstmals auf KI

OMD und PENNY haben eine KI-basierte Audio-Kampagne für Hörfunkspots eingeführt, um die Angebotskommunikation zu revolutionieren. Mit Hilfe der KI-Technologie S.A.R.A. (Speech & Audio Realtime Automation) werden wöchentlich aktualisierte, hyperlokale Spots mit einer synthetischen, aber realistisch klingenden Stimme erzeugt und programmatisch im jeweiligen Einzugsgebiet ausgestrahlt. Die KI erlaubt es, Spots schnell und kosteneffizient zu individualisieren, wobei Orts- und Zeitdaten in Kombination mit einem speziellen Marketing-Tool genutzt werden. Die Kampagne ist auf Plattformen wie Spotify und SoundCloud sowie bei verschiedenen Radiosendern zu hören und wird im Laufe von drei Monaten mehr als 2.500 Spots produzieren.

und Audio beeinträchtigen, was wiederum ethische Fragen zur künstlerischen Integrität und kulturellen Vielfalt aufwerfen kann.

- » **Bezahlung:** Wenn Stimmen von echten Menschen künstlich nachgeahmt und dann zum Beispiel in Podcasts, Radioprogrammen oder Hörbüchern eingesetzt werden, um die menschliche Arbeitszeit zu sparen, entstehen Fragen der Bezahlung. Ein Mensch hat seine Stimme zur Verfügung gestellt, wird nun aber zum Beispiel nicht mehr zum Einsprechen neuer Inhalte

benötigt. Wie wird seine Leistung honoriert?

Um diese ethischen und rechtlichen Bedenken anzugehen, ist es wichtig, klare Richtlinien und Vorschriften für die Verwendung von KI in der Audioindustrie zu entwickeln. Dies sollte in enger Zusammenarbeit zwischen der Industrie, Regierungsbehörden, Künstlerinnen und Künstlern und Technologiefachkräften erfolgen, um sicherzustellen, dass die Vorteile der KI genutzt werden, ohne die oben genannten Bedenken zu vernachlässigen.

Autorin



Andrea Anders: langjährige Marketing- und Verkaufsleiterin bei großen Medienverlagen, ist bei amily GmbH für die Aufgabenbereiche Produktentwicklung, Consulting, Projektmanagement und Sales zuständig. Und bei aprile consulting GmbH ist sie als Consultant für Medien- und Vermarktungsthemen tätig. Obwohl Andrea Anders auch einschlägige Erfahrung in anderen Mediengattungen mitbringt, schlägt ihr Herz besonders für Audio. So bekleidete sie in den vergangenen 30 Jahren leitende Marketing- und Vertriebsfunktionen, u. a. bei OK Radio, Radio Hamburg und zuletzt bei der VMG Verlags- und Medien GmbH & Co. KG in Magdeburg mit den Sendern SAW, Rockland sowie 20 Webradios.



Abstract

Die Rolle der Künstlichen Intelligenz in der Audiobranche wächst rasant. Vor zwei Jahren nutzen nur 13 % der deutschen Unternehmen KI; heute sind es 15 %, wobei drei Viertel der Unternehmen die Technologie für wettbewerbsentscheidend halten. Insbesondere im Radiobereich hat KI bereits Einzug gehalten: So gibt es mit „Eva“ von Baden.fm eine KI-gesteuerte Stimme, die Wetter- und Verkehrsdaten verliest, und auch vollständig KI-generierte Radioprogramme wie „Absolut Radio AI“ oder „BigGPT“ sind bereits on air.

Die Qualität künstlich erzeugter Stimmen verbessert sich stetig, angetrieben durch fortschrittliche KI-Modelle und riesige Datensätze. Obwohl sie bisher kaum Emotionen vermitteln können, könnte ihre Qualität in den 2030er-Jahren mit der menschlichen Stimme gleichziehen. Dies hat insbesondere Potenzial für Hörbücher und für die Barrierefreiheit.

Doch KI ist nicht nur im Bereich Sprache aktiv. Musikgeneratoren wie „AIVA“ oder „Soundful“ können komplette Songs kreieren, und in der Audioproduktion ermöglichen KI-Tools wie „TechSmith Audiate“ oder „Descript“ eine effiziente und qualitativ hochwertige Nachbearbeitung von Sprachaufnahmen.

Trotz dieser Fortschritte bleiben ethische und rechtliche Fragen, etwa im Bereich des Urheberrechts oder des Datenschutzes, die dringend geklärt werden müssen.