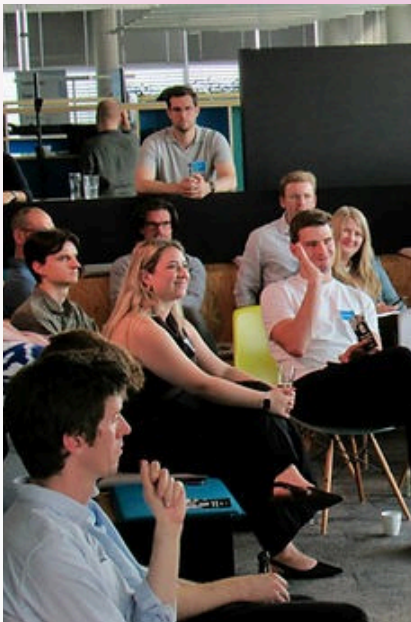


AI AUDIO WHITEPAPER

Erkenntnisse und interdisziplinäre Perspektiven vom AI AUDIO DAY
(#AIAD) 2026 zum KI-Audio-Einsatz in der Medienbranche

2026



Die Zukunft vorhersagen, indem wir sie bauen

Zukunft entsteht dort, wo Menschen den Mut haben, Ideen auszuprobieren, voneinander zu lernen und gemeinsam etwas zu bauen. Am 2. Juli 2026 kamen im o2 Tower in München genau solche Menschen zusammen: Redakteure, Podcaster, Plattformbetreiber, Technologiepartner und Kommunikatoren. Menschen, die KI-Audio nicht nur diskutieren, sondern schon heute in konkrete Produkte oder Formate übersetzen.

Sie alle verbindet eine Frage: Wie können wir mit KI besseren Journalismus und bessere Kommunikation schaffen? Eine ebenso anspruchsvolle wie ermutigende Mission.

Der AI Audio Day ist bewusst ein Forum, klein genug für ehrliche Gespräche und groß genug für unterschiedliche Perspektiven. Gerade im Lokaljournalismus entsteht daraus eine besondere Offenheit: Husumer können von Heidenheimern lernen, Augsburger von Aachenern – ohne sich Leser, Hörer oder Nutzer streitig zu machen.

Wissen wird nicht kleiner, wenn man es teilt. Innovation wird dadurch schneller.

Für den Blick in die Zukunft halten wir es mit dem Informatiker Alan Kay:

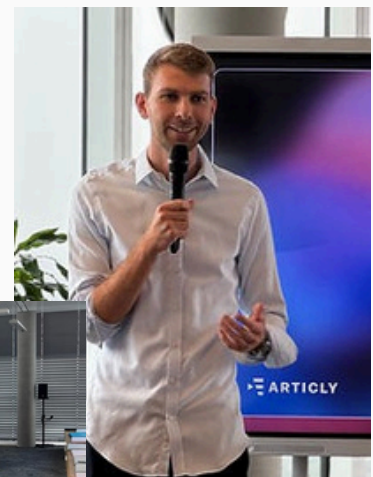
“The best way to predict the future is to invent it.”

Darin liegt der Anspruch von ARTICLY und des AI Audio Days: nicht abzuwarten, welche Rolle KI und Audio künftig spielen, sondern diese Entwicklung mit Optimismus, Verantwortung und guten Produkten aktiv mitzugestalten.

Dieses Whitepaper verdichtet die wichtigsten Ideen, Erfahrungen und Use Cases des Tages. Es soll Orientierung geben, Mut machen und Lust darauf wecken, selbst loszulegen.

Denn die Zukunft von Audio wird nicht nur vorhergesagt. Sie will gestaltet werden.

Wolf Weimer, AIAD Host



Der AIAD in sieben Thesen

1 Podcast wird Leitmedium. Der Werbemarkt hinkt hinterher.

Bei 18- bis 49-Jährigen hat Podcast das Radio überholt, für über ein Drittel ist er Hauptquelle der Nachrichtenaufnahme. Vermarktet werden aber nur 7 Prozent der aktiven Podcasts: Die Nutzung ist zwar da, die Erlöse entstehen langsamer.

2 Das Modul löst das Format ab.

150 Jahre war Audio linear. KI bricht diese Linearität: Die Zukunft gehört Audio-Bausteinen, um Inhalte mehrfach zu verwerten. Einmal produziert können diese neu kombiniert werden als Briefing, E-Paper, Playlist oder personalisierte Podcasts.

3 Vertrauen ist die Währung und sie hat klare Regeln.

Marken, denen Menschen vertrauen, bauen ihren Vorteil im KI-Zeitalter weiter aus. Wer gewinnt, hat eine direkte Beziehung zum Publikum; wer KI einsetzt, kennzeichnet sie von Anfang an - 87 Prozent der Hörer lehnen unklare Herkunft ab. Hörbare KI wird angenommen, das künstliche Gesicht (noch) nicht.

4 Der lokale Podcastmarkt ist noch weitgehend unbesetzt.

Echter, vor Ort produzierter Lokal-Content macht nur einen kleinen Teil des regionalen Audioangebots aus, der Rest ist recyceltes Radio. Die Nachfrage nach lokalen Informationen ist da. Nur das Angebot fehlt.

5 Gute Stimmen ersetzen keine guten Inhalte.

KI-Stimmen klingen inzwischen überzeugend real - sogar im Dialekt. Entscheidend ist der Text: Ein Skript, das fürs Hören geschrieben ist, senkt Aussprachefehler und macht einen hörbaren Unterschied.

6 Der Mensch bleibt am Steuer.

Die vollautomatisierte KI-Sendung der WELT wird erst nach menschlicher Freigabe veröffentlicht. Es geht nicht darum, wen KI ersetzt, sondern wie Redakteure die neu gewonnene Zeit besser einsetzen.

7 Machen schlägt Warten.

Der Prototyp ersetzt das Konzeptpapier: drei Tage Bauen bringen oft mehr als drei Monate Prozess. Wer früh die richtige Infrastruktur schafft, kann schnelle technologische Sprünge forcieren.

1 Das große Bild: KI verändert die Wertschöpfung

Inspiziert von Redebeiträgen durch Olivier Boder & Stefan Christoph (AWS) und Paulina Kiekenbeck (Project A)

Vom Konzeptpapier zum Prototypen

Wie schnell kann ein großes Medienhaus heute ein KI-Audioprodukt bauen? Die Antwort, die Olivier Boder und Stefan Christoph von AWS mitbrachten beruht auf einer wahren Geschichte. Der Kunde war ein europäisches Medienhaus mit zahlreichen Radiostationen. Dieses Haus brachte für drei Tage sechs Teams aus verschiedenen Ländern zusammen, mit der gemeinsamen Aufgabe. Die Teams sollten einen Prototypen eines funktionierenden, KI-gestützten Radiostreams bauen. Kein PowerPoint, kein Sechsseiter, sondern ein laufendes Produkt.

Die Teams verteilten sich entlang der gesamten Wertschöpfungskette einer Radiosendung: Aufbereitung von Rohnachrichten zur Sendung, Musikplanung, Verkehrsinformationen, Einbindung der Community über Kanäle wie WhatsApp, lokale und wetterabhängige Werbung sowie ein KI-Sprecher, der die Sendung moderiert. Am Ende der drei Tage lief der Stream. Er lief sogar so gut, dass die im Hintergrund spielende Musik des Hackathon-Raums irgendwann der KI-generierte Stream selbst war, ohne dass es auffiel.

Das eigentliche Learning liegt jedoch nicht in der Technologie. 1. Tempo: In drei Tagen entstand, woran sonst rund drei Monate gearbeitet worden wäre, Abstimmungsschleifen, Meetings und manuelles Coden inklusive.

2. Besetzung: **In den Teams saßen** nicht nur Techniker, sondern **auch Domain-Experten**. Im Gewinnerteam war ein echter Radiomoderator mit am Tisch. Die Kommunikation zwischen Menschen, die das Fachproblem verstehen, und denjenigen, die die Technik verstanden, war der Beschleuniger, während die KI im Hintergrund baute. 3. Kulturwandel: Das traditionsreiche „Working Backwards“-Verfahren, also erst FAQ und Pressemitteilung schreiben, dann bauen, wird zunehmend ergänzt durch den umgekehrten Weg: **erst einen Prototypen bauen**, der überzeugt, dann den Prozess nachziehen.

Wichtig bleibt die Einordnung, die beide Architekten mehrfach betonten: Ein Prototyp ist kein Produkt. Wer per „Vibe Coding“ in Stunden etwas Funktionierendes zusammenstellt – Stefan Christoph demonstrierte einen nachgebauten KI-Radiostream, erstellt in rund zwei Stunden – braucht anschließend weiterhin Entwickler, die daraus etwas Wartbares, Erweiterbares und Produktionsreifes machen. Dabei bleibt der Mensch verantwortlich. KI ist ein neues Werkzeug im Werkzeugkasten, das die Einstiegshürde senkt und keine Instanz, an die man Entscheidungen delegiert.

„KI ändert nichts daran, was ein ordentliches Produkt ist – sie ändert, wie schnell ich dahin komme.“

Olivier Boder & Stefan Christoph, AWS

Beyond the Hype: Was der Mittelstand gerade lernt

Paulina Kiekenbeck arbeitet bei A11, der Beratungseinheit, die der Berliner Frühphasen-Investor Project A 2025 als eigenständiges Unternehmen ausgegründet hat. Bei A11 arbeiten rund 150 Leute operativ mit Portfolio-Unternehmen und großen Häusern in der Transformation. Zwei Tage vor dem AI Audio Day hat sie den KI-Gipfel in Stuttgart mitgestaltet: über 1.000 Teilnehmer in der MHP Arena, bewusst außerhalb der Berliner Tech-Bubble, mitten im Mittelstand.

Ihre wichtigste Beobachtung ist eine Verschiebung der Fragestellung. Vor einem Jahr suchte der Mittelstand noch nach ersten Use Cases: „Was könnten wir mit KI machen?“. In diesem Jahr lautete die Frage: „Was hat es aus der Pilotphase in den produktiven Einsatz geschafft, und **was bringt messbar Mehrwert?**“. Das Motto des Gipfels brachte es auf die Formel: „einfach machen“. Die Zahlen stützen den Wandel. Der Anteil mittelständischer Unternehmen, die KI einsetzen, ist binnen eines Jahres um gut die Hälfte gestiegen.

Für Medien zog sie eine klare Trennlinie zwischen Gewinnern und Verlierern des KI-Umbruchs. Seit über einem Jahr übersteigt laut Studien der KI-generierte Content im Netz den von Menschen erstellten. Wer sein Publikum über Suchmaschinen-Traffic erreicht, verliert; wer eine **direkte Beziehung** zur eigenen Leserschaft hat, **gewinnt** – sichtbar etwa an den **Abonnementrekorden** von Titeln wie der ZEIT oder der New York Times. Die Währung dahinter: Vertrauen, hergestellt durch spürbaren Human Touch. Als Negativbeispiel diskutierte der Saal die Debatte um KI-generierte Anteile in Reden des Bundesdigitalministers. Wo Menschen den menschlichen Anteil nicht mehr spüren, fühlen sie sich getäuscht.

Zwei Fehler, so Kiekenbeck, könne sich derzeit kein Unternehmen leisten. KI zu ignorieren und keine Berufseinsteiger mehr einzustellen. Denn während global nur rund jeder Sechste regelmäßig KI nutzt, sind es unter Studierenden über 80 Prozent. Die junge Generation bringt genau die Kompetenz mit, die der Wandel verlangt.

Take Aways – Kapitel 1

- Der Prototyp ersetzt das Konzeptpapier: Drei Tage Bauen mit Domain-Experten ersetzen drei Monate Prozess, aber ein Prototyp ist noch kein Produkt.
- Die Marktphase hat gewechselt: nicht mehr „Welche Use Cases gibt es?“, sondern „Was skaliert aus der Pilotphase heraus?“.
- Im KI-Zeitalter gewinnt, wer eine direkte, vertrauensvolle Beziehung zum Publikum hat. Human Touch ist Strategie, nicht Stilfrage.



2 Aus der Redaktion: KI-Audio im journalistischen Alltag

Inspiziert von Redebeiträgen durch Thibault Krause (WELT) und Kirsten Stünkel & Jennifer Truhn (NOZ Digital)

Case 1 – “Going Local”: Die WELT baut ein lokales Informationstool



Live-Interview auf der Bühne des AI Audio Day 2026, Wolf Weimer (links) und Thibault Krause (rechts)

Die Idee hinter „Going Local“ ist bewusst bescheiden formuliert: kein Recherche-, sondern ein Informationstool. Für die Pilotstadt Stendal bündelte ein KI-System die öffentlich verfügbaren lokalen Quellen darunter offizielle Seiten von Stadt, Gemeinde, Polizei und Feuerwehr sowie lokale Social-Media-Kanäle. Daraus spielte es zunächst kurze, meist einminütige Avatar-Videos als „News-Snack“ zu Wetter, Veranstaltungstipps und Verkehrslage aus. Die lokale Tageszeitung wurde bewusst nicht als Quelle genutzt. Die WELT setzt stattdessen auf Kooperationsmodelle mit Verlagen, erste Gespräche in anderen Städten gibt es auch schon. Der Anspruch, so Krause, sei ausdrücklich kein Gegenmodell zum Lokaljournalismus, sondern der Versuch, dort ein Angebot zu schaffen, wo ein Vakuum entsteht.

Was dann geschah, überraschte selbst das Team:

schnelles organisches Wachstum und Nutzer, die dem Kanal von sich aus Informationen zuschickten über eine gesperrte Straße, die neue Pizza des lokalen Italieners und viele mehr. Der Endnutzer wurde vom Konsumenten zum Mitproduzenten, nach dem Prinzip einer Blitzer-Warn-App. Je mehr Menschen etwas melden, desto verlässlicher die Information.

Das zentrale Learning des Projekts betrifft jedoch die Akzeptanz, nicht die Technik. Sobald Nutzer erkennen oder erfahren, dass KI im Spiel ist insbesondere bei der visuellen Komponente des Avatars stellt sich reflexhaft Skepsis ein, unabhängig von der inhaltlichen Qualität. Diese war aber geringer als erwartet und sank mit der Zeit deutlich.

In der Redaktion hat das Projekt eine Grundsatzdebatte befeuert, die Krause offensiv führt. Statt nur KI-Inhalte zu kennzeichnen, könnte man den Spieß umdrehen und menschengemachte Inhalte mit einer Art Gütesiegel versehen, analog zum Bio-Siegel im Supermarkt.

Der zweite Erfahrungsschatz der WELT ist die wöchentliche KI-Sendung : fünf bis zehn Minuten, freitags, klar als KI gekennzeichnet. Die KI wählt Themen aus, schreibt die Skripte, schneidet und vertont selbst. Die Sendung kommt im Prinzip fertig produziert aus der Maschine. Veröffentlicht wird trotzdem nichts ohne menschliche Freigabe. Ein Chef vom Dienst prüft jede Ausgabe inklusive Fact-Checking.

Dieses Prinzip, betont Krause, wird bleiben. Auf die Frage nach Verdrängungsängsten antwortet Krause mit einer Umdeutung, die sich durch den ganzen Tag zog: Die falsche Frage sei „Wen kann ich mit KI ersetzen?“, die richtige: „Was schaffe ich mit der neu gewonnenen Zeit?“ Wer statt 40 nur noch 10 Stunden pro Woche schneidet, hat 30 Stunden für originelleren Content. Seine Prognose für die kommenden drei bis fünf Jahre ist, dass Schnitt und Sendungsaufbau weitgehend von der KI übernommen werden; Moderation und Vertonung bleiben menschlich. Gerade die Imperfektion und die regionalen Färbungen echter Stimmen haben Wert. Und den Reporter vor Ort, der im Krisengebiet Informationen einsammelt, wird eine KI absehbar nie ersetzen.

Case 2 – NOZ Digital: Was Nutzer von KI-Audio wirklich erwarten

Die NOZ Digital kommt beim Thema KI-Audio aus der Praxis. Der Morgen-Podcast „Guten Morgen SH“ – täglich um 5:30 Uhr, mit erheblichem Produktionsaufwand – wurde gemeinsam mit ARTICLY schrittweise zu einem KI-gestützten Portfolio: kürzere Produktionswege, lokalere Ausspielung bis auf Stadtebene, Tests in Nischen wie Blaulicht-Themen. Statt jedoch Feature um Feature aus dem Bauch heraus zu bauen, hat das Team die Nutzerschaft systematisch befragt mit dem Kano-Modell, das Produkthanforderungen in Basis-, Leistungs- und Begeisterungsmerkmale unterteilt. Die Ergebnisse zeichnen ein präzises Bild der Erwartungen. Hier die wichtigsten Befunde:

Audio ist Alltagsbegleiter, nicht Primärmedium. Die Nutzer verstehen Audio als Ergänzung, konsumiert parallel zu anderen Tätigkeiten: 45 Prozent beim Autofahren, 45 Prozent bei der Hausarbeit und 23 Prozent während der Morgenroutine. Für die Produktentwicklung heißt das, dass CarPlay, Hintergrundwiedergabe und ein vollwertiger Player keine Kür sind, sondern Pflicht.

Transparenz ist eine unverhandelbare Basisanforderung. 87 Prozent lehnen eine unklare redaktionelle Herkunft ab. Das heißt, dass ein Produkt, das nicht erkennbar in Website, App und Marke eingebettet ist, durchfällt. 80 Prozent stört eine fehlende KI-Kennzeichnung. Diese sollte am Besten direkt zu Beginn kommen und nicht als Hinweis am Ende.

Aussprache als Kompetenzbeweis. Gerade die Zielgruppe 45–64 bewertet die korrekte Aussprache regionaler Begriffe und Namen als Qualitätskriterium. Wer die Holtenauer Hochbrücke in Schleswig-Holstein falsch betont, verliert an Glaubwürdigkeit. Für die Qualitätssicherung wurde eigens eine „Ortsnamen-Testreise“ durch Schleswig-Holstein vertont und bestanden.



Zusammenfassungen ja, aber mit Kontext und Verweis. 48 Prozent begrüßen Audio-Summaries; als ideale Länge kristallisierten sich rund zwei Minuten pro Einheit heraus, beim morgendlichen Briefing drei bis fünf Minuten. Komplexe Themen dürfen länger sein und Hörer wünschen sich die direkte Verlinkung zum vollständigen Artikel.

Personalisierung: Vorauswahl statt Filterblase. Der vielleicht feinste Befund: In qualitativen Interviews wünschen sich Nutzer zunächst volle Selbstzusammenstellung. Auf Nachfrage zeigt sich jedoch die Sorge, in einer Blase zu landen und das gemeinsame Lagerfeuer der Zeitung zu verpassen. Eine kuratierte Vorauswahl mit regionaler Eingrenzung gilt als Leistungsmerkmal, intelligente Vorschläge als Begeisterungsmerkmal.

Aus diesen Befunden entsteht derzeit ein MVP: Ein morgendliches Briefing von drei bis fünf Minuten mit den Elementen Stadt, Welt und Events. Dieses wird prominent platziert und in die bestehenden Apps und Portale integriert statt als separate App.

Parallel arbeitet das Team daran, wie die Nähe zur Redaktion hörbar wird, etwa indem der Chefredakteur die Anmoderation spricht. Das menschlich produzierte Podcast-Portfolio bleibt bewusst getrennt. KI-Audio und Premium-Podcasts sind zwei Säulen, die sich ergänzen, nicht ersetzen.

Take Aways – Kapitel 2

- Sichtbare KI (Avatare) erzeugt Skepsis, hörbare und lesbare KI wird angenommen. Die Kennzeichnung gehört an den Anfang, die Einbettung in die vertraute Marke ist Pflicht.
- Nutzerforschung schlägt Bauchgefühl: Das Kano-Modell liefert eine belastbare Rangfolge von Pflicht (Transparenz, Player, Aussprache) über Kür bis Begeisterung.
- Die produktivste Frage der KI-Nutzung lautet: „Was machen wir mit der gewonnenen Zeit?“

3 Beyond TTS: Wie modulares KI-Audio Content verändert

Inspiziert von Redebeiträgen durch Lukas Paetzmann (ARTICLY) und Philipp Munich (ProSiebenSat.1)

Das Ende der Linearität

Lukas Paetzmans Vortrag begann mit einer Zeitreise, die zugleich eine Demutsübung war: 2016 klangen Text-to-Speech-Stimmen wie ein Anrufbeantworter; vor einem Jahr feierte der AI Audio Day Stimmen, die erstmals lachen konnten; heute sprechen KI-Stimmen so überzeugend Bayrisch, dass der Unterschied zu einem echten Menschen kaum noch zu hören ist. Die jährliche Pointe:

Worüber man heute staunt, wirkt in zwölf Monaten schon veraltet. Die Stimmqualität ist als Engpass praktisch verschwunden. Die eigentliche These setzt tiefer an: Audio war 150 Jahre lang linear. Ein Format, ein Anfang, ein Ende und wer hören wollte, musste nehmen, was kam. Eine Zeitung kann man aufschlagen und gezielt den Artikel unten rechts lesen; ein kurzer Newscast hat möglicherweise sieben Minuten irrelevante Themen, bis der Inhalt kommt, der einen wirklich interessiert. 9

KI bricht diese Linearität, weil sie es erlaubt, nicht mehr in Formaten denken zu müssen, sondern in Modulen.

Die kleinste Einheit dieses Systems ist die Audio-Summary: eine kondensierte, fürs Hören geschriebene Version eines Artikels. Paetzmanns Beispiel ein Servicetext über freie Ferienunterkünfte mit über sieben Minuten Vorlesedauer zeigt das Prinzip: Kaum jemand hört sieben Minuten Ortsnamen und Bettenzahlen; die einminütige Summary transportiert den Kern. Aus diesem einen Baustein lassen sich beliebig viele Produkte schneiden: eine bessere Vorlesefunktion, kuratierte Playlists, ein Audio-E-Paper, redaktionell kuratierte Tages-Highlights, lokale Briefings bis hin zum vollautomatisch personalisierten Podcast. Die Ökonomie dahinter: Das Modul wird einmal produziert und vielleicht 25-mal verwendet; das Zusammenschneiden verursacht nur marginale Kosten und skaliert auf hunderttausende Varianten pro Tag.

Wie Personalisierung konkret aussieht, zeigte ein Prototyp mit der Deutschen Bahn: Der Reisende wählt Themen (Politik, Wirtschaft, Wetter - bitte kurz), das System sucht alle passenden, bereits vertonten Module, ergänzt eine persönliche Begrüßung und nach rund **acht Sekunden** liegt der individuelle Podcast für die Fahrt im ICE vor. Wichtig ist die Reihenfolge. Niemand startet mit personalisierten Podcasts. Der Weg führt stufenweise von der Audio-Summary über Briefings und Newscasts zum Audio-Hub und erst dann zur Personalisierung. Fünf Artikel zu personalisieren lohnt nicht, das gesamte Tagesangebot schon.

Dass KI-Formate und menschlich produzierte Podcasts keine Konkurrenten sind, illustrierte das Beispiel Kicker: Das tägliche KI-Audioformat ist der „Habit Builder“, der Hörer in Routine und ins Audio-Universum der Marke bringt, wo sie dann auf die wöchentlichen Interview-Formate und den Evergreen-Content stoßen, die menschlich produziert sind. Das tägliche Format ist Zubringer, nicht Ersatz.

Die unterschätzte Kunst: Das Skript

Das ehrlichste Learning betraf nicht Stimmen, sondern Text. Ein TTS-Modell funktioniert wie ein Sprachmodell: Es arbeitet mit Wahrscheinlichkeiten. Es gibt Sätze, die es sicher aussprechen kann, und solche, bei denen es scheitert. Wer der KI ein starres Korsett anlegt („Lies exakt diesen Satz vor, der fürs Lesen geschrieben wurde“), erzwingt Aussprachefehler. Ein fürs Hören geschriebenes Skript mit kürzeren Sätzen, reduzierter Komplexität, natürlichen Formulierungen klingt nicht nur besser, es ist auch fehlerärmer.

Die Kunst liegt in der Balance: 100 Prozent Nähe zum Original macht jede Chefredaktion glücklich, klingt aber vorgelesen; volle hörsprachliche Freiheit klingt großartig, wirft aber die Frage der faktischen Treue auf. ARTICLY arbeitet an dem Punkt dazwischen. Drei Mechaniken sichern das ab: ein Tonalitätstraining, bei dem das System aus jeder redaktionellen Korrektur lernt; eine saubere Extraktion der Fakten des Artikels; und mehrstufige Verifikation, bei der Sprachmodelle die Aussagen der Zusammenfassung nacheinander gegen das Original prüfen. Ein Nebenfund mit Sprengkraft: Die Annahme, Summaries müssten kurz sein, war falsch. Entscheidend ist nicht, ob fünf Sätze stimmen, sondern was weggelassen wurde und wenn die KI 70 Prozent der Originallänge braucht, darf sie sie nehmen.

„Text-to-Speech ist keine Audio-strategie. Wer den Vorleseknopf drückt, ist längst zum Artikel navigiert. Der Wechsel zwischen Lesen und Hören sollte immer möglich sein.“

Lukas Paetzmann, ARTICLY

Zum Schluss zog Paetzmann Bilanz über seine Vorjahres-Hypothesen: Stimmen realer - eingetreten, inklusive Dialekt. Infrastruktur schlägt Status quo - bestätigt:

Wer **vor zwei Jahren investiert** hat, baut **personalisierte Podcasts in drei Monaten**; wer wartete, wartet immer noch. Personalisierung technisch möglich - ja, mit Voraussetzungen. Qualitativ überzeugende Open-Source-TTS-Modelle auf Deutsch - anders als erwartet noch nicht da. Der Dialog mit dem Artikel - technisch möglich, im Alltag noch nicht. Neue Vermarktungsmodelle - leider langsamer als antizipiert.

Liquid Content: die Entertainment Perspektive

Dass modulares Denken nicht nur ein Verlagsthema ist, zeigte das Gespräch mit Philipp Münich von ProSiebenSat.1, wo es der Begriff „**Liquid Content**“ spätestens seit dem ersten AI Audio Day bis in die Konzernstrategie geschafft hat. Die Audio-Geschichte des Konzerns reicht von der Podcast Factory (ca. 2018) über eine eigene Plattform (FYEO) bis zur heutigen Strategie: Premium-Podcasts vermarkten durch die Übernahme des Studio Bummens und seit 2025 auch eigene produzieren. Unabhängig davon wird eine KI-Audio Schiene angedacht. Podcasts sind bei ProSiebenSat.1 ein Premium-Produkt mit Gesichtern und Fangemeinde, während KI-Audio die zweite, neue Säule sein kann.

Die strategische Logik heißt intern „Everything, everyone, everywhere“. Das bedeutet, dieselben Content-Assets über alle Zielgruppen, Nutzungsszenarien und Plattformen maximal streuen. Denn das klassische Bewegtbild verliert Menschen in genau den **Momenten, die Audio gehören** - beim Joggen, Kochen, Pendeln. Audio erschließt eine komplementäre Reichweite. Konkret denkbar sind Companion-Formate und Zusammenfassungen zu langen Sendungen, Spin-offs starker Marken und perspektivisch sogar die Lieferung von Inhalten an Dritte. Die maximale Ausbaustufe beschrieb Münich als „Entertainment Data Cloud“: nicht mehr in Formaten planen, sondern ein **Content-Piece als Asset** begreifen, das in unterschiedlichen Längen, für unterschiedliche Zielgruppen, auf unterschiedlichen Plattformen oszilliert.

Das geht bis hin zur Idee, mit KI aus hunderten Folgen einer Daily Novela einzelne Handlungsstränge zu extrahieren und als Short-Drama für Social auszuspielen. Vor zehn Jahren scheiterte Eins-zu-eins-Personalisierung an der Produzierbarkeit; dieses Problem ist mit KI jetzt gelöst. Die größere Hürde ist kulturell: die emotionale Barriere, das Denken in festen Formaten loszulassen. Und eine rote Linie zog Münich deutlich. Die eigene Premium-Marke darf durch KI-Ausspielung nicht beschädigt werden.

Take Aways – Kapitel 3

- Das Audio-Modul ist die neue Grundeinheit: einmal produziert, beliebig oft rekombiniert, vom E-Paper bis zum in acht Sekunden erstellten personalisierten Podcast.
- Der Engpass ist vom Klang zum Skript gewandert: Fürs Hören geschriebene Texte sind der Hebel für Qualität, Fehlerarmut und redaktionelle Freigabefähigkeit.
- Liquid Content gilt vom Regionalverlag bis zum Entertainment-Konzern. Die Technik steht, die Herausforderung ist kulturell.



4 Markt & Daten: Was Audio Intelligence verrät

Inspiziert von Redebeiträgen durch Max Franke (Podius) und Michael Mörs (Podmon)

Der Podcast-Markt 2026 im Reality-Check

Max Franke, CEO der Podcast-Analytics- und Mediaplanungs-Plattform Podius, lieferte die Marktvermessung des Tages entlang von Hörermarkt, Angebot und Werbemarkt.

Hörermarkt: 63 Prozent der 18- bis 49-Jährigen hören Podcasts; im vergangenen Jahr hat der Podcast das Radio in der Nutzung dieser Zielgruppe erstmals überholt. Der Markt wächst doppelt: Neue Hörer kommen dazu, zunehmend auch in älteren Zielgruppen, und wer einmal hört, hört immer mehr. Inhaltlich liegen Nachrichtenformate mit Abstand vorn. Das Fundament sind die besonderen Attribute des Mediums: hohe Glaubwürdigkeit, die parasoziale Verbindung zwischen Host und Hörer, und die Rolle als Gegenpol zur verkürzten Aufmerksamkeit. Die Folge: Mehr als ein Drittel der Hörer nennt Podcasts eine Hauptquelle für Informationen zum Zeitgeschehen. Podius' These: Der Podcast entwickelt sich zum Leitmedium.

Video-Frage: Muss jetzt jeder Podcast Video sein? Die Daten mahnen zur Differenzierung. Von rund 31.000 aktiven deutschsprachigen Podcasts hat etwa ein Fünftel mindestens eine Videoepisode; rund 10 Prozent veröffentlichen die Hälfte ihrer Episoden als Video, 6 Prozent sind video-only.

Auf Konsumentenseite geben 40 Prozent an, Videopodcasts zu nutzen, aber ein Drittel dieser Nutzer beachtet das Video kaum oder schaltet es aus. Video wird wichtiger, gerade für Sichtbarkeit und Teilbarkeit, ob es in Deutschland zum Synonym für Podcast wird wie in den USA, ist offen.

Werbemarkt: Von den rund 31.000 aktiven Podcasts werden nur etwa 2.200 vermarktet. Das entspricht nur rund 7 Prozent. Host-Reads dominieren mit rund zwei Dritteln der Werbeerlöse bei international beneidenswert hohen TKPs. Die Schwachstelle ist der Mittelbauch. Hier braucht es skalierende Formate wie Presenter-Sponsorings und Rotationsbuchungen. Zur Gesamtgröße kursieren unterschiedliche Zahlen von rund 75 Millionen Euro über rund 100 Millionen bis zu einem Report, der den deutschen Podcast-Werbemarkt bei rund 239 Millionen US-Dollar sieht. Der Maßstab im Medium Audio bleiben die rund 710 Millionen Euro des Radiowerbemarkts. Frankes Einordnung: Podcasts folgen dem klassischen Medien-Playbook, erst der Goldrausch, dann die Konsolidierung, dann die Kreativität. Wir treten in Phase drei ein.

Der Audio-Footprint: Regionalität wird messbar

Michael Mörs baut mit Podmon seit 2021 so etwas wie ein „Google Alerts für Podcasts“: eine Datenbank von rund 200.000 deutschsprachigen Podcasts, davon rund 31.000 aktiv, mit über drei Millionen durchsuchbaren Episoden und etwa 5.000 neuen pro Tag. Der Kern-Use-Case: Marken, PR- und Medienbeobachter finden heraus, wo, wie und in welchem Kontext über sie gesprochen wird. Die Klassifizierung arbeitet in 30-Sekunden-Schritten über 140 Themenkategorien und trennt relevante thematische Auseinandersetzung vom bloßen Nebensatz-Rauschen.

Für den AI Audio Day hatte Mörs eine eigens erstellte Analyse mitgebracht, deren Frage bei einer Zugfahrt entstand: Lässt sich Regionalität in Podcasts überhaupt messen? Denn im RSS-Feed steht nicht, wo ein Podcast zu Hause ist. Über Hilfsindikatoren (Städtenamen, regionale Bezeichnungen, lokale Marken und Vereine, Mundart) analysierte Podmon den „Audio-Footprint“ von vier Städten über zwei Monate: Erfurt, Jena, München und Regensburg.

Die Befunde: Jede Stadt hat ein eigenes, unterscheidbares Audio-Profil. Erfurt ist politisch geprägt, Jena wissenschaftlich, Regensburg kulturell und München ist ein Resonanzraum mit 20 mal mehr Erwähnungen als die Vergleichsstädte. In diesem Raum lässt sich kaum ein spezifischer Footprint isolieren. Die Profile sind dynamisch und reagieren auf Ereignisse, wobei Fußball alles überlagert. Tipp: Wer Regionalität messen will, muss Sport erst herausfiltern.

Ein methodischer Nebenbefund: Mörs bezweifelt den einfachen linearen Zusammenhang, dass Themen erst durch Social Media und dann in Podcasts wandern. Laut ihm sei die **Beziehung reziprok**. Die für Verlage wichtigste Erkenntnis steckt in der Zusammensetzung des lokalen Angebots: Ein erheblicher Teil dessen, was regional nach Podcast aussieht, ist recyceltes Radio, also stündlich wiederholte Nachrichten, die das Bild verzerren.

Der Anteil an echtem, vor Ort produziertem lokalem Podcast-Content ist klein; inhaltlich unterscheiden sich beide Welten deutlich. Radio ist gleichförmig (Sicherheit, Politik, Wirtschaft), originärer **Podcast-Content facettenreich** und **regional individuell**. Das Narrativ, dass Städte eigene Content-Hubs mit eigenem Profil werden können, ist bestätigt — und es gibt **viel unbesetzten Raum für** alle, die echten **lokalen Audio-Content** bauen wollen.

Ergänzend zeigte ein Kundencase die Skalierung: Für das Stadtmarketing Düsseldorf analysierte Podmon nach der Fußball-EM 17.400 Podcasts mit 170.000 Episoden aus den Teilnehmerländern und fand über 40.000 Erwähnungen Düsseldorfs im Fußballkontext — kondensiert zu einer Auswertung, wie internationale Gäste die Stadt erlebt und beschrieben haben.

Take Aways – Kapitel 4

- Der Podcast wird Leitmedium: Er hat das Radio bei den 18- bis 49-Jährigen überholt und ist für über ein Drittel der Hörer Hauptinformationsquelle, während der Werbemarkt noch hinterher hinkt.
- Video wächst rasant im Angebot, wird aber von einem Drittel der Nutzer kaum beachtet.
- Regionalität ist messbar geworden: Städte haben eigene Audio-Profile, und der Markt für echten lokalen Podcast-Content ist nachweislich unterversorgt.



Was bis zum AI Audio Day 2027 wichtig wird

Verdichtet man die Prognosen des Tages, entsteht ein klares Bild der nächsten zwölf Monate. Die ersten personalisierten Audio-Formate erreichen den Regelbetrieb, nicht als Experiment, sondern als Produkt auf Basis der Modul-Infrastruktur, die jetzt gebaut wird. ProSiebenSat.1 will bis zum nächsten AIAD erste Sendungs-Companions und Spin-offs im Markt haben; die WELT skaliert „Going Local“ auf weitere Städte; die NOZ Digital bringt ihr forschungsbasiertes Briefing-MVP in die App.

Der Dialog mit dem Inhalt – die Zeitung, die man anrufen kann – bleibt dagegen voraussichtlich noch ein bis zwei Entwicklungsstufen entfernt, ebenso wie überzeugende deutschsprachige Open-Source-Stimmen. Und die Vermarktung wird zur entscheidenden Baustelle: Die Nutzung ist da, die Erlösmodelle für KI-Audio-Formate und Podcasts im Allgemeinen entstehen langsamer als die Produkte.



Audio-as-a-Service & Liquid Content

ARTICLY baut die Infrastruktur, mit der Verlage, Medienhäuser und Unternehmen aus ihren bestehenden Inhalten modulares KI-Audio machen – von der fürs Hören geschriebenen Audio-Summary über lokale Briefings bis zum vollautomatisch personalisierten Podcast. Nicht der Vorleseknopf, sondern das nahtlose Zusammenspiel von Lesen und Hören steht im Zentrum.

Referenzen und laufende Projekte reichen von Regionalverlagen wie der NOZ, Mediahuis oder Pforzheimer Zeitung bis zu Partnerschaften mit Audi oder der Deutschen Bahn. Auf Wunsch zeigen wir in einer kurzen Demo, wie wir in wenigen Minuten aus Ihren Inhalten Hörerlebnisse machen – Faktengeprüft und ohne redaktionellen Mehraufwand.



Dank & Ausblick 2027

Der AI Audio Day 2026 wurde ermöglicht durch die Premium-Partner **Velora** und **Podmon**

Save the date

AI AUDIO DAY:
15. Juli 2027

Fotogalerie



Speaker- und
Partneranfragen



wolf@articly.de

