

GEISST

GENERATOR
FÜR EMOTIONAL
INDIVIDUALISIERBARE
SYNTHETIK-STIMMEN

Prof. Dr. Petra Grimm | Susanne Kuhnert M. A. | Marcel Schlegel M. A.
November 2025 • Zweite, überarbeitete Fassung

Synthetische und geklonte Stimmen im Nachrichtenjournalismus

Ethisches Gutachten und ethische Leitlinien für das
Forschungsprojekt GEISST – gefördert durch das BMFTR

Förderkennzeichen: FKZ 01|S23014 A-C



IDE

HOCHSCHULE
DER MEDIEN

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt



Synthetische und geklonte Stimmen im Nachrichtenjournalismus

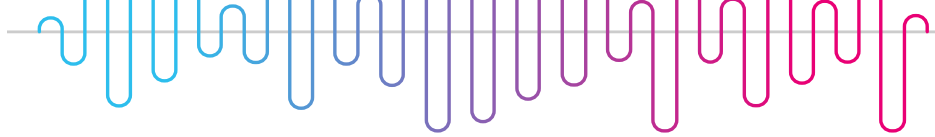
Technologischer Wandel im Nachrichtenjournalismus

Der Journalismus befindet sich im Zuge der Digitalisierung, dem Aufkommen der Sozialen Medien und zuletzt durch die Einführung von Künstlicher Intelligenz (KI) in die Medienproduktion in einem tiefgreifenden strukturellen Wandel. Neben algorithmengestützten Recherchesystemen, automatisierten Textgeneratoren und personalisierten Distributionsmechanismen gewinnen auch synthetische Sprachsysteme zunehmend an Bedeutung. Diese KI-basierten Systeme ermöglichen die Herstellung von menschlich klingenden Stimmen durch die sogenannte Text-to-Speech-Technologie (TTS) sowie das Klonen von Stimmen tatsächlich existierender Menschen. Insbesondere im audiovisuellen (Online-) Journalismus eröffnen synthetische Stimmen zahlreiche Möglichkeiten – zum Beispiel im Kontext von News-Podcasts, Sprachassistenzsystemen, der automatisierten Erzeugung von Nachrichtensendungen oder der barrierefreien Aufbereitung journalistischer Inhalte. Sie erlauben unter anderem eine kosteneffiziente und skalierbare Produktion, eine hoch personalisierte Ansprache unterschiedlicher Zielgruppen sowie die schnelle Übersetzung von Inhalten in verschiedene Sprachen. Mithilfe von KI-Systemen könnten Medienanbieter künftig also Arbeits- und Produktionsabläufe automatisieren und dadurch vereinfachen, Ressourcen (Kosten, Zeit, Personal etc.) einsparen und somit Effizienzen heben.

Das Forschungsprojekt „GEISST – Generator für emotional individualisierbare Synthetik-Stimmen“, gefördert vom Bundesministerium für Forschung, Technologie und Raumfahrt (BMFRT), hatte das Ziel, KI-basierte Technologien in der Audiotbearbeitung sowie in Bezug auf mögliche Einsatzszenarien von synthetischen und/oder geklonten Stimmen im Nachrichtenbereich aus einer ethischen Perspektive zu betrachten. Als Grundlage für die Einschätzung wurden bestehende digital- und journalismusethische Prämissen berücksichtigt. Die Erkenntnisse, die, dem werteesensitiven Ansatz des Value Sensitive Design folgend, auch durch empirische Verfahren und unter Einbindung relevanter Bezugsgruppen (Interviews mit Medienschaffenden) gewonnen wurden, wurden schließlich in ethische Leitlinien für einen sozialverträglichen und moralisch redlichen Einsatz von Generativer KI (GenKI) im Nachrichtenjournalismus überführt. Die Empfehlungen werden im vorliegenden Gutachten vorgestellt. Sie können, wie es der anwendungsorientierte Ansatz des Ethics By Design vorschlägt, (proaktiv) in künftige Medienprojekte einfließen.

Jüngere Zielgruppen konsumieren Nachrichten kaum noch über traditionelle Einzelmedien (TV, Print, Radio etc.), sondern empfangen auch journalistische Informationen häufig über Soziale Medien. Ein Trend, der sich seit Jahren allgemein feststellen lässt: Über alle gesellschaftlichen (Alters-)Gruppen hinweg zeigen sich markant veränderte Mediennutzungsgewohnheiten – hin zu Online-Medien. Der nachrichtlich ausgerichtete





Journalismus kann sich nun gezwungen sehen, seine Produktions- und Distributionsmodalitäten an das mediale Verhalten seiner Rezipienten und Rezipientinnen anzupassen. Eine Folge könnte sein, dass verschiedene Versionen eines Medieninhalts entlang von zielgruppenspezifischen Gewohnheiten, Bedarfen und Präferenzen (Kanal, Format, Länge etc.) erstellt und verteilt werden. Um diesem enormen Konzeptions- und Produktionsaufwand, der bislang durch menschliche Arbeit bewerkstelligt wird, zukünftig besser begegnen zu können, wird der Einsatz von Software aus dem Bereich der Generativen KI (GenKI) erprobt.

KI-Stimmen: Begriffsbestimmung und Relevanz

Im Folgenden verwenden wir „synthetische Stimme“ als Oberbegriff für computergenerierte Sprache, unabhängig davon, ob es sich dabei um eine technisch neu geschaffene oder um eine geklonte Stimme handelt. Beide Varianten sind technisch verwandt, beide fußen auf KI-Verfahren, unterscheiden sich allerdings in ihrer sozialen Reichweite, im Grad ihrer menschlichen Bezogenheit und im Hinblick auf ethische Ableitungen. Während synthetische Stimmen keine Entsprechung in der Realität aufweisen, basieren KI-Klone auf Stimmen lebender Menschen. Bislang werden primär öffentlich bekannte Personen geklont, also Prominente, Moderatoren, Nachrichtensprecherinnen, Schauspieler etc. Gleichwohl lässt sich im Grunde von jeder Person ein Stimmklon generieren. Vereinfacht gesprochen werden dafür Aufnahmen einer Originalstimme technisch repliziert, dies mit KI-Verfahren wie dem „Voice Cloning“. Aus dem unterschiedlichen Grad an menschlicher Echtheit ergeben sich unterschiedliche ethische Anforderungen: unter anderem im Hinblick auf personenbezogene Werte (Privatheit, Authentizität, Transparenz, Integrität, Identität etc.) sowie dazugehöriger Prinzipien (Kontextintegrität, Zugriffskontrolle, Wahrhaftigkeit usw.) und auch bezüglich rechtlicher Vorgaben (Entscheidungsfreiheit bzw. Autonomie im Hinblick auf Verwertung und Verwendung der Stimme der stimmgebenden Person, Datenschutz, Urheberrecht). Beispielsweise könnte eine bekannte TV- oder Radio-Journalistin den legitimen Bedarf äußern, dass ihre geklonte Stimme möglichst einen Wiedererkennungswert aufweist, sprich mit ihrer natürlichen Stimmen weitgehend übereinstimmt. Auch könnte sie Sorge darum haben, dass ihr Replika für unerwünschte Zwecke verwendet oder in nicht vorhergesehen Zusammenhänge gerät.

Aktuelle Entwicklungen und Beispiele

Synthetische Stimmen haben längst Eingang gefunden in die Medienproduktion. Sie kommen in Sozialen Medien ebenso vor wie im Angebot traditioneller Journalismus- oder



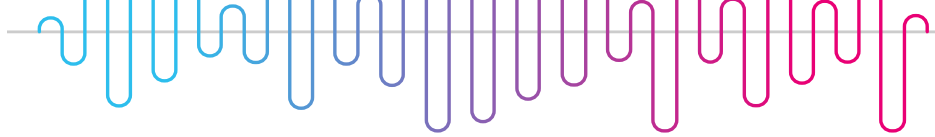


Medienanbieter, erstellt durch Akteure aus privaten Medienorganisationen genauso wie durch den öffentlich-rechtlichen Rundfunk. Der Einsatzbereich reicht dabei von der Verwendung von Voice Clones als Sprecherstimmen in Radioformaten oder Podcasts, über Kommentierungen von audiovisuellen Beiträgen in den Social Media bis hin zu ersten Videobeiträgen, die (anteilig) auf künstliche Avatare setzen, welche dann mit der sogenannten „Deep Fake“-Technologie¹ generiert wurden. Auch ist davon auszugehen, dass gänzlich neue Darstellungsformen entstehen, für die KI-Technologie die Grundlage bietet. Daran zeigt sich, dass der Einsatz von KI im Online-Journalismus erhebliche Potenziale aufweist. Gerade Anwendungen aus dem Bereich der GenKI könnten unter anderem dafür eingesetzt werden, Medienproduktionen zu vereinfachen, diese günstiger, schneller und vor allem automatisiert umzusetzen. Darunter fällt auch die Möglichkeit, Konzepte für Texte, Videos oder Audios zu erstellen, zu strukturieren oder diese Medienerzeugnisse herzustellen, zu überprüfen und schließlich kanal- sowie zielgruppengerecht auszuspielen. Auch als Recherche- oder Analysetool kann KI verwendet werden. Dies, weil riesige Datensätze die Grundlage für derlei Modelle bieten und letztere technisch fähig dazu sind, erhebliche Mengen von Informationen zielgerichtet auszuwerten.

Gleichzeitig häufen sich Berichte, die von einer zunehmend missbräuchlichen Verwendung von KI zeugen. So lassen sich mit GenKI-Tools auch gezielt Desinformationen generieren, mediale Debatten verfälschen, öffentliche Diskurse stören und (indirekt) politische Prozesse sabotieren. Ebenso mit geklonten Stimmen oder Deep Fakes, mit denen beispielsweise auch Erpressungskampagnen umgesetzt werden können. Dies etwa, wenn Prominente zum Opfer von Kriminellen werden, die auf gefälschten Social-Media-Accounts und mithilfe von KI-generierten Klonen destruktive Inhalte veröffentlichen. Entwicklungen, die gerade für den Journalismus – als maßgebliche Institution der öffentlichen Kommunikation – in vielerlei Hinsicht eine Herausforderung darstellen. Bezieht man diese Entwicklungen zum Beispiel auf den Umstand, dass die Ähnlichkeit von synthetischen/geklonten mit menschlichen Stimmen durch die rasant fortschreitende technische Optimierung der KI-basierten Sprachsynthese sukzessive zunimmt, stellen sich Fragen im Hinblick auf typische Werte der zwischenmenschlichen wie medialen Kommunikation – etwa nach Wahrheit und Wahrhaftigkeit oder nach Vertrauen und Vertrauenswürdigkeit. Jedenfalls dürften sprachsynthetische Erzeugnisse dem Publikum die Unterscheidung zwischen faktischer oder medial vermittelter Realität und deren technisch-fiktiven Nachahmung erschweren, das heißt: zwischen authentischen Wiedergaben von Wirklichkeit und deren auch bewusst veränderten Kopien.

¹ Deep Fakes sind mithilfe von KI generierte oder manipulierte audiovisuelle Inhalte. Meistens werden Personen oder Lebewesen in Form von Videos oder Bildern realistisch nachgebildet. Ziel ist es, reale Szenen täuschend echt zu imitieren oder zu verändern. Dabei werden große Datenmengen genutzt, um die mimischen, sprachlichen und gestischen Muster einer Zielperson zu erlernen und anschließend künstlich zu reproduzieren.





Ethische Relevanz von Stimmengenerierung

Allgemein gesprochen stellen Stimmen weder rein funktionale Werkzeuge der Informationsübermittlung dar, noch sind sie von neutraler Qualität. Vielmehr repräsentieren, charakterisieren und individualisieren menschliche Stimmen eine Person in ihrer Persönlichkeit. Da jede Stimme so einzigartig wie ein Fingerabdruck ist, fungiert sie in sozialen Kontexten als Identifikations- und Wiedererkennungsmerkmal, das untrennbar mit einer bestimmten Person verbunden bleibt. In diesem Sinne kann die Stimme als selbstreflexives Merkmal des Menschen verstanden werden: als Ausdruck seiner Identität, mit dem er sich selbst identifiziert und von anderen identifiziert werden kann. Aufgrund dieser Unverwechselbarkeit, die selbst für den oder die jeweilige(n) Stimmträger oder Stimmträgerin schwer zu verändern ist, lässt sich die Stimme als Manifestation von Authentizität begreifen. Sie verweist auf die leiblich verankerte, nicht substituierbare Dimension menschlicher Präsenz und Individualität: Sie unterscheidet Menschen von unbelebten Artefakten.

In der auditiven Wahrnehmung übernehmen Stimmen zentrale soziale Funktionen. In persönlichen Begegnungen bilden sie häufig den ersten Zugang zu einer anderen Person. Sie stiften Beziehungen, erzeugen Nähe oder Distanz und prägen wesentlich die Wahrnehmung des Gegenübers. Über die Stimme – und auch über Sprache – transportieren Menschen Vertrauen, Sympathie und Zugehörigkeit – etwa durch individuelle Sprachklänge, geteilte Dialekte oder charakteristische Sprechweisen. Damit tragen Stimmen Spuren von Herkunft in sich und vermitteln ferner weitere personenbezogene Metainformationen, beispielsweise biografische oder demografische Merkmale (Alter, körperliche Verfassung usw.). In Tonalität, Rhythmus und Ausdrucksweise beinhalten Stimmen zwangsläufig emotionale Nuancen, die über das Gesagte hinausweisen. So gelangen Gefühle, Stimmungen und Beziehungseinstellungen in einer Stimme buchstäblich zum Ausdruck. In diesem Sinne fungieren sie nicht nur als Indikatoren für Aufrichtigkeit, sondern etablieren bei Adressatinnen und Adressaten auch Wahrnehmungen von Wahrhaftigkeit, Vertrauenswürdigkeit, Glaubwürdigkeit und stiften soziale Verbundenheit, wodurch sie die Stabilität zwischenmenschlicher Beziehungen mit absichern. Die zentrale Bedeutung einer ethischen Reflexion von KI-basierter Sprachsynthese, auch im Journalismus, begründet sich zum einen darin, dass KI-erstellte Stimmen in einen grundlegenden Bereich der menschlichen Kommunikation eingreifen: die gesprochene Sprache. Auch Sprache ist nicht nur Träger von Information, sondern auch Ausdruck von individueller und sozialer Identität, die immer auch emotionale Aspekte beinhaltet.

GenKI kann Sprache und Stimmen nun technisch reproduzieren. Damit verschiebt sich die Herstellung von gesprochener Sprache von einem menschlich-körperlichen Ausdruck hin zu einem technisch vermittelten Prozess. KI-Stimmen können dadurch das Erleben von Sprache verändern – und damit die zwischenmenschliche Kommunikation selbst, auch solche, die medial erstellt oder vermittelt ist.





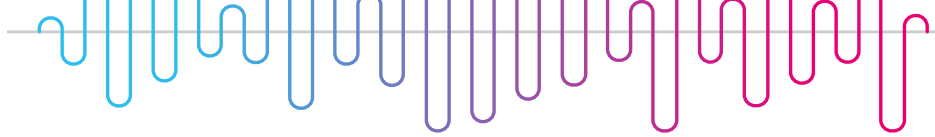
In journalistischen Kontexten berühren synthetische Stimmen nun grundlegende Werte und Prinzipien der journalistischen Berufsethik und der öffentlichen Kommunikation (siehe S. 11 und 12), mit denen sich die journalistische Rolle in Demokratien sowie die Qualität journalistischer Erzeugnisse einerseits absichern und andererseits begründen lassen. Im Journalismus muss Sprache, auch in ihrer in Text gegossenen Form, daher als ein zentrales Instrument von Integrität aufgefasst werden. Kommt bei Empfängerinnen oder Empfängern von Medienbotschaften nun Zweifel an der Echtheit der Stimme bzw. an ihrem (vermeintlichen) Träger auf, kann darunter das Medienvertrauen leiden – ein Nährboden für Verunsicherung und Misstrauen. Dies kann sich im Folgenden auf die Einschätzung der vermittelten Informationen selbst auswirken, weiterhin auf die journalistische Senderinstanz und, zu Ende gedacht, auf die Institution des Journalismus in Gänze. Wie kann ich wissen, dass eine Botschaft faktisch wahr ist, wenn ich die Echtheit, Wahrhaftigkeit, Vertrauenswürdigkeit und hiernach Glaubwürdigkeit des Senders nicht einschätzen oder sie als unecht deklarieren muss? – So ließe sich ein solcher Zweifel vereinfacht formulieren. Die Stimme steht somit an der Schnittstelle von Körper, Identität und Technologie. Ihre technische Reproduzierbarkeit durch KI eröffnet neue Möglichkeiten, wirft aber zugleich Fragen nach der Grenze zwischen menschlicher Präsenz und maschineller Simulation auf. Eine verantwortungsvolle Auseinandersetzung mit synthetischen Stimmen verlangt daher nicht nur technische, sondern auch ethische und kulturelle Perspektiven.

In einer technisierten Umgebung, in der Stimmen synthetisch erzeugt, imitiert oder verändert werden können, stellt sich die Frage nach der Integrität dieses Ausdrucks. Eine Stimme ist nicht nur ein beliebiger akustischer Output, sondern ein Ausdruck von leiblicher Präsenz und zugleich von personaler Verantwortung. Die Stimme vermittelt Inhalte und Botschaften und sie transportiert Antworten. Die Verwendung von geklonten Stimmen im Rahmen technischer Systeme berührt daher grundlegende ethische Fragen nach Verantwortung, Integrität und Authentizität.

Wenn die Stimme, beispielsweise von einer etablierten Nachrichtensprecherin, durch eine synthetische Version ersetzt oder geklont wird, werden grundlegende ethische Fragen tangiert:

- Wie verändert sich das Vertrauensverhältnis zwischen Medium und Publikum?
- Welche Rolle spielt Authentizität in der Rezeption von Nachrichten?
- Wie transparent müssen synthetische Stimmen gekennzeichnet sein?
- Wem gehört eine Stimme aus rechtlicher Sicht und aus einer ethisch-moralischen Perspektive?





Diese Fragen berühren zum einen Grundprinzipien der journalistischen Ethik. Zum anderen betreffen sie rechtliche Implikationen (Persönlichkeitsrechte, Urheberrechte etc.) und nicht zuletzt kulturelle Aspekte – wie jene nach der Repräsentation menschlicher Merkmale in den und durch die Medien.²

Journalismus zwischen Innovation und Vertrauenskrise

Die Entwicklungen rund um synthetische und geklonte Stimmen sind Teil einer größeren Transformation journalistischer Produktionsprozesse. Sie verweisen auf ein neues Spannungsverhältnis zwischen technologischem Fortschritt und einem wachsenden Vertrauensverlust. Während automatisierte Sprachsysteme enorme Potenziale zur Demokratisierung von Information bieten, insbesondere durch Barrierefreiheit, Mehrsprachigkeit und Skalierbarkeit, laufen Nachrichtenmedien zugleich Gefahr, ihren Vertrauensanker zu verlieren, wenn die Authentizität medialer Erzeugnisse, das heißt deren menschliche Urheberschaft, nicht mehr eindeutig erkennbar ist. Der journalistische Beruf lebt von Stimmen, die für etwas stehen. Wenn diese Stimmen durch Maschinen ersetzt oder manipuliert werden, muss neu verhandelt werden, wie Vertrauen, Verantwortung und Wahrheit in einer digitalisierten Öffentlichkeit definiert werden.

Ein zentrales Problem im Nachrichtenbereich liegt in der wachsenden Verunsicherung der Rezipientinnen und Rezipienten, da reale Inhalte zunehmend schwerer von KI-generierten Inhalten zu unterscheiden sind. Besonders deutlich zeigt sich dies bei synthetischen Stimmen: Wo vormals die menschliche Stimme unmittelbar als ein authentisches Zeichen für eine bestimmte Präsenz und als ein Indiz von Glaubwürdigkeit galt, kann nunmehr zunehmend ein Moment der Verunsicherung, der Zurückhaltung oder des Misstrauens entstehen, ausgelöst von wahrheits- und wahrhaftigkeitsbezogenen Fragen der Rezipientinnen und Rezipienten. Stammt diese Information von einer verlässlichen, ja menschlichen und vor allem professionell agierenden Quelle? Und wenn nicht, ist diese Information dann wahr, wenn ich die Wahrhaftigkeit des Senders nicht mehr selbst einschätzen kann. Derlei Fragen könnten sich stellen. Hörende könnten vereinzelt nicht mehr sicher sein, ob es sich um die Stimme einer realen Person handelt oder um ein KI-generiertes Produkt. Das schwindende Vertrauen und die Ununterscheidbarkeit von menschlichen und synthetischen Stimmen untergräbt die vormals dominant vorherrschende intuitive Sicherheit der Rezipientinnen und Rezipienten zunehmend und kann das Fundament der Glaubwürdigkeit im Journalismus nachhaltig schwächen.

² Im Forschungsprojekt GEISST wurde für die rechtlichen Fragen ein eigenes Rechtsgutachten von Dr. Elisabeth Clausen-Muradian (armedis Rechtsanwälte, Hannover) erstellt: „Die rechtliche Einordnung des KI-basierten Sprach-Synthese-Systems GEISST. Rechtsgutachten im Auftrag der nachtblau GmbH im Rahmen des Forschungsprojektes GEISST – Generator für emotional individualisierbare Synthetik-Stimmen“.





Medienethische Einschätzung zum Unterschied beim Einsatz von künstlich generierten Stimmen im Entertainmentbereich und im Nachrichtenjournalismus

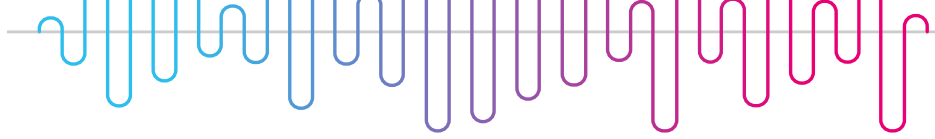
Aus medienethischer Sicht unterscheiden sich der Nachrichtenjournalismus und der Entertainmentbereich in ihren normativen Grundlagen, Zielsetzungen und Publikumserwartungen grundlegend. Der Einsatz von synthetischen Stimmen unterliegt daher in beiden Feldern unterschiedlichen ethischen Maßstäben.

Unter Unterhaltungsmedien (Entertainment) verstehen wir mediale Inhalte, deren vorrangige Funktion die Vermittlung von Emotionen, Stimmungen, Erfahrungen und Erzählungen ist. Im Zentrum steht dabei die Zerstreuung der Rezipierenden, primär durch narrative Inhalte. Zugleich erfüllen Unterhaltungsmedien häufig soziale und kulturelle Funktionen, etwa indem sie gemeinschaftsstiftende Bezugspunkte schaffen, Identitätsangebote bereitstellen oder reflexive Auseinandersetzungen mit gesellschaftlichen Themen anregen. Obwohl Unterhaltung dabei im Vordergrund steht, können Unterhaltungsmedien somit auch Bildungs- und Orientierungsleistungen erbringen. Anders als im Journalismus müssen sich Angebote aus dem Bereich des Entertainments nicht notwendigerweise auf überprüfbare Fakten oder gesellschaftlich relevante Themen beziehen lassen. Gleichwohl *können* sie einen Bezug zur tatsächlichen Realität und bspw. zu politischen Zusammenhängen aufweisen. Dazu zählen etwa Filme, Serien, Hörspiele, Computerspiele oder auch unterhaltende Formate in Radio oder Sozialen Medien. Der Nachrichtenbereich hingegen erfüllt eine andere, eine für die Stabilität von Demokratien fundamentale Funktion. Hier steht die Bereitstellung verlässlicher Informationen anhand von überprüfbaren Fakten über aktuelle Ereignisse im Vordergrund, die für die politische Meinungsbildung und für das kollektive Urteilsvermögen relevant sind. Während Unterhaltung ein legitimes Feld künstlerischer und kommerzieller Gestaltung ist, beruht der Nachrichtenjournalismus auf den Grundprinzipien Faktizität, Glaubwürdigkeit und öffentlichem Auftrag. Diese Differenz bildet den Maßstab, an dem auch der Einsatz synthetischer Stimmen jeweils gesondert zu bewerten ist.

Normative Zielsetzung und Wahrheitsanspruch

Der Nachrichtenjournalismus operiert im Rahmen eines gesellschaftlichen Auftrags, der sich aus der Rolle der Medien in einer demokratischen Öffentlichkeit ableitet. Zentrale Werte und Normen wie Wahrhaftigkeit, Sorgfalt, Transparenz oder Verantwortlichkeit sind nicht nur berufsethische Ideale, sondern auch die konstitutiven Elemente der wesentlichen Funktion, die Nachrichten im öffentlichen Diskurs übernehmen. Eine Stimme im Nachrichtenkontext stellt nicht nur ein technisches Transportmittel für Informationen, sondern auch einen Teil der „Be-





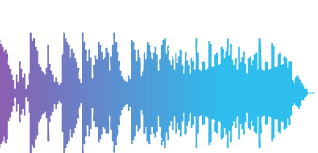
glaubigung“ des Gesagten dar: Sie trägt Authentizität, verweist auf eine verantwortliche Redaktion und steht in einer indexikalischen Beziehung zu einer realen, überprüfbaren Quelle.

Im Entertainment-Bereich hingegen gilt kein systematischer Wahrheitsanspruch. Hier steht die ästhetische, emotionale oder narrative Wirkung im Vordergrund. Die Stimme erfüllt in erster Linie performative Zwecke und dient der Erzeugung einer gewünschten Stimmung oder der glaubhaften Verkörperung einer Figur. Die Bindung an die Wirklichkeit ist aufgehoben und selbst bei realitätsnahen Darstellungen darf inszeniert, stilisiert oder etwas erfunden werden, solange dies dem intendierten ästhetisch-narrativen Zweck dient.

Publikumserwartung und Vertrauensdimension

Das Publikum tritt an Nachrichten und deren Sender mit einer gewissen Erwartungshaltung heran. Beispielsweise möchten Zuschauer, Leserinnen usw. sich darauf verlassen können, dass nachrichtliche Informationen zuverlässig recherchiert, faktisch richtig und objektiv ausgelegt sind. Stimmen werden in diesem Zusammenhang – bewusst oder unbewusst – als Träger einer redaktionellen Identität wahrgenommen. Eine synthetische Stimme im Nachrichtensegment verändert diese Beziehung: Sie kann den Eindruck erwecken, dass es keine Personen mehr gibt, die unmittelbar hinter den gesprochenen Worten stehen. Diese Entkoppelung kann sich auf die sendende Instanz auswirken und potenziell schwächen, insbesondere wenn Herkunft und Erzeugungsprozesse nicht offengelegt werden. Dem ist so, weil deren Vertrauenswürdigkeit, Wahrhaftigkeit und Glaubwürdigkeit (mitsamt der von ihr transportierten Botschaften) nicht mehr eingeschätzt werden können, wenn sie sich von einem Menschen entfernt.

In Unterhaltungsformaten erwarten Rezipientinnen und Rezipienten hingegen keine Authentizität im journalistischen Sinn, sondern Kohärenz mit der erzählten Welt. Eine synthetische Stimme könnte hier sogar als kreatives Stilmittel begrüßt werden, etwa wenn sie zur Charakterisierung einer Figur beiträgt oder neue künstlerische Ausdrucksformen eröffnet. Das Vertrauen des Publikums bezieht sich in diesem Fall nicht auf Faktentreue, sondern auf die Qualität der Inszenierung. Trotzdem stellt die Stimme – etwa von bekannten Schauspielerinnen oder prominenten Synchronsprecherinnen – auch in Unterhaltungsformaten ein eindeutiges Identifikationsmerkmal dar, das gleichzeitig Attribute vermittelt, die mit diesen Personen in Verbindung gebracht werden. Dieser Wiedererkennungswert kann für Vertrauen sorgen und die Beziehung zwischen Medienakteur und Publikum festigen. Wird diese Relation indes durch die nicht-gekennzeichnete oder sogar kriminelle Verwendung eines KI-Klons gestört, kann dieses auf Verlässlichkeit, Glaubwürdigkeit und Authentizität basierende Vertrauensbeziehung geschädigt werden und damit der Vertrauensvorschuss insgesamt, den Mediennutzende medial vermittelten Personen notwendigerweise entgegenbringen müssen.





Verantwortung und Transparenzpflicht

Nachrichtenredaktionen unterliegen einem hohen Maß an ethischer Verantwortung. Sie müssen sicherstellen, dass eine gesendete Information nicht nur faktisch korrekt, sondern auch verantwortungsvoll zustandekam (Recherche, Kontrolle, Produktion). Dazu gehört, dass der Produktionsweg für das Publikum nachvollziehbar ist, insbesondere wenn neue Technologien wie KI-Stimmen zum Einsatz kommen. Aus dieser Verantwortung folgt eine erhöhte Transparenzpflicht: Der Einsatz synthetischer Stimmen im Nachrichtenkontext erfordert klare Kennzeichnung, um Täuschung zu vermeiden und um die Zuschreibung von Verantwortlichkeiten zu sichern.

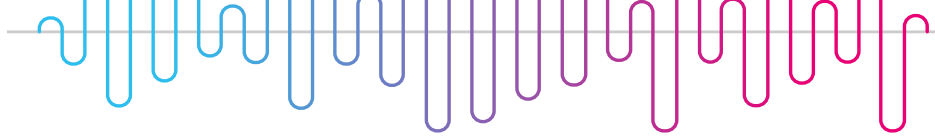
Im Entertainment-Bereich ist die Verantwortung in erster Linie auf urheber- und persönlichkeitsrechtliche Fragen sowie auf den Schutz vor irreführender Darstellung gerichtet, wenn reale Personen imitiert werden. Transparenz kann flexibler gehandhabt werden, solange keine dokumentarischen oder faktischen Ansprüche erhoben werden und alle beteiligten Personen vor Missbrauch geschützt werden.

Ethische Implikationen der Indexikalität

In der Medientheorie wird die Stimme im journalistischen Kontext als indexikalisches Zeichen verstanden: Sie verweist auf eine reale Person, die mit dem Gesagten in Verbindung steht und für dessen Richtigkeit einsteht. Wird diese Stimme synthetisch erzeugt, kann diese Bedeutung verlorengehen oder verzerrt werden, weil durch die Künstlichkeit der Verbindung diese zentrale Verknüpfung zwischen einer Botschaft und einem zur potenziellen Verantwortungsübernahme (moralisch, rechtlich, berufsethisch) verpflichteten menschlichen Sender verändert respektive aufgehoben wird. Noch einmal gewendet: Anders als in der medial vermittelten Beziehung zwischen tatsächlich existierendem, folglich menschlichem Sender und dessen Rezipierenden fehlt in der Maschine-Mensch-Relation synthetisch geprägter Konstellationen der Beleg für Faktizität, Glaubwürdigkeit und Wahrhaftigkeit des Sprechers (eben durch die Abwesenheit des menschlichen Absenders). Sie muss dann auf anderen Wegen aufgebaut werden, etwa durch die Offenlegung des KI-Einsatzes oder die Letztkontrolle eines KI-Erzeugnisses durch einen Menschen. Dies wirft Fragen der Authentizität und Verantwortungszuschreibung auf, die im Nachrichtenbereich ethisch relevant sind.

Im Entertainment ist die Stimme hingegen in der Regel ikonisch oder symbolisch: Sie steht für eine Figur, nicht für eine real existierende Quelle. Der Verlust an Indexikalität ist hier kein ethisches Problem, sondern ein gestalterischer Spielraum.





Risiko- und Chancenbilanz

Im Nachrichtenbereich bestehen aus ethischer Sicht unter anderem die folgenden Risiken: Vertrauensverlust, mögliche Täuschung, Entkopplung von Verantwortung und die Gefahr, dass publizistische Standards untergraben werden. Chancen bestehen vor allem in barrierefreien Zugängen, multilingualen Ausspielwegen und in klar abgegrenzten Service-Formaten, sofern diese transparent als synthetisch erstellt gekennzeichnet sind. Ein für die journalistische Ethik maßgeblicher Wert, der ganz grundsätzlich zur Disposition steht, wenn KI-Technologie im Journalismus eine dominante Verwendung findet, ist Unabhängigkeit. Um ihre Integrität abzusichern, dürfen sich journalistisch arbeitende Medienorganisationen – einer bekannten Maxime nach – mit keiner Sache gemein machen. Diese Prämisse lässt sich auf die technische Dimension erweitern. So besteht die Gefahr, dass Medienanbieter zunehmend abhängig werden von den Anbietern der KI-Modelle und von darauf basierender Tools, vornehmlich also von Technologie-Konzernen aus den USA, die schon die Plattformen und Sozialen Medien „beherrschen“.

Im Entertainment-Bereich weist GenKI also zahlreiche Potenziale auf: synthetische Stimmen können neue kreative Formate ermöglichen, Produktionskosten senken, Übersetzungen und Synchronisation deutlich erleichtern und individuelle Nutzererlebnisse ermöglichen. Aber auch hier gibt es ethische Risiken, die betrachtet werden müssen: der potenzielle Missbrauch realer Stimmen, unzureichende Rechteklärung, Jobverluste von Sprecherinnen und Sprechern sowie Fälle, in denen Fiktion als Fakt inszeniert werden könnte.

Im Folgenden werden nun jene Werte referiert, die für die öffentliche, mediale und journalistische Kommunikation maßgeblich sind. Aus Werten lassen sich Prinzipien ableiten bzw. letztere mit ersteren begründen. Prinzipien lassen sich wiederum in allgemeine Normen überführen. Werden diese nun auf einen spezifischen Kontext übertragen, sprechen wir von Leitlinien.

Nachrichtenjournalismus

Zentrale Werte:

- Wahrheit
- Vertrauen
- Transparenz
- Verantwortung
- Unabhängigkeit
- Neutralität
- Integrität



- 
- Vielfalt, Repräsentation und Gleichheit
 - Gerechtigkeit und Teilhabe

Ziele und dazugehörige Prinzipien:

- Wahrhaftigkeit und Objektivität
- Glaubwürdigkeit und Ehrlichkeit
- Irrtumsfreiheit und Faktizität
- Vertrauenswürdigkeit und Verantwortlichkeit
- Unparteilichkeit, Fairness und Meinungspluralismus
- Partizipation
- Zugangsgerechtigkeit, Inklusion und Barrierefreiheit
- Diskriminierungsfreiheit und Jugend- und Opferschutz
- Authentizität des journalistischen Absenders (Rolle, Funktionen etc.)
- Nachvollziehbarkeit, Offenlegung, Quellentransparenz und Rechenschaftspflichten
- Abwehrrecht bezüglich Privatsphäre und Quellenschutz

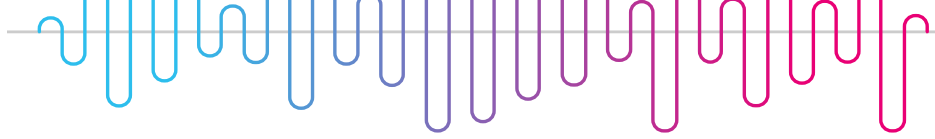
Mögliche Risiken bei Einsatz von synthetischen Stimmen:

- Vertrauens-, Reputations- und Legitimitätsverlust
- Täuschungspotenzial und Zweckentfremdung
- Entkoppelung von Verantwortung im Verhältnis Sender und Botschaft
- Abhängigkeit (ggü. Dritten wie KI-Anbieter) und Verlust von institutioneller Autonomie und Selbstbestimmung
- Verlust von Entscheidungsfreiheit in Bezug auf Verwendung und Verwertung von journalistischer Information und Daten
- Zugangs-, Zugriffs- und Kontextkontrolle (bzgl. Informationen und Daten)
- Verlust von Echtheit, Originalität, Faktizität und Authentizität

Mögliche Standards als Gegenmaßnahmen (Vorschläge, Auswahl):

- Kennzeichnung und Offenlegung eines KI-Einsatzes
- journalistische Ziel- und Zweckbindung der KI-Produkte
- Human-in-the-Loop, menschliche Endkontrolle und Letztverantwortung
- Zugangs-, Zugriffskontrolle und Kontextbindung bzgl. KI-relevanter journalistisch verwendeter Daten und Informationen
- Datensicherheit, Datensparsamkeit, Provenienz-Metadaten
- privatheitssichernde Selbstbestimmungsrechte bezüglich der Verwendung von Klonen





Entertainment-Bereich

Zentrale Werte:

- Unterhaltung und Zerstreuung
- Kreativität und Ästhetik
- Freiheit und Autonomie (der Kunst)
- Verantwortung (gegenüber Kunst und dessen Publikum)
- Respekt und Menschenwürde (gegenüber den dargestellten Personengruppen)
- Vielfalt, Repräsentation und Gleichheit
- Gerechtigkeit und Teilhabe
- Autonomie (des Publikums in seiner Entscheidungsfindung und Beurteilung)
- Empathie (mit dargestellten Personengruppen)

Ziele und dazugehörige Prinzipien (ergänzend):

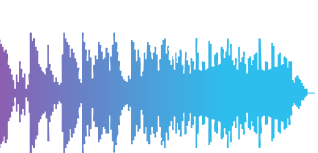
- Sicherstellung von Kunstfreiheit und von ästhetischer Qualität
- Urheberschaft und Rechtewahrung der Kuntschaffenden
- Wertschätzung von kreativer, kunst- und kulturschaffender Arbeit
- Diskriminierungsfreiheit, Jugend-, Opfer- und Minderheitenschutz (Ethnie, Geschlecht, sexuelle Orientierung etc.)
- Zugangsgerechtigkeit, Inklusion und Barrierefreiheit
- Gewaltfreiheit (Zumutbarkeit relativ zur jeweiligen Gruppe)
- Achtung der Gefühle und Perspektive anderer
- Manipulationsfreiheit (Täuschung nur zur Unterhaltung, nicht zur Persuasion)

Mögliche Risiken beim Einsatz von synthetischen Stimmen:

- Persönlichkeits-/Urheberrechtsverletzungen
- Einschränkung von menschlicher Kreativität
- Quantität vor Qualität und Verlust von Qualität in der Produktion
- Jobverluste und damit Verlust von menschlichem Wissen und Fähigkeiten

Schlussfolgerung

Aus medienethischer Sicht ist der entscheidende Unterschied, dass der Nachrichtenjournalismus Teil einer normativ gebundenen Infrastruktur in einer demokratischen Öffentlichkeit ist, während der Entertainment-Bereich in einer primär ästhetischen, fiktionalen und marktorientierten Logik operiert. Daraus folgt, dass für den Einsatz synthetischer Stimmen





im Nachrichtenjournalismus strengere Transparenz-, Sorgfalts- und Verantwortlichkeitsstandards gelten müssen als im Entertainment. Der ethische Maßstab richtet sich nicht allein nach der Technologie, sondern nach der Funktion des Mediums und den berechtigten Erwartungen der Öffentlichkeit.

Der Verband Deutscher Sprecher:innen (vds) und der Verband United Voice Artists (UVA) haben Statements veröffentlicht, in denen sie eine kritische Sicht äußern im Hinblick auf den Einsatz von GenKI zur Stimmsynthese und zum Klonen von Stimmen und die Auswirkungen auf die menschliche Kreativität. Der Druck, der auf Sprecherinnen und Sprechern lastet und deren Zukunftsängste sind real und sehr ernst zu nehmen.³ Dies trifft vor allem auf den Entertainmentbereich zu, denn Synchronsprecherinnen und Synchronsprecher könnten zukünftig durch KI ersetzt werden und hier ist eine öffentliche Debatte notwendig, um sich über die kreative Zukunft des Menschen im Zeitalter von KI zu verständigen. Gerade hier stellt sich die Frage nicht nach dem, was technisch möglich ist, sondern was aus gesellschaftlicher und menschlicher Sicht eigentlich wünschenswert ist.

Das Forschungsprojekt GEISST untersucht synthetische und geklonte Stimmen im Nachrichtenbereich und aus diesem Grund wird in den folgenden ethischen Handlungsempfehlungen auch nur dieser Bereich adressiert. Der Entertainment-Bereich muss in weiteren, zukünftigen Forschungsprojekten gesondert betrachtet werden.

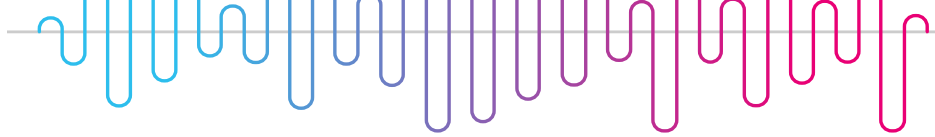
Ethische Leitlinien für den Einsatz von synthetischen und geklonten Stimmen im Nachrichtenjournalismus

Präambel

Der Einsatz von KI-generierten Stimmen im Journalismus birgt große Potenziale, etwa im Hinblick auf Barrierefreiheit, Effizienz durch Automatisierung, sich einer daraus ergebenden erhöhten Wirtschaftlichkeit und vielem mehr. Gleichzeitig stellen synthetische oder geklonte Stimmen eine tiefgreifende Herausforderung für das Mediensystem und die Institution des Journalismus dar, weil dadurch charakteristische professionsspezifische und demokratieerhaltende Güter bedroht werden können – namentlich Glaubwürdigkeit, Wahrhaftigkeit, Vertrauenswürdigkeit, Verantwortlichkeit, Integrität, Transparenz und weitere.

³ Online unter: https://www.unitedvoiceartists.com/wp-content/uploads/2023/05/230529_UVA_Manifesto_Deutsch_52923_TF.pdf und online unter: <https://www.sprecherverband.de/ki-statement/>.





Die primäre Aufgabe und Leistung von Journalistinnen und Journalisten es ist, eine freie und unabhängige Meinungsbildung der Bürgerinnen und Bürger zu gewährleisten und damit öffentliche Diskurse zu initiieren und Debatten abzubilden. Weiterhin fungiert der Journalismus als „Vierte Gewalt“ – dergestalt, dass er öffentliche und vor allem politische Vorgänge kontrolliert, Missstände aufdeckt, bedeutsame Sachverhalte und Zusammenhänge aufklärt, erklärt und interpretiert.

Unter Journalismus verstehen wir dabei vorwiegend professionelle Praktiken der objektiven, faktenbasierten sowie auf Pluralismus ausgerichteten Informationsvermittlung. Deren Richtwert stellt die gesellschaftliche Relevanz von Themen und Sachverhalten dar. Als zentraler Orientierungsmarker der Informationsvermittlung ist Wahrheit zu nennen. Als wesentliche Prinzipien der journalistischen Arbeit und der journalistischen Rolle sind Wahrhaftigkeit, Glaubwürdigkeit, Objektivität, Faktentreue und weitere Maximen oder Pflichten zu erwähnen. Die Legitimität dieser verfassungsrechtlich abgesicherten Rolle ergibt sich aus der Erfüllung der genannten funktionalen Leistungen zum Wohle des Erhalts der Stabilität einer Demokratie. Vertrauen in den und Vertrauenswürdigkeit in die Vertreterinnen und Vertreter des Journalismus sichern die exponierte Stellung des Journalismus als institutionalisierter Sprecher der Öffentlichkeit ab.

Journalismus unterscheidet sich von informeller, werblicher oder in ähnlichem Maße zweckmäßiger, meinungslastiger oder einseitiger Kommunikation insbesondere in seiner Neutralität, Unabhängigkeit und Faktizität. Journalistinnen und Journalisten sind idealtypisch frei von Parteilichkeit und einzig gegenüber der Öffentlichkeit verantwortlich und/oder dieser verpflichtet. Die Wahrung des Medienvertrauens ist damit fundamental für die Legitimierung der öffentlichen Rolle und basiert schließlich auf Verantwortung vor der Profession und der Gesellschaft.

Die im Folgenden referierten Leitlinien verfolgen einen teleologisch-verantwortungsethischen Ansatz, welcher sich an Max Webers Verantwortungsethik orientiert. Sie fordern eine an Ziel und Kontext gebundene Risikobewertung des KI-Einsatzes in den Medien und betrachten Journalismus als einen spezifischen Modus von öffentlicher Kommunikation sowie als eine Form von politischer Praxis mit gesellschaftlichem Auftrag. Ferner verstehen sie sich als Konkretisierung bestehender professionsethischer Kodizes wie allen voran dem Pressekodex des Deutschen Presserates.⁴ Auch bewegen sich die Leitlinien im Rahmen der europäischen Grundrechteordnung und der KI-Verordnung der Europäischen Union.⁵

4 Mehr dazu: https://www.presserat.de/pressekodex.html?file=files/presserat/dokumente/pressekodex/Presse-kodex_Leitsaetze_RL12.1.pdf

5 Hier nachzulesen: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A32024R1689>





Zum Begriff der teleologischen Verantwortungsethik

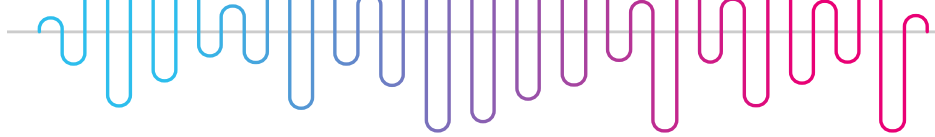
Die vorliegenden Leitlinien verwenden bewusst den Begriff einer teleologischen Verantwortungsethik, um zwei ethische Traditionslinien miteinander zu verbinden:

Zum einen umfasst dies die von Aristoteles geprägte Idee eines auf das „gute Leben“ ausgerichteten Handelns, das in konkreten Situationen Maß, Mitte und Zielorientierung reflektiert. Der Begriff des „telos“ wird demnach ausdrücklich in einem aristotelischen Sinn verwendet und bezieht sich auf die Ziele eines „guten“ Journalismus, der seinen normativen Aufgaben in erschöpfendem Maße nachkommen kann. Zum anderen stützen sich die Leitlinien auf die von Max Weber in „Politik als Beruf“ formulierte Verantwortungsethik, die das Abwägen von Handlungsfolgen und das Übernehmen von Verantwortung für reale Auswirkungen verlangt, weil der Journalismus durch seinen exklusiven Zugriff auf die massenmediale Vermittlungsinfrastruktur und seinen Zugang zu politischen Entscheidern einerseits erhebliche Machtpotenziale auf sich akkumuliert. Als öffentlicher Kommunikator oder Kommentator übt er andererseits selbst Macht aus oder kann die Macht anderer begrenzen, vornehmlich jene des politischen Systems. So können journalistische Recherchen und Publikationen handlungsleitend sein für Exekutive, Legislative oder Judikative. Ebendeshalb erkennt Weber im Journalismus explizit auch eine politische Instanz. Macht geht nun zwangsläufig mit Verantwortung einher. Weber fordert daher von journalistischen Akteuren auch, ihre kommunikativen Aktivitäten stets auf mögliche Folgen zu beziehen oder die Konsequenzen des eigenen Handelns zumindest abzuwägen, dies im Sinne der eigenen Aufgabe und des Gemeinwohls. Das gilt auch für den Technikeinsatz oder die Verwendung von KI.

Auch wenn Weber seine Verantwortungsethik von der Gesinnungsethik abgrenzt, betont er letztlich, dass beide in einem reifen ethischen Bewusstsein zusammengehören. In diesem Sinn versteht sich die hier formulierte teleologische Verantwortungsethik als ein Gestaltungsansatz, wie ihn prominent etwa Martha Nussbaum mit ihrem „Fähigkeitenansatz“ vertritt, und der nicht nur auf Folgen achtet, sondern auch auf die sittliche Haltung und Zielorientierung des Handelns – hier beispielweise im Kontext von digitalen respektive selbstlernenden Technologien.

Diese Haltung ist eng verwandt mit Ansätzen wie Value Sensitive Design oder Ethics by Design, die eine kontextbezogene, wertorientierte Technikgestaltung fordern. Die vorliegenden Leitlinien möchten daher nicht normativ „von außen“ bewerten, sondern Räume für verantwortbare, kontextgerechte Entscheidungen innerhalb redaktioneller und medialer Praxis schaffen. Sie sind Empfehlungen, die einen allgemeinen Handlungskorridor benennen und den entscheidenden Akteuren in der Praxis bewusst Handlungsspielräume ermöglichen, aber keine Richtlinien, deren Umsetzung verpflichtend wären.





Der hier verwendete Begriff der teleologischen Verantwortungsethik ist nicht mit einem einfachen Nützlichkeitsdenken gleichzusetzen. Vielmehr knüpft er an die aristotelische Tradition an, verstanden als Orientierung auf ein Gutes, das nicht bloß individuell, sondern auch gesellschaftlich begründet ist. Im Kontext journalistischer Praxis bedeutet dies: Der Einsatz synthetischer Stimmen ist nicht an sich gut oder schlecht. Er wird vielmehr daran gemessen, ob er der journalistischen Funktion (zum Beispiel sachliche Information, Orientierung oder demokratische Aufklärung) eines unabhängigen, zugänglichen und demokratisch verantworteten Journalismus dient. In einer medienökonomisch herausgeforderten Landschaft – etwa im Lokaljournalismus – könnte der gezielte, offen gelegte Einsatz von KI-Stimmen ein Mittel sein, um die journalistische Grundversorgung der Region sowie die Qualität der regionalen Medien zu sichern. In gebührenfinanzierten öffentlich-rechtlichen Kontexten hingegen, deren Legitimation auf Gemeinwohlorientierung und Vertrauen beruht, wäre der Einsatz von KI-Stimmen anders zu beurteilen, etwa wenn er primär auf eine Kostenreduktion oder auf eine Personalisierung von Angeboten zielt. Hier müsste geprüft werden, ob der Auftrag des öffentlich-rechtlichen Rundfunks dann womöglich noch ausreichend erfüllt wird.

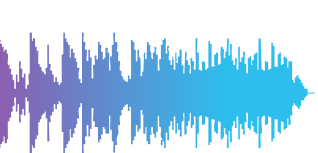
Die in diesen Leitlinien vorgeschlagene Form der Zielbindung ist also nicht rein technik- oder nutzenzentriert, sondern normativ eingebettet: Sie fragt, welcher Beitrag zur öffentlichen Orientierung durch den KI-Einsatz geleistet wird und ob dieser Beitrag mit dem demokratischen Auftrag des Journalismus vereinbar ist. Die hier formulierte Verantwortungsethik bezieht sich nicht nur auf klassische journalistische Institutionen, sondern auch auf Akteure, die öffentlich meinungsbildend wirken, beispielweise politische Influencerinnen und Influencer oder reichweitenstarke Plattformakteure. Auch hier wird die Meinungsbildung beeinflusst und Vertrauen beansprucht und dadurch entsteht eine ethische Verantwortung.

1. Zielbindung und Kontextualität

- Der Einsatz synthetischer oder geklonter Stimmen muss an eine klar definierte journalistische Funktion gebunden sein. Eine Generalanwendung ist ausgeschlossen.
- Der Einsatz ist nur dann ethisch vertretbar, wenn er der Informationsvermittlung, Aufklärung oder der Zugänglichkeit für spezifische Zielgruppen dient und somit durch nachvollziehbare Ziele verständlich wird.
- Jeder Einsatz muss im jeweiligen Kontext geprüft und dokumentiert werden.
- Die Abwägung darf sich nicht allein auf Effizienz- oder Ökonomiekriterien stützen, sondern muss der journalistischen Auftragserfüllung dienen.

2. Transparenz und Kennzeichnungspflicht

- Synthetische und geklonte Stimmen sollten eindeutig und unmissverständlich als solche gekennzeichnet werden.
- Eine Kennzeichnungspflicht besteht insbesondere bei gesprochenen Beiträgen, die in





Form und Intonation menschliche Sprecherinnen und Sprecher imitieren. Eine solche Offenlegung des KI-Einsatzes kann in Medienformaten, die auf Hör- und Sprechbeiträgen basieren (Hörfunk, Podcast etc.), auch dadurch erfolgen, dass eine KI-Stimme sich selbst offenbart; ein KI-Klon also selbst „sagt“, dass sie eine KI ist.

- Die Herkunft und technische Erzeugung der Stimme sollten dokumentierbar und offenlegbar sein.

3. Schutz vor Täuschung und Manipulation

- Die Verwendung von KI-Stimmen soll nicht der emotionalen Beeinflussung oder politischen Instrumentalisierung dienen.
- Nachrichtenbeiträge sollten keine Stimmprofile verwenden, die gezielt Vertrauen suggerieren (zum Beispiel durch bekannte oder besonders sympathische Stimmen), ohne dass dies offengelegt wird.
- Besonders sensible Inhalte (beispielsweise Krisen, Katastrophen, politische Konflikte) sollten nicht primär durch synthetische Stimmen vermittelt werden, da dies das mediale Verantwortungsverhältnis abschwächen könnte.

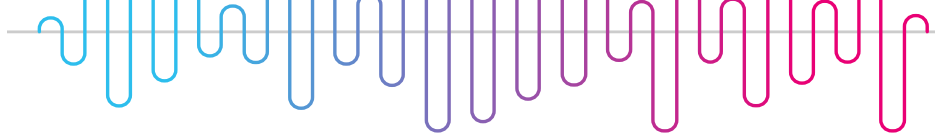
4. Verantwortung und Rechenschaftspflicht

- Die redaktionelle Verantwortung bleibt bei den journalistischen Akteuren. Der Einsatz von KI entbindet nicht von der ethischen und rechtlichen Verantwortung für die Inhalte.
- Es sollten nachvollziehbare Entscheidungsprozesse dokumentiert werden, die den Einsatz von KI-Stimmen legitimieren.
- Medienorganisationen sollten Beschwerdemechanismen für Rezipientinnen und Rezipienten bereitstellen, die sich durch KI-Stimmen getäuscht oder irritiert fühlen.

5. Schutz menschlicher Sprecherinnen und Sprecher und Rechte am Stimmprofil

- Geklonte Stimmen dürfen nur mit ausdrücklicher, informierter und widerrufbarer Zustimmung der betroffenen Person eingesetzt werden.
- Stimmprofile dürfen nicht über den ursprünglich vereinbarten Kontext hinaus verwendet oder modifiziert werden.
- Die Rechte an einer Stimme sind schützenswert und als persönlichkeitsrechtlich relevant zu behandeln.
- Es ist aus einer ethischen Perspektive unzulässig, synthetische Stimmen zu erzeugen oder einzusetzen, die gezielt einem realen Stimmprofil stark ähneln, ohne dass eine ausdrückliche Zustimmung der betroffenen Person vorliegt – auch wenn die Stimme technisch neu generiert wurde. Die absichtsvolle Nachbildung bekannter Stimmen untergräbt Persönlichkeitsrechte und kann das Vertrauensverhältnis zwischen Medium und Publikum verletzen.





6. Ethische Gestaltung affektiver Sprechmodulation

- Die affektive Gestaltung (zum Beispiel Tonfall, Tempo, Stimmfarbe) muss dem journalistischen Auftrag angemessen sein und darf nicht gezielt auf emotionale Steuerung ausgerichtet werden.
- Emotionalisierte synthetische Stimmen dürfen nicht zur Vertrauensgenerierung missbraucht werden.
- Eine sachlich-informative Tonalität ist im Nachrichtenbereich vorrangig.

7. Reflexion, Schulung und Weiterentwicklung

- Journalistinnen und Journalisten sollten in der Lage sein, die Wirkungen synthetischer Stimmen zu reflektieren und in ihren redaktionellen Entscheidungen zu berücksichtigen.
- Medienhäuser sollten Schulungen und Leitfäden zum ethischen Einsatz von KI-Stimmen bereitstellen.
- Medienhäuser sollten sich aktiv an der Erforschung der affektiven und kognitiven Wirkung synthetischer Stimmen beteiligen – insbesondere im Hinblick auf Vertrauen, emotionale Wahrnehmung und Urteilsbildung. Die Ergebnisse sollten systematisch in die redaktionelle Praxis und Gestaltung einfließen.
- Leitlinien sollten regelmäßig überprüft und weiterentwickelt werden, insbesondere im Hinblick auf neue technische Entwicklungen.

8. Medienkompetenz

- Medienkompetenz wird als Fähigkeit verstanden, Informationen nicht nur zu konsumieren, sondern kritisch einzuordnen und zu bewerten. Dafür relevante Fähigkeiten, Fertigkeiten und Wissensbestände sollten stets gefördert und dergestalt sichergestellt werden, dass journalistische Akteure durch den Einsatz von KI nicht an journalistischen Basiskompetenzen einbüßen („Deskilling“).
- Eine transparente und offene Kennzeichnung synthetischer Stimmen kann diese Kompetenz auch auf Seiten von Nutzerinnen und Nutzern stärken, weil dadurch Lern- respektive Qualifizierungseffekte hinsichtlich des Erkennens von KI-Produkten eingeleitet werden könnten.
- Allgemein gesprochen steht der Journalismus damit auch in der Verantwortung, die Rezipientinnen und Rezipienten in die Lage zu versetzen, informierte Entscheidungen zum Umgang mit KI-Stimmen zu treffen. Diese Form der indirekten Befähigung des Publikums zur mündigen Reflexion von KI kann letzteres, als Nutzerinnen und Nutzerin, selbst zu einem souveränen und reflektierten KI-Einsatz befähigen.
- Medienhäuser könnten beispielsweise Modelle oder Darstellungsformate prüfen, bei denen Nutzerinnen und Nutzer grundsätzlich wählen können, ob sie eine KI-generierte Stimme oder eine menschliche Stimme bevorzugen, um Autonomie, transparenzbasiertes Vertrauen und einen reflektierten Medienkonsum zu fördern.
- Die kontinuierliche Prüfung der Wirkung von KI-Stimmen erfordert darüber hinaus mehr als statistische Auswertung oder Meinungsumfragen: Sie braucht Räume für akti-





ven Austausch, in denen Rezipientinnen und Rezipienten ihre Erfahrungen, Irritationen und Erwartungen artikulieren können – als Teil einer demokratischen Medienkultur.

- Medienhäuser sollten deshalb neue Dialog- und Austauschformate entwickeln, um qualitative Rückmeldungen der Nutzerinnen und Nutzer zu ermöglichen – etwa über moderierte Feedbackrunden, Ethikdialoge oder partizipative Foren zur Wirkung von KI-Stimmen.

Erweiterter ethischer Bezugsrahmen im Hinblick auf Medienkompetenz und die Verantwortung von Nutzerinnen und Nutzern

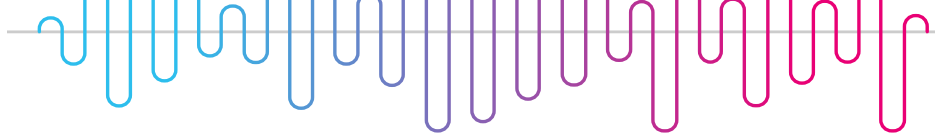
Die hier vertretene ethische Perspektive steht in gewisser Hinsicht in inhaltlicher Nähe zum „Fähigkeitenansatz“ (im Original: „Capability Approach“) von Martha Nussbaum, der ebenso auf aristotelischen Fundamenten beruht. Auch dort steht das Ziel im Zentrum, Menschen Bedingungen für ein gelingendes Leben zu ermöglichen – in Freiheit, Teilhabe und Würde. Der Journalismus trägt zu diesen Bedingungen bei, indem er Informationen bereitstellt, Orientierung ermöglicht und Öffentlichkeit herstellt. In diesem Sinne ist die Funktion des Journalismus nicht primär pädagogisch, sondern ermöglichend: Er schafft Voraussetzungen für individuelle und gesellschaftliche Urteilskraft.

Medienkompetenz wird in dieser Perspektive nicht zum Selbstzweck erhoben, sondern verstanden als eine Fähigkeit, die durch die Qualität und Transparenz journalistischer Angebote gestärkt werden kann. Eine verantwortungsethische Beurteilung des KI-Einsatzes trägt dazu bei, diese Fähigkeit nicht zu untergraben, sondern zu fördern – gerade durch Offenheit, Klarheit und ethische Zielbindung im Medieneinsatz.

Dabei kommt auch den Rezipientinnen und Rezipienten eine aktive Rolle zu. Transparenz schafft Vertrauen – aber nur, wenn die Nutzenden befähigt werden, informierte Entscheidungen zu treffen. Deshalb sollten Medienhäuser auch Möglichkeiten der Mitbestimmung und Wahlfreiheit prüfen, beispielsweise in Form von bewussten Wahlmöglichkeiten für Nutzende über Opt-in-Funktionen für synthetische Stimmen oder alternative Darstellungsformate, die es den Anwenderinnen und Anwendern erlauben, sich selbstbewusst für KI-Stimmen oder menschliche Stimmen zu entscheiden. Medienkompetenz bedeutet in diesem Verständnis nicht nur Informationen zu konsumieren, sondern aktiv einordnen und bewerten zu können.

Die Wahrnehmung und möglichen (Aus-)Wirkungen von synthetischen Stimmen werden bislang wissenschaftlich nur unzureichend verstanden. Gerade ihre affektive Wirkung – im Hinblick auf Vertrauen, emotionale Nähe oder Irritation – kann stark variieren je nach Altersgruppe, Kontext und individueller Medienerfahrung. Eine verantwortungsvolle Nutzung erfordert daher nicht nur statistische Nutzerakzeptanz, sondern qualitative Forschung zur





Rezeption und Wirkung. Medienhäuser sollten sich aktiv an der Erforschung dieser Effekte beteiligen und ihre Praxis auf neue Erkenntnisse abstimmen.

Die ästhetische Gestaltung synthetischer Stimmen sollte weiterhin nicht gezielt auf ein Idealbild menschlicher Kommunikation ausgerichtet sein, das möglicherweise ein übersteigertes Vertrauen erzeugen könnte. Stimmen, die zu perfekt wirken, könnten eventuell die Wahrnehmung von Echtheit verzerren und das Vertrauen in reale menschliche Sprecherinnen und Sprecher in Zukunft untergraben. Diese Überlegungen sind zum gegenwärtigen Zeitpunkt spekulativ und es gibt keine ausreichenden wissenschaftlichen Untersuchungen und Beweise. Aus diesem Grund müssen die Entwicklungen sorgfältig wissenschaftlich begleitet werden, um mögliche gesellschaftliche Risiken und Auswirkungen frühzeitig zu erkennen.

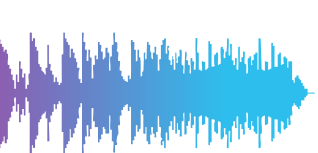
Eine verantwortungsethische Beurteilung des KI-Einsatzes trägt dazu bei, Medienkompetenz nicht zu untergraben, sondern zu fördern: durch Offenheit, Klarheit und ethische Zielbindung im Medieneinsatz. Die Perspektive der teleologischen Verantwortungsethik stellt sich dabei nicht in Gegensatz zu rechtlichen Vorgaben, sondern ergänzt sie: Sie fragt danach, wie die normativen Grundlagen der Demokratie konkret geschützt und gefördert werden können.

Diese Perspektive ist besonders relevant angesichts der Unsicherheiten, die mit dem Einsatz von KI-Systemen einhergehen. Viele ihrer Auswirkungen auf Wahrnehmung, Vertrauen, Urteilsbildung, menschliche Kreativität, Arbeitsmarkt und Arbeitsformen oder gesellschaftliche Diskurse sind noch nicht absehbar. Deshalb versteht sich die hier vorgeschlagene Verantwortungsethik als eine „wachsame Ethik.“ Sie fordert zu einer kontinuierlichen Prüfung der Zwecke und Ziele auf, denen der KI-Einsatz dient, dies erfordert die Überprüfung von deren Wirkung in der Praxis und kann gegebenenfalls ein rasches Umdenken und Anpassen des Handelns an die erkennbaren Folgen abverlangen.

Schlussbemerkung

Diese Leitlinien verstehen sich als ethisches Fundament für eine differenzierte, verantwortungsvolle Nutzung von KI-Stimmen im Nachrichtenjournalismus. Sie bewerten synthetische Stimmen nicht als Bedrohung, sondern als Gestaltungselement: Leitlinien können journalistische Verlässlichkeit stärken, Teilhabe ermöglichen und Freiräume für echte menschliche Kommunikation schaffen, sofern sie verantwortungsvoll eingesetzt werden. Diese Leitlinien stellen keine Verbote auf, sondern rahmen die Bedingungen, unter denen ein ethisch legitimer Einsatz möglich ist.

Ihre Grundlage ist eine teleologische Verantwortungsethik, die dem journalistischen Handeln Orientierung in einer digitalen Medienlandschaft geben will, die sich immer schneller transformiert. Sie verstehen sich zugleich als Einladung zur aktiven Gestaltung einer Medienethik,





die nicht nur reagiert, sondern Verantwortung in die Zukunft hineintragen möchte: Indem sie Verantwortung konkret einfordert von Redaktionen, durch transparente Entscheidungen und Schulungen; von Anbietern, durch Kennzeichnung und Offenlegung; und von Nutzerinnen und Nutzern, die ihre Urteilkraft sorgfältig ausbilden und ihre Wahlfreiheit selbstbestimmt anwenden.

Dabei gehen diese Leitlinien über bestehende Rahmenwerke hinaus, indem sie auch die affektive Wirkung synthetischer Stimmen in den Blick nehmen möchten und die noch unzureichend erforschte Rezeption nicht als Nebenaspekt, sondern als ethisch relevante Größe anerkennen. Zum Schluss soll noch dieser Punkt hervorgehoben werden: Die mediale Macht synthetischer Stimmen entsteht nicht nur durch Inhalte, sondern ebenso durch Wiederholung, Ästhetik und algorithmische Auswahl. Die Macht der Gewöhnung sollte nicht unterschätzt werden. Eine ethische Medienpraxis muss in Zukunft auch diese neuen Formen symbolischer Autorität reflektieren und regulieren können.

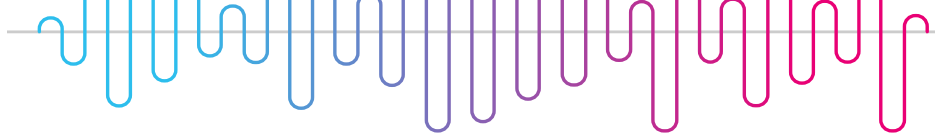
Zusammenfassend lässt sich sagen: Die hier vertretene teleologische Verantwortungsethik fordert eine Rechenschaftspflicht über den Einsatz von synthetischen oder geklonten Stimmen im Journalismus. Wer mithilfe von Künstlicher Intelligenz generierte Stimmen einsetzt, muss darlegen können, zu welchem Ziel dieser Einsatz dient und warum keine menschlichen Stimmen verwendet werden. Die bloße Möglichkeit oder Effizienz ohne Zweck und somit ein Einsatz ohne eine vernünftige Begründung sind aus einer ethischen Sicht im Journalismus zu vermeiden. Eine Begründung für den Einsatz von KI-Stimmen ist notwendig, um die Voraussetzung für Verständnis und Akzeptanz zu schaffen. Der Einsatz von Künstlicher Intelligenz darf nicht zu einer Entwertung der journalistischen Verantwortung oder zur beliebigen Steigerung von Content führen. (Medien-)Vertrauen entsteht nicht durch Masse, sondern durch Sinn, Maß, Haltung und Verständnis.

Weitere Erläuterungen zu den ethischen Leitlinien im Projekt GEISST

Begründung des ethischen Theoriebezugs

Die ethischen Leitlinien, die dem vorliegenden Gutachten zugrundeliegen, sind systematisch innerhalb eines ethischen Theoriebezugs verankert, der sowohl die Begründung als auch die Anwendung der Leitlinien transparent und nachvollziehbar macht. Diese bewusste Entscheidung unterscheidet unsere Herangehensweise von Leitlinien-Dokumenten, die auf eine theoretische Rahmung verzichten und sich auf pragmatisch formulierte Handlungsempfehlungen konzentrieren.





Wir verweisen explizit auf Aristoteles, Max Weber und Martha Nussbaum, weil sie drei sich ergänzende Dimensionen verantwortlichen Handelns aufzeigen:

- **Aristoteles:** Teleologische Ethik und Tugendorientierung

Aristoteles bietet mit seiner Ethik und der Idee des telos (Ziel/Zweck) eine Orientierung an, die nicht nur auf starre Regeln, sondern auf eine situationsbezogene Urteilskraft und das Streben nach einem guten Leben ausgerichtet ist. Es geht darum, das richtige Maß und die richtige Mitte zu finden, um die Fähigkeit und Notwendigkeit zur ausgewogene Entscheidungen, die sich an Situationen und Kontexten orientieren. In einem dynamischen Feld wie dem KI-gestützten Journalismus bietet dieser Ansatz ein tragfähiges Konzept, um technisches Handeln an einem übergeordneten Ziel – der Wahrheitsvermittlung und der informierten Öffentlichkeit – auszurichten. Gleichzeitig erlaubt es, nicht nur zu fragen, was erlaubt ist, sondern, was gut für den Menschen und die Demokratie ist. Und schließlich lässt es einen kontextabhängigen Handlungsspielraum zu, weil es darum geht, das richtige Maß zu finden.

- **Max Weber:** Verantwortungsethik als Haltung im politischen Kontext

Webers Begriff der Verantwortungsethik fordert dazu auf, die Folgen des eigenen Handelns genau abzuwägen und bereit zu sein, die Verantwortung auch für unbeabsichtigte Nebenwirkungen zu übernehmen. Im politischen Kontext, und Weber setzt den Journalismus, als beobachtende und darauf rekurrierende Instanz, in Relation zum politischen System, ist dieses ethische Handeln als eine Haltung zu verstehen, entlang derer Handelnde bereit sein sollen und wollen, Antworten zu geben und sich ihrer verantwortlichen Rechenschaftspflicht bewusst zu sein. Dieser Ansatz ist besonders relevant im Kontext automatisierter Systeme, in denen die Trennung zwischen menschlicher Absicht und technischer Wirkung zunehmend verschwimmt.

- **Martha Nussbaum:** Fähigkeitenansatz als Gestaltungsgrundlage

Der Capabilities Approach von Nussbaum, der auf einer Interpretation der aristotelischen Ethik basiert, erweitert die Diskussion um eine normative Orientierung, im Rahmen derer die Handlungsfähigkeit, Teilhabe und Würde der Medienkonsumentinnen und Medienkonsumenten von Bedeutung sind. Nussbaum erinnert uns daran, dass eine ethische Gestaltung von Technik nicht nur abstrakten Standards folgen sollte, sondern sich immer auch an den konkreten Bedürfnissen und Schutzinteressen realer Menschen orientieren muss. Dies gilt insbesondere in Bezug auf die Werte des Journalismus, die zur Befähigung und Teilhabe der einzelnen Personen am politischen und gesellschaftlichen und somit demokratischen Diskurs beitragen sollen.





Die Vorteile eines expliziten Theoriebezugs:

- Er macht die normative Grundlage der Leitlinien explizit und unterscheidet sie von üblichen Compliance-Vorgaben.
- Er schafft eine reflektierte Kohärenz zwischen Einzelforderungen und übergeordneten Zielen.
- Er fördert Verantwortungsbewusstsein nicht nur auf der Handlungsebene, sondern auch auf der Haltungsebene.
- Er erlaubt eine argumentative Weiterentwicklung der Leitlinien bei zukünftigen technischen oder gesellschaftlichen Veränderungen.
- Und schließlich stärkt er die kulturelle Anschlussfähigkeit der Leitlinien: beispielsweise im öffentlichen Diskurs, in Bildungsprozessen oder im interdisziplinären Dialog zwischen Technik, Ethik und Gesellschaft.

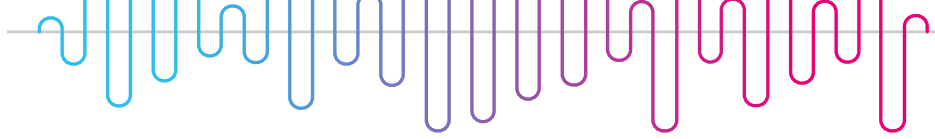
Medienethik und Medienkompetenz

Vor dem Hintergrund des Einsatzes von GenKI und von synthetischen und geklonten Stimmen im Journalismus wird eine Weiterentwicklung von Konzepten für eine tragfähige Medienkompetenz notwendig. Es genügt nicht mehr, lediglich grundlegende Recherchefähigkeiten zu vermitteln, wie das Auffinden und Überprüfen von Quellen. Vielmehr muss in Bildungskontexten verstärkt die Fähigkeit geschult werden, die Aufmerksamkeit bewusst zu lenken und die jeweilige kommunikative Situation richtig einzuordnen. Dazu gehören ein geschärftes Bewusstsein für die Bedingungen der Medienproduktion, ein Verständnis für die Funktionsweisen synthetischer Inhalte sowie die Kompetenz, Informationen und Praktiken unter Berücksichtigung von Kontextbedingungen kritisch zu deuten. Nur durch diese umfassenden Fähigkeiten lässt sich die wachsende Verunsicherung durch nicht mehr eindeutig unterscheidbare Stimmen im Nachrichtenbereich abfedern und die Urteilsfähigkeit der Rezipientinnen und Rezipienten stärken.

Die folgenden Fähigkeiten sollten dafür geschult und trainiert werden:

- 1. Aufmerksamkeitskompetenz:** die Fähigkeit, bewusst auf Nuancen, Kontexte und potenzielle Brüche in der medialen Darstellung zu achten, um Unsicherheiten früh wahrnehmen zu können.
- 2. Kontextualisierungskompetenz:** die Fähigkeit, mediale Inhalte nicht isoliert, sondern in ihren Produktions- und Verbreitungskontexten einzuordnen (z. B. wer spricht, in welchem Medium, mit welcher Intention).
- 3. Recherchefähigkeiten:** die Fähigkeit Inhalte und Quellen selbstständig zu überprüfen.
- 4. Technologisches Grundverständnis:** ein Basiswissen über Funktionsweise, Möglichkeiten und Grenzen synthetischer Stimmsysteme, das vor unreflektierter Akzeptanz schützt.





5. Urteils- und Reflexionskompetenz: die Fähigkeit, Unsicherheit auszuhalten, kritisch zu reflektieren und eine begründete Einschätzung vorzunehmen, auch wenn eine eindeutige Unterscheidung nicht möglich ist.

Medienkompetenz muss in Zeiten von synthetischen Stimmen über ein reines Fakten-Checken hinausgehen. Sie erfordert ein Zusammenspiel von technologischem Wissen, kritischer Aufmerksamkeit und reflektierter Urteilskraft.

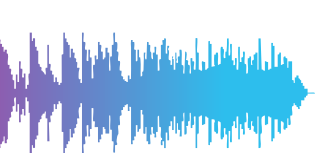
Eine zusätzliche Schwierigkeit im Hinblick auf Medienkompetenz und das veränderte Nutzungsverhalten liegt in der stetig sinkenden Aufmerksamkeitsspanne beziehungsweise in der stetig sinkenden Bereitschaft zur Aufmerksamkeit von Seiten der Rezipientinnen und Rezipienten. Der Medienkonsum verschiebt sich zunehmend hin zu kurzen, schnell konsumierbaren Social-Media-Inhalten, die kaum Raum für eine vertiefte Rezeption oder eine kritische Einordnung lassen. Damit verstärkt sich ein Widerspruch: Gerade im KI-Zeitalter wird eine intensive und kontextbewusste Auseinandersetzung mit Nachrichteninhalten erforderlich, um synthetische von menschlichen Stimmen und Darstellungsformen unterscheiden zu können. Der vorherrschende Trend zur Oberflächenwahrnehmung wirkt dieser Notwendigkeit jedoch entgegen und verschärft dadurch die bestehende Verunsicherung. Für diese Herausforderung liegen bislang weder im Journalismus noch in der Medienpädagogik ausreichende Lösungsansätze vor.

Rechenschaftspflicht im Rahmen einer teleologischen Verantwortungsethik

Rechenschaftspflicht bezeichnet im Sinne einer teleologischen Verantwortungsethik die moralische Verpflichtung von Akteuren, insbesondere journalistischen oder technologischen Entscheidungsträgern, für die Zwecke, Mittel und Folgen ihres Handelns öffentlich sichtbar und erklärbar zu sein. Sie umfasst dabei nicht nur die rückblickende Reaktion auf bereits eingetretene Wirkungen, sondern auch die vorausschauende Berücksichtigung möglicher Konsequenzen innerhalb eines normativ gerahmten Zielsystems.

Im aristotelischen Sinne ist Rechenschaftspflicht eng mit der Idee des guten Handelns im Hinblick auf das telos verbunden, also auf das Ziel einer tugendhaften Praxis, die das Gemeinwohl im Blick behält. Eine Entscheidung gilt dann als ethisch verantwortlich, wenn sie in Übereinstimmung mit einem sittlich guten Ziel (in diesem Fall qualitativ hochwertigem Journalismus und einer informierten Öffentlichkeit) getroffen wurde und die Mittel diesem Ziel angemessen sind. Die Rechenschaftspflicht ergibt sich aus der Zumutung an das Subjekt, sein Handeln begründbar, kohärent und in einem öffentlichen Diskurs zu gestalten.

In der Perspektive Max Webers Verantwortungsethik ergänzt sich diese Haltung um den Gedanken der realistischen Folgenabschätzung: Rechenschaft bedeutet hier, sich nicht hinter





guten Absichten oder einer Gesinnung, Systemlogiken oder einer vermeintlichen Technikneutralität zu verstecken, sondern die antizipierbaren Wirkungen der eigenen Entscheidungen mitzutragen, auch wenn diese nicht vollständig intendiert waren. Wer handelt, trägt in diesem Sinne eine Verantwortung für die absehbare Wirkung und nicht nur für eine gute Absicht.

In den ethischen Leitlinien zu KI-Stimmen im Nachrichtenjournalismus konkretisiert sich Rechenschaftspflicht daher in folgenden Aspekten:

- die Verantwortung für Sinnverschiebungen infolge automatisierter Textkürzungen,
- die Verantwortung für die Wirkung von KI-Stimmen auf das Vertrauen der Rezipientinnen und Rezipienten,
- die Verantwortung für die Gestaltung von Schnitt, Ton und Text im Hinblick auf das Ziel, wahrhaftige, verständliche und integre Nachrichten zu vermitteln,
- die Verantwortung für Transparenz und eine adäquate Kennzeichnung, um die Zuschreibung von Urheberschaft und redaktioneller Kontrolle öffentlich nachvollziehbar zu machen,
- und die Verantwortung für die Menschen: die Sprecherinnen und Sprecher, die ein Recht auf ihre eigene Stimme haben und in Zukunft nicht einfach durch KI ersetzt werden sollten.

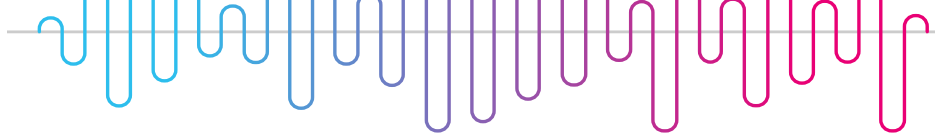
Rechenschaftspflicht bedeutet damit nicht nur Reaktion – sondern Reflexion. Sie ist eine Haltung, die ethische Orientierung nicht an formalen Zuständigkeiten festmacht, sondern an der Bereitschaft, für das eigene Tun erkennbar, ansprechbar und begründungsfähig zu bleiben, auch dann, wenn technische Systeme beteiligt sind.

Zentraler Use Case im Projekt GEISST und dessen ethische Beurteilung anhand der Leitlinien

Erstellung eines Profils für eine professionelle Moderationsstimme im Nachrichtenbereich

In primären Anwendungsfall, der im Rahmen des Forschungsprojektes GEISST untersucht wurde, geht es um die Veränderung eines Originalbeitrags, der für ein bestimmtes Nachrichtenformat produziert wurde und dann für weitere Medien und Kanäle automatisiert abgeändert werden soll, ohne dass die Sprecherin oder der Sprecher erneut im Studio anwesend sein und den Beitrag neu vertonen muss. So können zusätzliche Versionen des Nachrichtenbeitrags beispielsweise für regionale Varianten, für das Internet oder gekürzte





Social-Media-Versionen mit einer synthetischen oder geklonten Stimme produziert werden. Insgesamt soll die gesamte Produktion von verschiedenen Varianten eines Originalbeitrags in Zukunft automatisiert ablaufen können.

Ethische Relevanz des Use-Cases

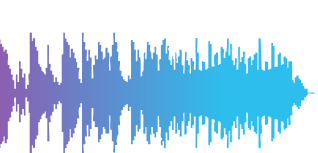
Die automatisierte Aufbereitung journalistischer Inhalte für unterschiedliche Ausspielkanäle – insbesondere für Soziale Medien – markiert einen bedeutsamen Paradigmenwechsel in der journalistischen Produktion. Während bislang menschliche Redakteurinnen und Redakteure den Prozess der Kürzung, Umformulierung, Bildauswahl und Vertonung verantworteten, verlagert der hier beschriebene Use-Case wesentliche Teile dieser Arbeit an ein KI-System. Dieses System übernimmt nicht nur die Textanpassung und den Bildschnitt, sondern erzeugt auch eine neue Tonspur mit einer KI-generierten Stimme. Die damit einhergehende Automatisierung betrifft zentrale journalistische Arbeitsprozesse und berührt grundlegende ethische Fragen

Die ethische Relevanz ergibt sich aus mehreren Kernaspekten:

- **Wahrhaftigkeit und Kontexttreue:** Die Auswahl und Umformulierung von Inhalten beeinflusst, wie ein Sachverhalt wahrgenommen wird. In der Kürzung kann sich eine Verschiebung der Gewichtung oder sogar eine unbeabsichtigte Sinnveränderung ergeben.
- **Transparenz und Kennzeichnung:** Das Ersetzen menschlicher Stimmen durch KI kann für das Publikum nur dann vertrauenswürdig sein, wenn die Herkunft klar und eindeutig gekennzeichnet ist.
- **Vertrauensschutz:** Die Stimme als Trägerin journalistischer Autorität ist ein zentrales Element der Glaubwürdigkeit von Nachrichten. Eine KI-Stimme verändert dieses Vertrauensverhältnis und muss daher besonders sorgfältig eingesetzt werden.
- **Verantwortung und Rechenschaftspflicht:** Durch die Automatisierung wird die Gefahr verstärkt, dass redaktionelle Verantwortung unklar verteilt oder vollständig an technische Systeme delegiert wird.

Im Hintergrund stehen dabei grundlegende Werte des journalistischen Selbstverständnisses:

- **Wahrheitspflicht:** Die Verpflichtung, korrekte und überprüfte Informationen zu liefern.
- **Unabhängigkeit:** Die Freiheit von unzulässiger Einflussnahme, auch durch technische Systeme, deren Funktionsweise nicht vollkommen transparent sind.
- **Verantwortlichkeit:** Die klare Zuweisung von Entscheidungskompetenz und Verantwortung an menschliche Akteure.





- **Publikumsorientierung:** Die Pflicht, Informationen so bereitzustellen, dass sie verständlich und im Kontext nachvollziehbar sind.

In der praktischen Umsetzung ergeben sich mehrere Herausforderungen:

- **Technische Unschärfen:** KI-gestützte Textkürzungen oder Bildschnitte können Nuancen falsch interpretieren, übersetzen oder unbeabsichtigt irreführende Schwerpunkte setzen.
- **Plattformlogiken:** Social-Media-Formate folgen eigenen Aufmerksamkeitsstrategien (z. B. Kürze, Emotionalität, visuelle Dynamik), die mit den Prinzipien des seriösen Nachrichtenjournalismus in Konflikt geraten können.
- **Kennzeichnungspflicht in fragmentierten Umgebungen:** In Social-Media-Feeds werden Inhalte oft ohne den vollständigen Kontext konsumiert, was die Möglichkeiten und Wirksamkeit von Kennzeichnungen erschwert.
- **Verlust der redaktionellen Kontrolle:** Bei starker Automatisierung kann die menschliche Endprüfung zu einem formalen Prozess werden, der nur noch Fehlerkorrektur leistet, ohne eine echte inhaltliche Verantwortung wahrzunehmen.

Eine ethische Prüfung dieses Use-Cases ist daher nicht nur angezeigt, sondern notwendig, um sicherzustellen, dass die Einführung von KI in die journalistische Produktionskette nicht zu einer schleichenden Absenkung redaktioneller Standards führt. Sie dient dazu, die eingesetzten Systeme und Prozesse so zu gestalten, dass sie den Grundwerten des Journalismus verpflichtet bleiben und das Publikum weiterhin auf die Integrität der vermittelten Informationen vertrauen kann.

Im folgenden Abschnitt werden die zuvor definierten Ethischen Leitlinien für den Einsatz von KI-Stimmen im Nachrichtenjournalismus auf diesen konkreten Anwendungsfall angewendet. Die Handlungsempfehlungen orientieren sich entlang der einzelnen Leitlinien, um zu prüfen, inwieweit der Use-Case diesen Anforderungen entspricht und wo Risiken, Anpassungsbedarfe oder Schutzmechanismen notwendig sind. Die Wahrung der Rechte der Sprecherinnen und Sprecher und deren Einverständnis werden als gegeben vorausgesetzt und nicht noch einmal separat aufgeführt.

Anwendung der Ethischen Leitlinien auf den Use-Case

Leitlinie 1: Zielbindung und Kontextualität

- **Bewertung:** Die automatisierte Kürzung und Neuformulierung birgt das Risiko einer Bedeutungsverschiebung, insbesondere wenn die KI-Inhalte nach algorithmischen Relevanzkriterien auswählt.



- 
- **Empfehlung:** Eine redaktionelle Endkontrolle ist verpflichtend, um sicherzustellen, dass alle wesentlichen Fakten korrekt und vollständig wiedergegeben werden. Kürzungen dürfen keine inhaltlichen Verzerrungen erzeugen.

Leitlinie 2: Transparenz und Kennzeichnungspflicht

- **Bewertung:** In Social-Media-Umgebungen, in denen Inhalte häufig isoliert konsumiert werden, ist eine deutliche Kennzeichnung der KI-Stimme besonders wichtig.
- **Empfehlung:** Jede Version – unabhängig von Plattform und Länge – muss eindeutig markieren, dass die Tonspur KI-generiert ist. Dies gilt sowohl für den visuellen als auch für den akustischen Bereich.

Leitlinie 3: Schutz vor Täuschung und Manipulation

- **Bewertung:** Die Stimme ist ein Schlüsselträger von journalistischer Glaubwürdigkeit. Eine KI-Stimme kann, wenn sie emotionalisiert oder stilistisch von der Nachrichtennorm abweicht, das Vertrauen untergraben oder gezielt zur Meinungsbildung eingesetzt werden.
- **Empfehlung:** Auswahl und Training der KI-Stimme müssen auf Neutralität, Klarheit und Seriosität ausgerichtet sein. Plattformtypische Emotionalisierung ist zu vermeiden, auch um die Trennlinie zu Entertainmentformaten deutlich zu wahren.

Leitlinie 4: Verantwortung und Rechenschaftspflicht

- **Bewertung:** Die KI übernimmt in dem Use-Case mehrere Bearbeitungsebenen (Text, Bild, Ton), die in ihrer Kombination erhebliche Gestaltungsmacht haben.
- **Empfehlung:** Die Verantwortung für die Endfassung liegt ausschließlich bei der Redaktion. Prüfprozesse müssen so gestaltet sein, dass menschliche Kontrolle nicht nur formal, sondern substantiell erfolgt.

Leitlinie 6: Ethische Gestaltung affektiver Sprechmodulation

- **Bewertung:** Die Anpassung an Social-Media-Formate darf nicht zu einer Verwischung der Grenze zwischen Nachrichten und Unterhaltung führen.
- **Empfehlung:** KI-generierte Schnitte, Bildauswahl und Textfassungen müssen stets im Nachrichtenkontext bleiben. Gestalterische Elemente, die eine Vermischung mit Entertainment-Inhalten fördern, sind zu vermeiden.



Fazit

Der beschriebene Use-Case kann unter Einhaltung der Leitlinien umgesetzt werden, erfordert jedoch konsequente redaktionelle Prüfprozesse, klare Kennzeichnungspflichten und eine bewusste Gestaltung der KI-gestützten Elemente im Hinblick auf Wahrhaftigkeit, Transparenz und Vertrauensschutz. Die größte Herausforderung liegt in der konsequenten Umsetzung dieser Standards in schnelllebigen Social-Media-Formaten, in denen Zeitdruck und Plattformlogiken häufig im Konflikt mit journalistischen Qualitätsmaßstäben stehen.





Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

© 2025

Institut für Digitale Ethik (IDE)
Hochschule der Medien (HdM)
Nobelstr. 20
70569 Stuttgart

Kontakt

Prof. Dr. Petra Grimm
grimm@hdm-stuttgart.de

Design & Layout

Hauke Janowsky