die
medienanstalten

Dr. Paul Grünke • Simon Litsche • Sandra Starchenko

Demokratiekompetenz stärken

Herausforderung Künstliche Intelligenz und die Vermittlung von Medienkompetenz

Demokratiekompetenz stärken

Herausforderung Künstliche Intelligenz und die Vermittlung von Medienkompetenz

Dr. Paul Grünke
Simon Litsche
Sandra Starchenko

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Alle Online-Verweise, die in der vorliegenden Publikation enthalten sind, wurden von uns sorgfältig geprüft. Sofern nichts anderes vermerkt ist, wurden sie am 01.03.2024 letztmalig abgerufen.

die medienanstalten – ALM GbR
Friedrichstraße 60
10117 Berlin
Tel: +49 30 206 46 90 0
Fax: +49 30 206 46 90 99
E-Mail: info@die-medienanstalten.de
Website: www.die-medienanstalten.de

Verantwortlich
Dr. Eva Flecken, Vorsitzende der Direktorenkonferenz der Landesmedienanstalten (DLM)

Herausgeber
Albrecht Bähr, Vorsitzender der Gremienvorsitzendenkonferenz der Landesmedienanstalten (GVK)
Martin Gorholt, stellv. Vorsitzender der Gremienvorsitzendenkonferenz der Landesmedienanstalten

Projektleitung und Redaktion
Dr. Shina-Nancy Erlewein, Referentin,
Gemeinsame Geschäftsstelle der Medienanstalten

Bildnachweis
sdecoret – stock.adobe.com

Gestaltung und Satz
Rosendahl Borngräber GmbH
Kastanienallee 71, 10435 Berlin
E-Mail: mail@rosendahl-berlin.de
Website: www.rosendahl-berlin.de

Stand: März 2024

Inhalt

Executive Summary

Die fortschreitende Entwicklung von Künstlicher Intelligenz (KI) und insbesondere generativer KI hat einen tiefgreifenden Einfluss auf den Medienbereich und verändert die Art und Weise, wie Nachrichten konsumiert und produziert werden. KI ist hierbei von ambivalenter Natur. Sie bietet einerseits Chancen, beispielsweise durch Personalisierung und Automatisierung, mehr Menschen mit relevanten Informationen zu erreichen und zu einer besser informierten Gesellschaft beizutragen. Andererseits bringt der Einsatz von KI auch Risiken mit sich, die sich beispielsweise in Form von Desinformation oder Filterblasen manifestieren können. Um die Veränderungen des Medienbereichs erfolgreich zu gestalten, müssen die Chancen, die KI für die Gesellschaft bietet, genutzt und zugleich die Risiken minimiert werden. Ein essenzieller Bestandteil ist dabei eine Stärkung der Medienkompetenz der Bürger:innen.

In modernen Gesellschaften fungieren – insbesondere digitale – Medien als zentrale Informationsquellen. Das Internet und soziale Medien werden bei einem stetig wachsenden Anteil der Bevölkerung, insbesondere bei jungen Menschen, zur Hauptnachrichtenquelle und auch die gesellschaftliche Meinungsbildung verlagert sich vermehrt auf digitale Plattformen. Informations- und Meinungsbildungskompetenz ist daher eine unverzichtbare Voraussetzung, um aktiv an der gesellschaftlichen Meinungsbildung und damit am demokratischen Prozess teilzunehmen, informierte Entscheidungen zu treffen und sich in einem pluralistischen Umfeld zu orientieren. Es ist daher essenziell, dass die Bürger:innen eine Medienkompetenz erlangen, die dem veränderten Medienbereich gerecht wird. Nur so können sie mündig politisch partizipieren, was wiederum notwendig für eine funktionierende Demokratie ist.

In diesem Gutachten werden die Auswirkungen von KI auf den Medienbereich untersucht und Maßnahmen zur Stärkung der Medienkompetenz vorgeschlagen. Dazu werden zunächst die Veränderungen betrachtet, die durch KI im Medienbereich bereits stattgefunden haben und noch erwartet werden. Hierbei werden einerseits konkrete Anwendungen diskutiert, die auf maschinellem Lernen (Machine Learning) und generativer KI basieren sowie andererseits das systemische Risiko von Desinformation betrachtet, das insbesondere durch generative KI an Relevanz und Gefahrenpotenzial zugenommen hat.

KI-basierte Anwendungen können im Medienbereich zu signifikanten Veränderungen führen. Ein zentraler Einsatzbereich sind algorithmische Empfehlungssysteme, die die Grundlage für die Inhaltsverteilung auf sozialen Medien und anderen Online-Plattformen bilden. Diese Systeme personalisieren die präsentierten Inhalte, oft mit dem Ziel, die Nutzer:innen länger auf der jeweiligen Plattform zu halten. Für eine individualisierte Benutzererfahrung können Inhalte an die persönlichen Vorlieben und Interessen angepasst werden.

Ein weiterer Einsatzbereich von KI ist das automatisierte Content-Management. Hierbei kann KI beispielsweise genutzt werden, um Plattforminhalte automatisch nach unerwünschten oder illegalen Inhalten zu filtern. Dies kann zur Schaffung sichererer und qualitativ hochwertigerer Onlineumgebungen beitragen, indem potenziell schädliche oder unangemessene Inhalte effektiv identifiziert und entfernt werden.

KI-Werkzeuge haben die Möglichkeiten für Journalist:innen bei der Datenanalyse und Recherche fundamental verändert. Sie werden dazu verwendet, Informationen zu sammeln, große Mengen an Daten zu analysieren und den Verifikations- oder Falsifikationsprozess von Inhalten zu unterstützen. Dies ermöglicht eine effizientere und präzisere Informationsbeschaffung.

Des Weiteren gewinnt die synthetische Medienproduktion an Bedeutung, bei der mittels generativer KI-Technologien synthetische Inhalte wie Bilder, Videos oder Texte erstellt werden.

Die Ambivalenz der Technologie zeigt sich deutlich in all diesen Anwendungen. KI kann als äußerst nützliches Werkzeug dienen, um den Zugang zu Informationen zu erleichtern, Inhalte präzise auf die individuellen Bedürfnisse der Konsument:innen zuzuschneiden und in kürzester Zeit relevante Inhalte zu erstellen. Diese Effizienzgewinne können dazu beitragen, die Informationsflut zu bewältigen und personalisierte Erlebnisse für Nutzer:innen zu schaffen. Gleichzeitig birgt KI in diesen Anwendungen auch erhebliche Herausforderungen und Risiken. Die Personalisierung von Inhalten kann zu einer Einschränkung der dargestellten Meinungsvielfalt oder sogar zur Darstellung falscher Informationen führen. So entstehen Filterblasen, die eine gesellschaftliche Polarisierung begünstigen. Bias in KI-Anwendungen sowie noch ungeklärte Fragen des Urheberrechts bei generativer KI sind weitere Herausforderungen.

Die potenziellen negativen Effekte von KI im Medienbereich werden insbesondere beim systemischen Risiko der Desinformation sichtbar, das auf mehreren der oben genannten Entwicklungen basiert. Desinformation ist zwar kein neues Phänomen, KI hat die Erstellung und Verbreitung von irreführenden Informationen jedoch signifikant vereinfacht. Es besteht die Befürchtung, dass das Internet zukünftig von Misinformation und Desinformation verschiedener Akteure mit vielfältigen Zielsetzungen überschwemmt wird. Dies könnte eine ernsthafte Bedrohung für die Informationsfreiheit darstellen. Rezipient:innen stünden vor der Herausforderung, die Faktizität von Inhalten zu beurteilen, können auf Social-Media-Plattformen aber nicht auf seriöse Medienintermediäre vertrauen. Vorschlagsalgorithmen von digitalen Plattformen können ausgenutzt werden, um Inhalte relevanter erscheinen zu lassen und ein breites Publikum zu erreichen. Die schiere und möglicherweise omnipräsente Menge an überzeugender Desinformation erschwert eine Bewertung durch die Nutzer:innen erheblich und könnte das Vertrauen in Online-Inhalte untergraben. Insbesondere im Kontext von Wahlen kann dies auch zur Bedrohung für Demokratien werden.

In diesem Gutachten werden **drei zentrale Handlungsfelder** identifiziert, um die Chancen von KI zu nutzen und gleichzeitig die durch KI entstehenden Risiken zu minimieren: Regulierung, technologische Maßnahmen und Stärkung der Medienkompetenz. Die ersten beiden Handlungsfelder verfolgen das Ziel, schädliche Inhalte aus den Medien zu verbannen bzw. synthetisch generierte Inhalte zu kennzeichnen, um eine Überforderung der Bürger:innen zu verhindern.

Regulierung
Die Politik muss Rahmenbedingungen setzen, um Rechtssicherheit bezüglich der Anwendung von KI zu schaffen und um insbesondere auf Social-Media-Plattformen die Verantwortlichkeiten zu klären. Hierfür hat die EU mit der Datenschutzgrundverordnung und dem Digital Services Act bereits zentrale Verordnungen verabschiedet und auch der geplante AI Act wird einen wichtigen Beitrag leisten, um Risiken durch KI zu minimieren und die Sicherung der Grundrechte zu stärken. In der Umsetzung der jeweiligen Verordnungen ist noch einiger Aufwand nötig, um ihre Durchsetzung

in den Mitgliedsstaaten zu ermöglichen. Langfristige oder implizite Risiken für die Gesellschaft sind in den Verordnungen bisher kaum adressiert und müssen weiter Thema der politischen Diskussion bleiben.

Öffentliche wie auch private Medienanbieter sollten sich über die Regulierung hinaus im Rahmen von Normungs- und Standardisierungsprozessen oder durch Selbstverpflichtungen damit beschäftigen, wie sie mit KI umgehen möchten. Dies ist notwendig, um den hohen journalistischen Standard und das daraus resultierende weiterhin hohe Vertrauen in etablierte Medien zu erhalten. Viele Medienhäuser haben bereits entsprechende Selbstverpflichtungen veröffentlicht.

Technologische Maßnahmen
Technische Maßnahmen zur Detektion und Kennzeichnung von KI-Inhalten sowie zur Authentifikation von menschengemachten Inhalten sind ein weiterer Baustein. Kennzeichnung von KI-generierten Inhalten oder Entscheidungen, die von bzw. mithilfe von KI getroffen werden, führen zu mehr Transparenz in der Kommunikation und gleichzeitig zu einem besseren Wissen bei den Nutzer:innen darüber, wo KI überall involviert ist. Neben Kennzeichnungen und Authentifizierungsmethoden sind insbesondere Werkzeuge hilfreich, die die Detektion von KI-erzeugten Inhalten ermöglichen und den Nutzer:innen selbst die Möglichkeit geben, Inhalte zu prüfen. Diese sind aktuell jedoch vor allem für Textinhalte noch nicht sehr verlässlich. Eine Authentifikation von Gesprächspartner:innen, Artikeln, Bildern, Videos etc. kann beispielsweise über Wasserzeichen oder andere Arten von Kennzeichnungen erfolgen, die direkt bei der Erstellung des Mediums vergeben werden und technisch so gestaltet sind, dass sie nur schwer gefälscht werden können. Dies kann zu mehr Vertrauen in die jeweiligen Inhalte führen und so sowohl für die Erstellenden einen Vorteil erzeugen als auch für die Nutzer:innen eine gute Orientierungshilfe bieten. Plattformbetreiber erhalten dadurch zusätzliche Möglichkeiten zur Filterung und Strukturierung von Inhalten. Der Umgang mit solchen Kennzeichnungen und Detektionswerkzeugen erfordert jedoch Technologie- und Medienkompetenz bei den Nutzer:innen. Die Technikgestaltung muss die Nutzer:innen aktiv mitdenken und es müssen Angebote geschaffen werden, die die Kennzeichnungen und Werkzeuge adressatengerecht erklären, um Akzeptanz und Vertrauen zu erzeugen. Technologische Ansätze können, ebenso wie regulatorische Maßnahmen, jedoch immer nur unterstützend wirken.

Medienkompetenz
Das dritte Handlungsfeld ist daher die Stärkung der Medienkompetenz. Dies betrifft sowohl die Medienschaffenden wie auch die Rezipient:innen. Das vorliegende Gutachten argumentiert, dass hierfür im Kontext von KI insbesondere eine Stärkung der Technologiekompetenz notwendig ist. Ein grundlegendes Verständnis von KI ist notwendig, um Kompetenzen wie Medienanalyse oder Medienkritik zu entwickeln. Technologiekompetenz meint in diesem Zusammenhang nicht nur das Verständnis technologischer Mechanismen, sondern beinhaltet auch die aktive Handhabung von Technologie. Im Zusammenhang mit KI bedeutet dies, Medieninhalte nicht nur passiv zu konsumieren, sondern die Algorithmen und Mechanismen hinter diesen Inhalten zu verstehen und die kreativen Potenziale der Technologie zu entdecken.

Medienkompetenz lässt sich durch verschiedene Maßnahmen fördern, die teilweise bereits erfolgreich umgesetzt werden und in Zukunft an die veränderte Medienlandschaft angepasst werden müssen. Eine effektive Strategie zeigt sich in praxisorientierten Angeboten, die einen interaktiven Austausch mit Fachexpert:innen ermöglichen. Dieser Ansatz findet bereits Anwendung, sei es in schulischen Projekten, Konferenzen oder durch Medienunternehmen, die Einblicke in ihre

Produktionsprozesse gewähren. Publikationen und Handbücher spielen eine maßgebliche Rolle bei der Vermittlung von Medienkompetenz, indem sie nicht nur Wissen vermitteln, sondern auch Leitfäden und praxisnahe Anleitungen bieten. Online-Beratungen und -Seminare sind ebenfalls zentral, da sie einen barrierefreien Zugang zu Bildungsinhalten ermöglichen, unabhängig vom geografischen Standort. Kampagnen zur Stärkung der Medienkompetenz können ein kritisches Verständnis für Medieninhalte fördern, indem sie gezielt zu Themen wie Faktenprüfung, sicherem Online-Verhalten und Gefahren von Desinformation aufklären. Diese Sensibilisierung ermöglicht eine sicherere Teilnahme der Bürger:innen im digitalen Raum.

KI kann für die Entwicklung neuer und innovativer Maßnahmen zur Vermittlung von Medienkompetenz sehr hilfreich sein. So ermöglichen KI-basierte Lernplattformen personalisierte Schulungen. Individuelle Lernpfade fördern eine differenzierte Medienkompetenzentwicklung, während die kritische Auseinandersetzung mit KI ein reflektiertes Verständnis der Technologie fördert. KI-gestützte Chatbots oder virtuelle Assistenten bieten die Möglichkeit, Fragen zur Medienkompetenz zu beantworten, Ratschläge zu geben und zielgerichtet Informationen bereitzustellen. Diese interaktiven Anwendungen fördern selbstgesteuertes Lernen und bieten einen unkomplizierten Zugang zu relevanten Quellen. Mithilfe von KI können personalisierte Nachrichtenaggregatoren erstellt werden. Gut umgesetzt können sie eine ausgewogene Informationsaufnahme fördern und die Nutzer:innen dabei unterstützen, verschiedene Perspektiven und Quellen zu berücksichtigen.

Damit die Maßnahmen zur Vermittlung von Medienkompetenz den gewünschten systemischen Effekt erzielen, ist es entscheidend, dass große Teile der Bevölkerung erreicht werden. Daher sollten die verschiedenen Ansätze möglichst vielen Menschen, mit vielfältigen Hintergründen und Lebenssituationen, möglichst leicht zugänglich gemacht werden.

Für Medienschaffende bieten KI-Anwendungen zahlreiche Potenziale. Diese zu nutzen, erfordert ein umfangreiches Verständnis im Umgang mit den neuen Technologien. Der Umgang mit KI-Anwendungen sollte daher fester Bestandteil der journalistischen Ausbildung sein, etwa in Journalismus-Schulen und Universitäten. Medienhäuser müssen sich auch intern digitalisieren, um ihre Daten für den digitalen Markt nutzbar zu machen.

Forschung zu Medienkompetenz ist entscheidend, um den technologischen Veränderungen in der Medienwelt gerecht zu werden. Basierend auf Forschungsergebnissen können Bildungsstrategien entwickelt, angepasst und evaluiert werden, die sicherstellen, dass die Bürger:innen die für den Umgang mit Medien notwendigen Fähigkeiten erwerben und somit souverän in der digitalen Welt agieren können. Auch bei den technischen Ansätzen zur Detektion und Authentifizierung von Medieninhalten ist weitere Forschung notwendig, um verlässliche und leicht verständliche Werkzeuge und Kennzeichnungen zu entwickeln.

Neben den klassischen Bildungseinrichtungen wie Schulen und Universitäten spielen für die Umsetzung der Maßnahmen zur Stärkung der Medienkompetenz private und öffentliche Initiativen und insbesondere auch die Medienanstalten eine zentrale Rolle. Letztere können als Förderer, Initiatoren, Kooperationspartner und Multiplikatoren tätig werden. Die Medienanstalten verbreiten relevante Informationen, Standards und bewährte Praktiken. Als Multiplikatoren tragen sie dazu bei, ein einheitliches Verständnis für regulatorische Anforderungen zu schaffen und für Transparenz in Bezug auf rechtliche Rahmenbedingungen zu sorgen. Lokal fördern sie praxisorientierte Medienkompetenzaktivitäten, veröffentlichen Publikationen und unterstützen Online-Beratungsplattformen. Als Förderer unterstützen Medienanstalten durch finanzielle Mittel für Projekte, die

die Medienkompetenz in der Bevölkerung stärken. Dies umfasst Kampagnen, Prüf- und Aufsichtspraxis sowie die Unterstützung von Bildungseinrichtungen. Schließlich agieren die Medienanstalten als Kooperationspartner für verschiedene Institutionen und Organisationen, um gemeinsame Ziele in der Medienbildung zu erreichen. Dies schließt Forschungsprojekte zur Medienkompetenz und die Förderung von KI in der Medienkompetenzvermittlung ein.

Die Medienlandschaft zeigt sich derzeit äußerst dynamisch und ist geprägt von kontinuierlichen Entwicklungen, neuen Regulierungen und Standards. Gleichzeitig gibt es nahezu täglich Nachrichten, die den Einfluss von KI auf die Medienlandschaft verdeutlichen. Die fortschreitende Integration von KI in den Journalismus hat nicht nur Auswirkungen auf die Arbeitsweise von Medienschaffenden, sondern prägt auch die Informationslandschaft und den öffentlichen Diskurs. Daher ist eine kontinuierliche Aufmerksamkeit für die neuesten Entwicklungen sowie eine regelmäßige Anpassung des Wissensstandes unerlässlich, um den Herausforderungen und Chancen in diesem dynamischen Umfeld souverän begegnen zu können.

1 Einleitung

Die rasanten Fortschritte in der Digitalisierung und insbesondere bei Künstlicher Intelligenz (KI) haben den Medienbereich in eine Ära der tiefgreifenden Veränderungen geführt. Diese Veränderungen haben sich durch jüngste Entwicklungen im Bereich der generativen KI intensiviert und beschleunigt. Daraus erwächst eine Vielzahl an Herausforderungen. Der Weltrisikobericht 2024 des World Economic Forum (WEF) identifiziert KI-generierte Desinformation als größtes kurzfristiges globales Risiko, welches nicht nur die politische und gesellschaftliche Polarisierung verstärken, sondern auch demokratische Prozesse beeinträchtigen kann.[1] Insbesondere im Jahr 2024, in dem nahezu die Hälfte der Weltbevölkerung zu Wahlen aufgerufen ist, stellt dies eine fundamentale Herausforderung für Demokratien dar.

Die Auswirkungen von KI-Anwendungen reichen weit über das erhöhte Risiko von Desinformation hinaus, bergen neben Risiken aber auch Chancen. Im Bereich des Journalismus manifestiert sich dieser Wandel unter anderem durch automatisierte Schreibprozesse, bei denen KI-Systeme Nachrichten generieren können. Die automatisierte Inhaltsproduktion kann Effizienzsteigerungen mit sich bringen, wirft gleichzeitig aber auch ethische Fragen bezüglich der Authentizität und Qualität der erstellten Inhalte auf.

Der Digital News Report 2023 zeigt, dass das Internet und insbesondere soziale Medien bei einem stetig wachsenden Prozentsatz der Bevölkerung, insbesondere bei jungen Menschen, zur Hauptnachrichtenquelle wird.[2] Der durch KI ermöglichte Aufstieg digitaler Plattformen als neue Intermediäre ist eine grundlegende Veränderung des Medienbereichs. Personalisierte Algorithmen und Empfehlungssysteme führen dazu, dass sich Nutzer:innen vermehrt in sogenannten Filterblasen bewegen, da ihnen bevorzugt Inhalte angezeigt werden, die ihren eigenen Ansichten entsprechen. Dieser individualisierte Zugang zu Informationen kann zu einer eingeschränkten Perspektivenvielfalt führen. Dieser Wandel dezentralisiert nicht nur die Informationsvermittlung, sondern bringt auch Herausforderungen in Bezug auf die Verantwortlichkeit für verbreitete Inhalte mit sich. Plattformbetreiber fühlen sich nicht zwangsläufig einem öffentlichen Auftrag verpflichtet und erfüllen nicht die Gatekeeper-Rolle, die der traditionelle Journalismus übernimmt.[3] Um den Bürger:innen einen sicheren und mündigen Umgang mit den Plattformen zu ermöglichen, sind daher sowohl Regeln als auch eine Stärkung der Initiativen zur Medienbildung und Medienkompetenzvermittlung erforderlich.

In Anbetracht der zunehmenden Verlagerung von Meinungsbildungsprozessen auf digitale Plattformen wird eine freie und mündige Partizipation an der digitalen Öffentlichkeit zur Voraussetzung für eine gelebte Demokratie. Medienkompetenz wird so auch zur Demokratiekompetenz. In diesem Gutachten werden die Auswirkungen von KI auf den Medienbereich untersucht und mögliche Reaktionen zur Stärkung der Medienkompetenz vorgeschlagen.

1 WEF 2024
2 Reuters 2023
3 Die Gatekeeper-Funktion ist hier positiv gemeint, in dem Sinne, dass Berufsjournalist:innen durch journalistische Qualitätsstandards sowie Unabhängigkeit und Ausgewogenheit der Berichterstattung eine Qualitätssicherung gewährleisten, die bei Inhalten sozialer Netzwerke oder Plattformen nicht gewährleistet werden kann.

2 Begriffe und Problemstellung

In diesem Kapitel werden die für das Gutachten zentralen Begriffe – Künstliche Intelligenz (Kapitel 2.1), Medienkompetenz (Kapitel 2.2) und Demokratiekompetenz (Kapitel 2.3) – kurz eingeführt und die Problemstellung des Gutachtens (Kapitel 2.4) formuliert.

2.1 Künstliche Intelligenz

Der Begriff Künstliche Intelligenz (KI) ist breit und wird sehr unterschiedlich verwendet. Es gibt auch in der Fachwelt keine einheitliche Definition.[4] In diesem Gutachten werden zwei zentrale Aspekte fokussiert, die für die Veränderungen im Medienbereich zentral sind: Maschinelles Lernen (Machine Learning, ML) und Generative KI.

Mit ML-Algorithmen werden in diesem Gutachten Algorithmen beschrieben, die aus Erfahrungen lernen und sich selbst verändern können.[5] Konzeptionell können die Ursprünge von ML in die 1940er und 50er-Jahre zurückdatiert werden. Einen großen Aufschwung erlebten die ML-Algorithmen allerdings erst in den 2010er-Jahren, als vorliegende Datenmengen und Rechenleistung groß genug wurden, um die Möglichkeiten von ML voll auszuschöpfen. Im Medienbereich ermöglicht maschinelles Lernen eine präzisere Analyse von Nutzerverhalten, die Personalisierung von Inhalten und die Vorhersage von Trends. Nachrichtenredaktionen können beispielsweise ML-Algorithmen nutzen, um Leserpräferenzen zu verstehen und maßgeschneiderte Inhalte zu erstellen, während Unternehmen durch ML personalisierte Werbung effektiver gestalten können. ML-Algorithmen sind die Grundlage für personalisierte Feeds auf Social-Media-Plattformen und haben deren Erfolg erst ermöglicht.

Generative KI hat seit der Veröffentlichung von ChatGPT besondere Aufmerksamkeit erregt. Diese Technologie ermöglicht es mithilfe von ML-Modellen, die auf umfangreichen Datensätzen trainiert wurden, auf Basis dieser Trainingsdaten neue Inhalte zu generieren.[6] Im Medienbereich können generative Modelle dazu verwendet werden, automatisch Texte, Bilder, Audioinhalte oder Videos zu erstellen. Diese Fortschritte bergen jedoch nicht nur Potenziale, sondern auch Herausforderungen, insbesondere in Bezug auf Fragen des Urheberrechts sowie der Authentizität von generierten Inhalten und die mögliche Verbreitung von Fehlinformationen.

2.2 Medienkompetenz

Medienkompetenz definiert sich für viele der für dieses Gutachten befragten Expert:innen über die Fähigkeit zum selbstbestimmten und verantwortungsbewussten Umgang mit Medienangeboten. Dieser setzt nicht nur technische Fertigkeiten voraus, sondern erfordert auch Kompetenzen zur kritischen und angemessenen Einordnung. Bei Dieter Baacke findet sich ein rein instrumentelles Verständnis von Medienkompetenz: „Medienkompetenz meint also grundlegend nichts anderes als die Fähigkeit, in die Welt aktiv aneignender Weise auch alle Arten von Medien für das Kommunikations- und Handlungsrepertoire von Menschen einzusetzen.“[7] Dieses Verständnis ist zunehmend in Kritik geraten, da es ein ökonomisches Konzept von Ausbildung nahelegt und nicht die Entfaltung

4 Für einen Überblick über die zentralen Begriffe und Konzepte von KI, siehe zum Beispiel: https://www.plattform-lernende-systeme.de/glossar.html
5 Die Literatur zu ML ist sehr umfangreich. Einen guten Überblick über den konzeptionellen und mathematischen Hintergrund bietet zum Beispiel Goodfellow et al. 2016. Einen kurzen historischen Überblick bietet zum Beispiel Foote 2022.
6 Für einen Überblick über die Funktionsweise von Generativer KI siehe zum Beispiel die Veröffentlichung „Große Sprachmodelle“ der Plattform Lernende System (Löser et al. 2023).
7 Siehe Baacke 1996, S. 119.

einer mündigen Person durch Lehren und Lernen als Zielsetzung versteht.[8] Eine neuere Interpretation des Medienkompetenzbegriffs hat die Kultusministerkonferenz in Form von sechs verschiedenen Kompetenzbereichen vorgestellt, die die „Kompetenzen in der digitalen Welt“ beschreiben: Suchen, Verarbeiten und Aufbewahren; Kommunizieren und Kooperieren; Produzieren und Präsentieren; Schützen und sicher Agieren; Problemlösen und Handeln; Analysieren und Reflektieren. Jeder dieser Bereiche hat ausdifferenzierte Unterpunkte, die die jeweiligen Kompetenzen genauer beschreiben.[9] Die von der Kultusministerkonferenz formulierten Kompetenzen beschreiben das aktuelle Zielbild der schulischen Bildung in Bezug auf Medienkompetenz.

Im Zusammenhang mit KI betonen die Expert:innen, dass aktuell insbesondere die Technologiekompetenz einen hohen Stellenwert einnimmt. Hierbei wird auf einen aktiven Umgang mit der Technologie Wert gelegt, um sich diese anzueignen und kreativ mit den gebotenen Möglichkeiten umzugehen. Kreative Anwendungen erfordern ein grundlegendes Verständnis von Funktionsweisen, Modularität und Erfolgsbedingungen technologischer Anwendungen sowie eine Experimentierlust. Die Technologiekompetenz im Kontext von KI stellt eine wesentliche Voraussetzung für die Fähigkeit zur Medienanalyse und Medienkritik dar. KI hat im Medienbereich eine derart dominante Rolle eingenommen, dass eine umfassende Analyse der Medien erst durch ein Verständnis von KI und ihrer Anwendung im Medienkontext möglich wird.[10] Auch die Ständige Wissenschaftliche Kommission der Kultusministerkonferenz argumentiert in einem Impulspapier dafür, Anwendungen, die auf generativer KI basieren, möglichst schnell in den Schulunterricht zu integrieren, um den Schüler:innen die Interaktion mit der Technologie zu ermöglichen.[11]

2.3 Demokratiekompetenz

Demokratiekompetenz ist von zentraler Bedeutung für eine funktionierende demokratische Gesellschaft. Sie beschreibt die Fähigkeit der Bürger:innen, aktiv und selbstbestimmt am demokratischen Prozess teilzunehmen, informierte Entscheidungen zu treffen und sich in einem pluralistischen Umfeld zurechtzufinden. Die Ausbildung einer Informations- und Meinungsbildungskompetenz ist eine dafür notwendige Voraussetzung.

In modernen Gesellschaften sind Medien, und dabei zunehmend digitale Medien, die primären Informationsquellen der Bürger:innen.[12] Auch die gesellschaftliche Meinungsbildung findet vermehrt auf digitalen Plattformen statt. Eine effektive Demokratie erfordert mündige, informierte Bürger:innen, die in der Lage sind, Quellen kritisch zu bewerten und Fakten von Meinungen zu trennen. Von einem medienethischen Standpunkt aus ist für eine zeitgemäße liberale Demokratie nicht allein die Informationsmündigkeit der Bürger:innen von Bedeutung. Ebenso ist es entscheidend, dass es einen Konsens darüber gibt, dass die Suche nach Wahrheit als unverzichtbares Gut gilt und die Anerkennung der Wissenschaft als Referenzsystem für Fakten unerlässlich ist.[13]

Eine Studie von Meßmer et al. unterstreicht die Verbindung zwischen digitaler Nachrichtenkompetenz und demokratischer Grundhaltung. Letztere beinhaltet die Bereitschaft, sich über Politik zu informieren, die Wertschätzung für unabhängigen Journalismus, ein grundlegendes Vertrauen

8 Vgl. Oeftering 2024
9 Vgl. KMK 2017
10 Vgl. Thiel / Rostalski 2021
11 Vgl. SWK 2024
12 Vgl. Reuters 2023
13 Vgl. Kimmel 2021

in Demokratie und Medien sowie die Fähigkeit, auch andere Meinungen zu tolerieren. Menschen, die diesen Einstellungen ablehnend gegenüberstehen, weisen laut der Studie auch eine geringere Nachrichten- und Informationskompetenz auf.[14]

Medienkompetenz ist eine entscheidende Voraussetzung für Demokratiekompetenz, indem sie es ermöglicht, Informationen zu recherchieren, zu bewerten und zu verstehen. Dies schafft eine solide Grundlage für die demokratische Teilhabe. Zusätzlich fördert Medienkompetenz die aktive Beteiligung an öffentlichen Diskussionen. Bürger:innen, die Medien als Plattform zur Artikulation ihrer Ansichten nutzen können, tragen zur Vielfalt des öffentlichen Diskurses und somit zu einer lebendigen und pluralistischen demokratischen Gesellschaft bei.

2.4 Problemstellung

Das vorliegende Gutachten verfolgt das Ziel, die vielfältigen Formen und Anwendungen von KI im Medienbereich zu identifizieren und ihre Auswirkungen auf den Prozess der Meinungsbildung zu analysieren. Ein besonderer Fokus liegt dabei auf der Erfassung von kommunikations- und gesellschaftspolitischen sowie medienrechtlichen Handlungsbedarfen, die aus der verstärkten Verwendung von KI resultieren. Das Gutachten soll den Medienanstalten als Grundlage dienen, um fundierte Diskussionen über notwendige Anpassungen in den relevanten Bereichen zu führen.

Das Gutachten richtet den Blick gezielt auf die Anwendung von KI im Kontext von Nachrichten- und Informationskompetenz. Hier liegt ein spezifisches Augenmerk darauf, wie KI den Zugang zu Informationen beeinflusst und welche Auswirkungen dies auf die Fähigkeit der Bürger:innen zur kritischen Analyse und Einordnung von Inhalten hat. Ein zentraler Aspekt sind hier die Herausforderungen, die sich durch die vermehrte Nutzung von Desinformationswerkzeugen ergeben.

Das übergeordnete Ziel besteht darin, einen Beitrag zur Stärkung der Medienkompetenz der Bürger:innen zu leisten. Dies ist von großer Bedeutung, um sie zur aktiven Partizipation an demokratischen Prozessen zu befähigen und sie in die Lage zu versetzen, sich informiert in der Gesellschaft zu positionieren. Angesichts der kontinuierlichen und starken Veränderungen in der Medienlandschaft, insbesondere durch den vermehrten Einsatz von KI, eröffnen sich neue Wege der Informationsgewinnung, die sowohl Chancen als auch Gefahren bergen. Die Förderung von Medienkompetenz wird als Schlüssel betrachtet, um eine aktive Teilhabe an Meinungsbildung und Entscheidungsprozessen in einer Demokratie zu gewährleisten. In diesem Gutachten werden Beispiele erfolgreicher Medienkompetenzvermittlung dargestellt sowie Bedarfe identifiziert, die sich durch die neuen Entwicklungen ergeben haben und noch nicht ausreichend adressiert wurden. Darüber hinaus werden Strategien für neue Möglichkeiten zur Vermittlung von Medienkompetenz diskutiert.

14 Vgl. Meßmer et al. 2021

3 Veränderung der Medienlandschaft durch KI – Chancen und Risiken

Die Integration von Künstlicher Intelligenz (KI) in die Medienlandschaft markiert eine Transformation, die sowohl Chancen als auch Herausforderungen mit sich bringt. Nachrichtenredaktionen weltweit investieren verstärkt in KI-basierte Anwendungen, die sowohl die Art und Weise der Recherche als auch den Konsum von Medien grundlegend verändern.[15] Im Fokus stehen dabei klassische Maschinelles Lernen (Machine Learning ML)-Algorithmen und generative KI. Die Anwendung von ML-Algorithmen ermöglicht eine effizientere Bewältigung von Recherchen durch Automatisierung, Big Data Clustering und die Verifikation von Inhalten. Gleichzeitig revolutionieren algorithmische Empfehlungssysteme in Online-Medien den Medienkonsum und bilden die Grundlage für den Erfolg neuer Medien. Diese Algorithmen eröffnen auch neue Möglichkeiten im Content Management, die zur Konformität und Reduzierung von Hate Speech beitragen.[16]

Seit der Einführung von ChatGPT Ende 2022 haben Millionen von Menschen mit einer generativen KI interagiert. Diese Technologie wird nicht nur von Privatpersonen genutzt, sondern auch von Medienschaffenden, die durch generative KI neue Wege der Automatisierung, Personalisierung und Regionalisierung erkunden.

Die Ergebnisse einer Studie von XPLR: MEDIA in Bayern aus dem Jahr 2023 unterstreichen die rasante Akzeptanz von KI in der bayerischen Medienbranche. 78 Prozent der Befragten geben an, dass in ihren Unternehmen bereits KI-Werkzeuge zum Einsatz kommen. Diese werden vor allem genutzt, um Erfahrungen zu sammeln und Zeit bei alltäglichen Aufgaben zu sparen. Trotz des Effizienzgewinns durch die Zusammenarbeit mit KI hegen die Mehrheit der Befragten (69 Prozent) Bedenken hinsichtlich ethischer Aspekte. Diese Sorgen beziehen sich insbesondere auf den potenziellen Einfluss von Bias in den Algorithmen, die zu verzerrten oder ethisch problematischen Inhalten führen können. Auch die Gefahr von Arbeitsplatzverlusten durch die Automatisierung von Aufgaben wird als Herausforderung wahrgenommen. Die Manipulation von Informationen mittels KI stellt eine weitere Bedrohung dar, da gefälschte Nachrichten und manipulierte Inhalte die Vertrauenswürdigkeit von Informationen beeinträchtigen können.[17]

Die Entwicklung neuer Technologien, die für die Informationserzeugung und -verbreitung relevant sind, ist immer ambivalent und bringt eine Vielzahl an Dynamiken und Herausforderungen mit sich. So kann man selbst bei der Erfindung des Buchdrucks rückblickend sehr unterschiedliche Narrative erzählen. Einerseits hat der Buchdruck Texte und Wissen für Millionen von Menschen zugänglich gemacht und war damit eine der Voraussetzungen für die Reformation, die Aufklärung und die heutige Wissenschaft. Gleichzeitig ermöglichte der Buchdruck auch die Verbreitung von Propaganda, Lügen und Hetzschriften. Die Rechtfertigung der systematischen Hexenverfolgung basiert beispielsweise zu signifikantem Teil auf der Veröffentlichung und Verbreitung des so genannten Hexenhammers, dem Malleus Maleficarum.[18] Auch die jetzigen Veränderungen der Medienlandschaft durch KI weisen diese Ambivalenz auf. Einerseits bieten sich Chancen und Potenziale für Innovationen, die beispielsweise eine weitere Demokratisierung von Wissen ermöglichen. Andererseits wird durch KI auch die Erzeugung und Verbreitung von Hassbotschaften und Desinformation vereinfacht und daher voraussichtlich stark zunehmen.[19]

15 Vgl. Müller-Brehm 2021
16 Vgl. Konstan / Riedl 2012.
17 Vgl. Lehmann / Förtsch 2023
18 Vgl. Behringer / Jerouschek 2000
19 Vgl. Müller-Brehm 2021

In diesem Kapitel wird eine detaillierte Untersuchung gegenwärtiger Anwendungen im Bereich der KI, speziell im Kontext des maschinellen Lernens (Kapitel 3.1), sowie der generativen KI (Kapitel 3.2) präsentiert. Innerhalb dieser beiden Kapitel werden unterschiedliche Beispiele diskutiert, welche bereits signifikante Veränderungen in der Medienlandschaft initiiert haben. Kapitel 3.3 widmet sich dem Risiko Desinformation, das aufgrund seines systemischen Charakters und seines möglichen Einflusses auf demokratische Gesellschaften besonderer Aufmerksamkeit bedarf. Die Chancen und Risiken der identifizierten Veränderungen werden in Kapitel 3.4 zusammengefasst.

3.1 ML-Algorithmen

Maschinelles Lernen (Machine Learning, ML) ist ein Teilgebiet der Künstlichen Intelligenz (KI), das Computern die Fähigkeit verleiht, Muster in Daten zu erkennen und darauf basierend Vorhersagen oder Entscheidungen zu treffen.[20] Die Einsatzmöglichkeiten von ML-Algorithmen sind vielfältig und reichen von personalisierten Inhalten über Empfehlungen bis hin zur gezielten Werbung. Diese Technologie prägt die Medienlandschaft maßgeblich, beeinflusst die Art und Weise, wie Inhalte erstellt, präsentiert und konsumiert werden, und erfordert auch ein höheres Maß an Medienkompetenz.[21] Medienunternehmen nutzen diese Technologien unter anderem zur Rechercheunterstützung, in Form von algorithmischen Empfehlungssystemen oder zum automatisierten Content-Management, um ihre Angebote zu optimieren und den sich wandelnden Ansprüchen ihrer Zielgruppen gerecht zu werden. Diese datengetriebene Herangehensweise ermöglicht nicht nur eine effizientere Content-Strategie, sondern auch die Maximierung von Werbeerlösen und die Schaffung einer personalisierten Benutzererfahrung. Diese Veränderungen haben Auswirkungen auf den Medienkonsum der Nutzer:innen. Um diese Auswirkungen als Nutzer:in zu verstehen und einordnen zu können, ist eine Stärkung der Medienkompetenz von entscheidender Bedeutung. Ein grundlegendes Verständnis von ML-Algorithmen und deren Einfluss auf die Auswahl und Präsentation von Informationen in den Medien ist dafür eine notwendige Voraussetzung.

Im nachfolgenden Kapitel wird dargestellt, wie ML zur Rechercheunterstützung (Kapitel 3.1.1) für algorithmischen Empfehlungssysteme (Kapitel 3.1.2) oder für das Content Management (Kapitel 3.1.3) angewandt wird.

3.1.1 Rechercheunterstützung

Die fortschreitende Entwicklung von KI revolutioniert die Art und Weise, wie wir Informationen recherchieren. Durch Automatisierung wird der Recherchevorgang effizienter, indem KI-Algorithmen repetitive Aufgaben übernehmen. Datenanalyse ermöglicht eine tiefgreifende Auswertung großer Informationsmengen, wodurch relevante Erkenntnisse schneller identifiziert werden können. In Bezug auf die Verifikation von Inhalten bietet KI präzise Werkzeuge zur Überprüfung der Authentizität von Informationen, was die Zuverlässigkeit von Rechercheergebnissen stärkt.[22]

Automatisierung
In der heutigen digitalen Ära spielen KI und Automatisierung eine entscheidende Rolle in journalistischen Arbeitsprozessen. Die fortschreitende Entwicklung von KI hat bahnbrechende Möglichkeiten für den Journalismus eröffnet, insbesondere im Bereich der Rechercheunterstützung. Durch die Textanalyse und

20 Vgl. Konstan / Riedl 2012
21 Vgl. Long / Magerko 2020
22 Vgl. Heesen et al. 2023

-verarbeitung können KI-Modelle enorme Mengen von Textdaten durchsuchen, Muster identifizieren und relevante Informationen extrahieren. Weiterhin kann KI auch dazu eingesetzt werden, automatische Zusammenfassungen zu erstellen. Dies bedeutet für Journalist:innen und Redaktionen eine erhebliche Effizienzsteigerung bei der Suche nach relevanten Inhalten sowie bei der Durchführung umfassender Analysen. Die zeitaufwendige Aufgabe der Informationsbewältigung wird durch die präzise Verarbeitung von KI-Modellen erheblich erleichtert.[23]

Ein weiterer bedeutender Beitrag von KI liegt in der Themenverfolgung und Trendanalyse. Die Technologie ermöglicht es, kontinuierlich Nachrichtenquellen und soziale Medien zu überwachen, um aktuelle Themen und Trends zu identifizieren. Diese Echtzeit-Analyse unterstützt Redaktionen dabei, stets auf dem neuesten Stand zu bleiben und ihre Berichterstattung an aktuellen Entwicklungen auszurichten. Die Anpassungsfähigkeit an sich wandelnde Themenlandschaften wird durch die automatische Erkennung und Priorisierung relevanter Inhalte optimiert.

Insgesamt transformiert die Integration von KI-Unterstützung die Art und Weise, wie Journalist:innen Informationen sammeln und analysieren. Trotz der zahlreichen Vorteile ist es jedoch wichtig, die Grenzen und möglichen Risiken der Technologie zu berücksichtigen, um sicherzustellen, dass die menschliche Redaktionsarbeit weiterhin durch kritische Überlegungen und ethische Standards geprägt ist.

Datenanalyse / Big Data Clustering
Maschinelles Lernen revolutioniert die Art und Weise, wie wir komplexe Datenmengen verstehen und nutzen können. ML ermöglicht nicht nur die Bewältigung riesiger Datensätze, sondern auch das Extrahieren von wertvollen Erkenntnissen und Mustern.[24] Ein prominentes Beispiel dafür ist die Identifizierung von Mustern in Satellitendaten, ein Bereich, der durch ML deutlich erleichtert wird. Die Identifikation von Mustern in Satellitendaten ist eine komplexe Aufgabe, die dank der Fortschritte im Bereich der KI nun effizienter bewältigt werden kann. Ein Beispiel hierfür lieferte das Online-Medienportal BuzzFeed. Durch die Analyse von Flugdaten, die von der Webseite Flightradar24 stammten, veröffentlichte BuzzFeed Informationen über Überwachungsflugzeuge, die unter zivilen Scheinfirmen registriert waren. Um dies zu ermöglichen, wurde ein ML-Algorithmus trainiert, der Flugmuster analysierte und bekannten Flugrouten des FBI oder des US-Heimatschutzministeriums ähnelte. Der Algorithmus berücksichtigte dabei Flugdaten wie Squawk-Codes, Flughöhe, -geschwindigkeit und Kurvengeschwindigkeiten. Diese Anwendung verdeutlicht die Fähigkeit von ML-Algorithmen, komplexe Muster in großen Datenmengen zu identifizieren und aufschlussreiche Erkenntnisse zu liefern. Die Enthüllung löste eine Debatte über die Balance zwischen nationaler Sicherheit und Datenschutz aus.[25]

Ein weiteres Beispiel für die Bewältigung großer Datenmengen durch KI-Systeme sind die erfolgreichen journalistischen Investigationen wie die Panama Papers oder OpenLux. Für Ersteres kamen Algorithmen zum Einsatz, um über 2,6 Terabyte Daten aus 11,5 Millionen Dokumenten zu sichten.[26] Die Verwendung von Graph-Datenbanken wie *Neo4j,* wie es bei den Panama Papers der Fall war,[27]

23 Vgl. Heesen et al. 2023
24 Vgl. Heesen et al. 2023
25 Vgl. Aldhous 2017
26 Vgl. Diakopoulos 2019; Goldhammer et al. 2019
27 Vgl. Hunger 2021

unterstreicht die technologische Dimension solcher Projekte. Einzelne Zeitungen haben inzwischen eigene Abteilungen für Datenjournalismus, die sich auf den Einsatz von KI-Systemen für die Analyse und Auswertung komplexer Daten spezialisiert haben.[28]

Eine Gruppe von Forschenden verwendete Satellitenbilder, GPS-Daten von Schiffen und Deep-Learning-Modelle, um die Aktivitäten von Industrieschiffen und Offshore-Energieinfrastrukturen in den weltweiten Küstengewässern zu kartieren. Mit diesem Vorgehen konnten sie nachweisen, dass rund zwei Drittel der weltweiten Fischereiflotten nicht erfasst werden. Ohne die Zuhilfenahme von KI wären diese Erkenntnisse nicht möglich gewesen.[29]

Insgesamt zeigen diese Beispiele, wie die Kombination von großen Datensätzen und ML nicht nur die Identifikation von Mustern in komplexen Datensätzen ermöglicht, sondern auch die effiziente Analyse großer Datenmengen für investigative und journalistische Zwecke vorantreibt. Allerdings kann man an diesen Beispielen auch verdeutlichen, dass Big Data Clustering trotz seiner leistungsstarken Anwendungen nicht frei von Herausforderungen ist. Die sorgfältige Auswahl von Algorithmen und die Berücksichtigung der Datenkomplexität sowie eine angemessene Auswahl der erklärenden Variablen sind entscheidende Aspekte, um die Qualität und Zuverlässigkeit von Clustering-Analysen zu gewährleisten.

Verifikation von Inhalten
Millionen neuer Artikel werden jeden Tag online veröffentlicht. Viele davon beinhalten falsche oder irreführende Inhalte. Dies macht es sowohl für Konsument:innen als auch Journalist:innen herausfordernd, die Faktizität von Inhalten zu beurteilen. Deshalb wird vermehrt auf innovative Technologien zurückgegriffen, um bei der Einschätzung der Glaubwürdigkeit von Quellen zu unterstützen. Ein exemplarisches Beispiel für eine solche Technologie ist das System *Truthmeter,* entwickelt von der Ludwig-Maximilians-Universität.[30] Dieses System dient als Hilfsmittel, um den Verifikationsprozess zu optimieren und Journalist:innen bei der Beurteilung der Vertrauenswürdigkeit von Informationen zu unterstützen. Der Truthmeter-Ansatz beruht auf fortschrittlichen Methoden des maschinellen Lernens, die es ermöglichen, die Integrität von Quellen zu analysieren und den Grad der Zuverlässigkeit einzuschätzen. Durch den Einsatz von Algorithmen und Mustererkennungstechnologien können Journalist:innen auf eine effiziente Weise überprüfen, ob die von ihnen genutzten Informationen aus vertrauenswürdigen Quellen stammen. Dies trägt dazu bei, die Qualität der journalistischen Arbeit zu verbessern und gleichzeitig das Vertrauen der Leser:innen in die vermittelten Informationen zu stärken.[31]

Verringerung von Media Bias
Eine zentrale Herausforderung für klassische Medien besteht darin, Media Bias zu erkennen und zu vermeiden, um eine ausgewogene und objektive Berichterstattung zu gewährleisten.[32] KI hat in diesem Kontext eine entscheidende Rolle übernommen, um gezielte Schritte zur Bekämpfung von Framing und zur Förderung ausgewogener Recherchen zu unternehmen.[33] Ein Beispiel ist das *Aijo Project* der London School of Economics and Political Science. Hier wurde ein KI-System entwickelt, um Bias in Medienorganisationen zu identifizieren und somit einen effektiven Mechanismus zur Überwachung journalistischer Praktiken zu bieten. Die Technologie zielt darauf ab, Framing durch

28 Vgl. Kreye 2021
29 Vgl. Paolo et al. 2024
30 Vgl. Fletcher et al. 2020
31 Vgl. Wilczek / Haim 2022
32 Vgl. Heesen et al. 2023
33 Vgl. Hamborg 2019

bestimmte Wortwahl oder Themenwahl aufzudecken und ermöglicht es, alternative Informationsangebote gezielt vorzuschlagen.[34] Mittlerweile setzen Nachrichtenorganisationen wie Reuters, Agence France Presse und die Deutsche Welle diese Methodik ein. Die Integration von KI zur Aufdeckung von Media Bias ist zu einem wichtigen Instrument geworden, um die Qualität der Berichterstattung zu verbessern und sicherzustellen, dass diverse Perspektiven angemessen berücksichtigt werden.

Die vermehrte Nutzung von KI zur Erkennung und Vermeidung von Media Bias in klassischen Medien kann potenziell negative Auswirkungen auf die journalistische Sorgfaltspflicht haben. Journalist:innen könnten sich auf automatisierte Systeme verlassen und dabei die sorgfältige Prüfung nuancierter Kontexte und individueller Aspekte eines bestimmten Themas vernachlässigen. Dies könnte zu oberflächlicher Berichterstattung, ohne eine entsprechende journalistische Sorgfaltspflicht, gründliche Recherche und kritische Analyse, führen. Es ist daher entscheidend, dass der Einsatz von KI in der Medienbranche mit Bedacht erfolgt und die journalistische Integrität gewahrt bleibt.

Verifikation in sozialen Medien

Die Verifikation von Informationen und Quellen in sozialen Medien ist in der heutigen digitalen Ära von entscheidender Bedeutung, insbesondere angesichts der Verbreitung von Falschnachrichten und manipulierten Inhalten. KI spielt dabei eine immer wichtigere Rolle, um die Authentizität von Informationen sicherzustellen und die Verbreitung von irreführenden oder gefälschten Inhalten einzudämmen.[35] Ein bedeutendes Anwendungsgebiet von ML-Algorithmen in diesem Kontext ist die Aufdeckung und Bekämpfung von Deepfakes (siehe auch Kapitel 3.3.2). Deepfakes sind künstlich generierte Medieninhalte, wie Videos, Bilder oder Audioinhalte, die mithilfe von KI erstellt werden. Um dieser Herausforderung zu begegnen, setzen Plattformen in sozialen Medien KI-Systeme ein, um Deepfakes zu identifizieren und zu kennzeichnen oder zu entfernen.

Forscher:innen und Entwickler:innen arbeiten kontinuierlich daran, leistungsstarke Algorithmen zu entwickeln, die in der Lage sind, verdächtige Inhalte zu erkennen und zu analysieren. Diese Algorithmen nutzen Mustererkennungstechniken und ML, um Anomalien in Bildern, Audiospuren oder Videos aufzudecken, die auf eine mögliche Manipulation hinweisen könnten. Durch den Einsatz von KI kann auf Plattformen automatisch nach Mustern gesucht werden, die von menschlichen Überprüfer:innen möglicherweise übersehen werden, und so eine schnellere und effizientere Verifikation ermöglichen.[36] Die Verifikation von Informationen in sozialen Medien erfordert jedoch nicht nur technologische Lösungen, sondern auch die Zusammenarbeit von Menschen. Menschliche Überprüfer:innen spielen eine wichtige Rolle, um komplexe Nuancen zu verstehen, die KI möglicherweise nicht erfassen kann. Die Kombination von menschlicher Expertise und KI-Technologie ermöglicht eine umfassendere Kontrolle von Inhalten.[37] Insgesamt tragen KI-Systeme dazu bei, die Verifikation von Informationen und Quellen in sozialen Medien zu verbessern, indem sie die Verbreitung von Deepfakes und anderen Falschnachrichten reduzieren. Es ist jedoch wichtig zu betonen, dass die Entwicklung und Anpassung dieser Technologien fortlaufend erfolgen müssen, um mit den sich ständig weiterentwickelnden Methoden der Desinformation Schritt zu halten.

34 Vgl. Heesen et al. 2021
35 Vgl. Heesen et al. 2020a
36 Vgl. Ozbay / Alatas 2020
37 Vgl. Reuter et al. 2018

Veränderung der Medienlandschaft durch KI – Chancen und Risiken

Umgang mit Deepfakes

Deepfake-Detektion bezieht sich auf die Identifizierung und Erkennung von Deepfakes (siehe dazu auch Kapitel 3.3.2). Bei Deepfakes werden beispielsweise Gesichter in Videos ausgetauscht, Stimmen imitiert und realistisch wirkende Inhalte erstellt, die schwer von authentischen Inhalten zu unterscheiden sind. Die Deepfake-Detektion setzt verschiedene Methoden und Techniken ein, um Anomalien oder Hinweise auf Manipulation in den Medieninhalten zu identifizieren. Dazu gehören die Analyse von Gesichtsmerkmalen, die Überprüfung von Lippenbewegungen, die Untersuchung von Beleuchtung und Schatten, sowie die Anwendung von ML, um Muster und Abweichungen zu erkennen. Forschende entwickeln kontinuierlich fortschrittliche Modelle und Algorithmen, um der ständigen Weiterentwicklung von Deepfake-Technologien beizukommen.[38] Verschiedene Unternehmen im Bereich der Softwareentwicklung haben ihre Aufmerksamkeit auf die Aufgabe der Deepfake-Detektion gerichtet.[39] Gleichermaßen engagieren sich auch Forschungsinstitutionen intensiv in diesem Themenfeld. Das Forschungsprojekt *Deep Fake Total* am Fraunhofer Institut für Angewandte und Integrierte Sicherheit (AISEC) fokussiert sich darauf, Systeme zu entwickeln, die in der Lage sind, Audio- und Videomanipulationen sicher und automatisiert zu detektieren.[40] Durch die Analyse verdächtiger Audio-Dateien werden Deepfakes erkannt, wobei verschiedene Erkennungsmodelle für Audio-Spoofing und Deepfakes zum Einsatz kommen. Audio-Spoofing bezieht sich auf die Praxis, Täuschungen oder Manipulationen in Bezug auf Audiodaten vorzunehmen. Dabei wird versucht, Audiosysteme oder Spracherkennungstechnologien zu überlisten, indem künstlich erzeugte oder veränderte Audiosignale verwendet werden. Das Ziel von Audio-Spoofing kann darin bestehen, automatisierte Systeme zu täuschen oder menschliche Sprecher zu imitieren.[41] Das Deepfake-Detektion-Tool von Fraunhofer AISEC ermöglicht nicht nur die Identifikation von gefälschtem Material, sondern fördert auch den Austausch automatisch erkannter Deepfakes in der Sicherheitsgemeinschaft.[42] Das Fraunhofer Institut betont zudem die Bedeutung der Möglichkeit, originale Inhalte zu kennzeichnen.[43] Hierbei spielt die *Content Authenticity Initiative* (CAI) eine zentrale Rolle. CAI ist eine Kooperation von Medien- und Technologieunternehmen, NGOs, Wissenschaftler:innen und anderen Akteuren, die sich für die weitreichende Einführung eines offenen Industriestandards zur Sicherung der Authentizität und Herkunft digitaler Inhalte einsetzen. Die Mission von CAI besteht darin, Fehlinformationen entgegenzuwirken und die Transparenz bei der Nutzung von KI zu erhöhen, indem leicht zugängliche Tools bereitgestellt werden, die anzeigen, wann KI zur Generierung oder Bearbeitung von Inhalten verwendet wurde.[44] Im Zentrum dieser Initiative steht ein kryptografisches Hash-Verfahren. Bilder werden mit Metadaten angereichert, die Informationen darüber speichern, ob das Bild beispielsweise durch KI generiert wurde. Dieser Ansatz wird auch von der Leica Camera AG in ihren Kameras unterstützt. In ihrem neuesten Kameramodell enthält jede Bildaufnahme eine manipulationssichere Signatur mit wichtigen Details wie Kameramodell, Hersteller und Bildinhalt. Kostenlose Open-Source-Tools von CAI ermöglichen es, zu überprüfen, ob ein Bild bearbeitet wurde oder ein Original ist, und gewährleisten somit eine durchgängige Kette der Authentizität vom Moment der Aufnahme bis zur Veröffentlichung.[45]

38 Vgl. Rana et al. 2022
39 Vgl. zum Beispiel https://deepfakedetector.ai/
40 https://deepfake-total.com/
41 Vgl. Balamurali et al. 2019
42 Vgl. Fraunhofer AISEC 2023
43 Vgl. Fraunhofer AISEC 2023
44 Vgl. https://contentauthenticity.org/
45 Vgl. Leica 2023

Fact Check Explorer
Der *Fact Check Explorer v*on Google[46] ist dafür konzipiert, die Arbeit von Faktenprüfer:innen, Journalist:innen und Forscher:innen bei der Unterscheidung bereits entlarvter und unbestrittener Behauptungen zu erleichtern. Er wird als eine Art Suchmaschine für Faktenchecks verstanden und bietet ein umfassendes Werkzeugset zur Unterscheidung zwischen sachlichen Informationen und fiktivem Inhalt. Er basiert auf dem sogenannten Claim-Review-Markup, einem Mechanismus, der die Erkennung und Darstellung von Faktenprüfinformationen ermöglicht. Diese Plattform aggregiert Faktenchecks, die von unabhängigen Organisationen weltweit überprüft wurden, und richtet sich an Journalist:innen und Faktenprüfer:innen, aber auch an Bürger:innen, die die Korrektheit einer Meldung oder Behauptung prüfen möchten. Angesichts der potenziellen Manipulation durch Bilder, die in einen falschen temporalen oder lokalen Kontext gesetzt werden, wurde in der Beta-Version eine neue Funktion eingeführt, die es Faktenprüfer:innen ermöglicht, den Kontext und die zeitliche Entwicklung eines Bildes zu untersuchen und Aufschluss über dessen Indexierungsgeschichte und spätere Verwendung zu erhalten. Google und YouTube haben eine finanzielle Unterstützung in Höhe von 13,2 Millionen US-Dollar für das International Fact-Checking Network (IFCN) zur Gründung eines globalen Faktenckeck-Fonds zugesagt. Durch den Fond werden aktuell über 35 Faktenprüforganisationen in 45 Ländern gefördert.[47]

3.1.2 Algorithmische Empfehlungssysteme

Algorithmische Empfehlungssysteme sind computergestützte Systeme, die auf Algorithmen basieren, um Nutzer:innen personalisierte Empfehlungen für Produkte, Dienstleistungen, Inhalte oder andere Elemente anzubieten. Diese Systeme analysieren das Verhalten, die Vorlieben und die Historie der Nutzer:innen, um Vorhersagen darüber zu treffen, welche Artikel oder Inhalte für sie am relevantesten oder interessantesten sein könnten.[48] Das System trifft Entscheidungen darüber, welche Informationen für die Nutzer:innen am wahrscheinlichsten interessant oder nützlich sind, indem es Muster und Präferenzen aus vergangenen Interaktionen, dem Verhalten ähnlicher Nutzer:innen oder anderen relevanten Daten analysiert. Algorithmische Empfehlungssysteme werden in verschiedenen Kontexten eingesetzt, einschließlich E-Commerce-Plattformen, Streaming-Diensten, sozialen Medien und anderen Online-Plattformen, um die Benutzererfahrung zu personalisieren und relevante Inhalte bereitzustellen.[49]

Personalisierung
Empfehlungssysteme können eine Vielzahl von Vorteilen bieten, insbesondere durch ihre Fähigkeit zur Personalisierung von Inhalten. Diese ermöglicht eine maßgeschneiderte Erfahrung, indem das System das individuelle Verhalten, die Vorlieben und die Historie der Nutzer:innen analysiert. Dies kann zu einer effizienteren Nutzung der Plattform führen, da Nutzer:innen nicht mehr eine Vielzahl von Optionen durchsuchen müssen, um relevante Informationen zu finden. Die Zeitersparnis und die Bequemlichkeit tragen erheblich zur Zufriedenheit der Nutzer:innen bei.[50]

46 https://toolbox.google.com/factcheck/explorer
47 Vgl. Babakar / Sud 2023
48 Vgl. Konstan / Riedl 2012
49 Vgl. Zweig et. al 2017
50 Vgl. Beam 2014

Einfluss auf die Meinungsbildung
Die Personalisierung von KI, insbesondere in der Content-Auswahl, kann zu Echokammern führen, in denen Nutzer:innen verstärkt mit Inhalten konfrontiert werden, die ihre bestehenden Meinungen und Überzeugungen widerspiegeln. Dies hat potenziell negative Auswirkungen auf die Meinungsvielfalt, da Nutzer:innen in einer informellen Filterblase gefangen sein können, die ihre bestehenden Ansichten bestätigt und die Darstellung alternativer Perspektiven reduziert. Nutzer:innen können Personalisierungsoptionen in den wenigsten Fällen aktiv steuern. Im Kontext des US-Wahlkampfs 2016 wurde festgestellt, dass auf Facebook Nachrichten mit möglicherweise geringerer Qualität, aber mit provokativen oder polarisierenden Inhalten, tendenziell vom System eher empfohlen wurden. Während der letzten drei Monate des US-Präsidentschaftswahlkampfs 2016 konnte zudem festgestellt werden, dass falsche Information über die Wahl, die auf Facebook verbreitet wurden, eine höhere Interaktionsrate aufwiesen als die Top-News renommierter und seriöser Nachrichtenquellen.[51]

Das Phänomen, dass sensationalistische oder kontroverse Inhalte oft mehr Aufmerksamkeit erhalten, wird häufig als Clickbait bezeichnet.[52] Es wirft Fragen auf bezüglich der Messung von Nachrichtenqualität und Relevanz in einer Ära, in der Klickzahlen und Aufmerksamkeit oft als Indikatoren für den Erfolg von Medieninhalten betrachtet werden. Wenn Nachrichtenalgorithmen dazu neigen, Inhalte zu bevorzugen, die auf Klicks und Aufmerksamkeit abzielen, könnten sie dazu beitragen, dass Nutzer:innen in ihren persönlichen Filterblasen gefangen bleiben.[53] Das bedeutet, dass Menschen eher Inhalte sehen, die ihren bestehenden Ansichten entsprechen oder diese verstärken, während abweichende Perspektiven vernachlässigt werden. Darüber hinaus besteht die Gefahr, dass diese personalisierte Content-Auswahl zu einer verstärkten Bildung von Echokammern führt. Echokammern bezeichnen soziale Strukturen, in denen Menschen vorwiegend mit Gleichgesinnten interagieren und ihre Ansichten gegenseitig verstärken, während abweichende Perspektiven vernachlässigt werden. Dies kann zu einer Verzerrung der wahrgenommenen Realität und zu einer eingeschränkten Vielfalt von Informationen führen. In Bezug auf den US-Wahlkampf 2016 könnte die Tendenz, kontroverse oder aufsehenerregende Nachrichten zu bevorzugen, dazu beigetragen haben, dass Menschen in ihren politischen Überzeugungen bestärkt wurden, indem sie vermehrt Inhalte sahen, die ihre Standpunkte unterstützten. Dies trägt zur Fragmentierung der Informationslandschaft und zur verstärkten Polarisierung bei, da unterschiedliche Gruppen unterschiedliche Versionen der Realität präsentiert bekommen.

3.1.3 Content Management

Ein bedeutender Einsatzbereich von KI liegt im Content Management, wo KI-Systeme bei der Auswahl und Sortierung von journalistischen Inhalten unterstützen.[54] Die effiziente Verwaltung von Informationen wird durch diese Technologie optimiert. Ein Beispiel hierfür ist die *ARD-Mining Platform*, die es Journalist:innen ermöglicht, in crossmedialen Suchen die digitalen Archive der Rundfunkanstalten zu durchsuchen.[55] Durch automatisierte Verschlagwortung von Textinhalten und die Digitalisierung audiovisueller Inhalte können Audio-, Video- und Textbeiträge in einer Suche erfasst werden.[56] Dies verdeutlicht den Fortschritt in der Erschließung von Archiven durch KI. Die Recherche von aktuellen Nachrichten profitiert ebenfalls erheblich von KI-Tools wie *Heliograf*, *News Tracer* und *CrowdTangle*. Diese Tools weisen auf Eilmeldungen, virale Beiträge und ungewöhnliche

51 Vgl. Silverman 2016; Silverman et al. 2016
52 Vgl. Potthast et al. 2016
53 Vgl. Zweig et al. 2017
54 Vgl. Heesen 2022
55 Vgl. Heesen et al. 2023
56 Vgl. Maroni et al. 2020

Datentrends hin, wodurch eine schnellere und effizientere Informationsbeschaffung möglich ist. Zudem ermöglicht der Einsatz von KI-Systemen die Analyse großer Datenmengen, die ohne technische Hilfe kaum realisierbar wäre.[57]

Erkennung von Hassrede durch ML-Algorithmen
Die zunehmende Präsenz von Social-Media-Plattformen hat zu einem verstärkten Aufkommen von Hassrede und diskriminierenden Inhalten geführt. In diesem Kontext spielen ML-Algorithmen eine entscheidende Rolle bei der Erkennung und Eindämmung solcher schädlichen Äußerungen. Diese Algorithmen durchlaufen umfassende Trainingsprozesse, bei denen sie mit großen Datensätzen von Fachpersonal als problematisch markierten Inhalten gefüttert werden. Die Algorithmen nutzen Natural Language Processing (NLP)-Modelle, um subtile Nuancen und Kontexte in Texten zu verstehen. Ein Beispiel hierfür ist die automatisierte Erkennung von Schimpfwörtern, rassistischen Begriffen oder beleidigenden Ausdrücken. Die Algorithmen lernen, die Absicht hinter den Wörtern zu verstehen und können so rasch auf Beiträge, die Hassrede enthalten, reagieren. Plattformen wie Facebook und X (vormals Twitter) setzen bereits erfolgreich ML-Algorithmen ein, um die Verbreitung von Hassrede zu minimieren. Mitunter haben ML-Algorithmen jedoch Schwierigkeiten, den Kontext und die Ironie in Äußerungen zu erfassen. Infolgedessen und auch weil aktuell teilweise mehr Inhalte gelöscht werden als juristisch notwendig wäre, kann es vorkommen, dass auch legitime Meinungsäußerungen fälschlicherweise als Hassrede markiert und zensiert werden. Dies ist für die Meinungsfreiheit problematisch und muss durch eine Verbesserung der Algorithmen und der Richtlinien und Prozesse adressiert werden.

Konformitätsprüfung durch ML-Algorithmen
Die Konformitätsprüfung von Inhalten auf Social-Media-Plattformen ist ein komplexes Unterfangen, das ML-Algorithmen unterstützen können. Diese Algorithmen übernehmen die Überprüfung von Inhalten auf Einhaltung von rechtlichen Vorschriften, Urheberrechtsbestimmungen und Community-Richtlinien. Beispielsweise können ML-Modelle automatisch nach urheberrechtlich geschützten Inhalten suchen und diese entsprechend markieren. YouTube setzt beispielsweise *Content ID*[58] ein, ein System, das mithilfe von ML-Algorithmen automatisch urheberrechtlich geschützte Musik und Videos erkennt. Dies ermöglicht den Inhaber:innen geistigen Eigentums, ihre Rechte durch automatische Monetarisierung oder Entfernung von nicht autorisierten Inhalten zu schützen. Darüber hinaus helfen ML-Algorithmen dabei, unzulässige Inhalte wie Gewaltdarstellungen oder explizite sexuelle Inhalte zu identifizieren. Sie können automatisch nach vordefinierten Mustern suchen und Inhalte herausfiltern, die gegen die Nutzungsbedingungen der Plattform verstoßen. Dies erhöht nicht nur die Konformität mit rechtlichen Standards, sondern schafft auch eine sicherere und angenehmere Umgebung für die Nutzer:innen.[59]

Ein etabliertes Instrumentarium in diesem Kontext ist das KI-Tool *KIVI*[60], eingeführt von der Landesanstalt für Medien Nordrhein-Westfalen und implementiert durch die Condat AG aus Berlin. Der Entwicklungsprozess dieses Tools wurde 2020 initiiert, wobei der Schwerpunkt zunächst auf dem Schutz der Menschenwürde und dem Jugendschutz lag. Konkrete Verstoßkategorien umfassen dabei Gewaltdarstellungen, Volksverhetzung, die Verwendung verfassungsfeindlicher Symbole sowie

57 Vgl. Wilczek/Haim 2022
58 Vgl. Neubauer 2014
59 Vgl. Arnold et al. 2019
60 Vgl. Landesanstalt für Medien NRW 2021

frei zugängliche Pornografie. Das KI-Tool ist in der Lage, trainierte Verstoßkategorien sowie die Herkunft eines Verstoßes zu erkennen, beispielsweise indem es aufzeigt, über welche internationalen Kanäle Nutzer:innen auf Deutsch angesprochen werden.

3.2 Generative KI / Synthetische Medienbeiträge

Die Diskussion über den Einsatz Künstlicher Intelligenz (KI) erfährt durch generative KI eine deutlich erhöhte Aufmerksamkeit. In diesem Kontext spielt ChatGPT, als öffentlich zugängliche Pionier-KI, eine bedeutende Rolle. Insbesondere durch die kostenlose Nutzungsmöglichkeit kann jede Person sich aktiv mit der Technologie auseinandersetzen.

Generative KI bezeichnet Systeme, die in der Lage sind, eigenständig Inhalte zu erstellen, sei es Text, Bilder, Audio oder Video. ChatGPT hat die Fähigkeit, auf Anfragen zu antworten und menschenähnlichen Text zu generieren. Die fortschreitende Entwicklung generativer KI hat die Medienlandschaft verändert und bietet vielfältige Möglichkeiten zur Effizienzsteigerung in der Content-Produktion. Während KI in der Lage ist, selbstständig neue Beiträge zu generieren, gibt es jedoch Herausforderungen bezüglich der Zuverlässigkeit der Informationen. Rechtliche Fragestellungen zur Haftung und Verantwortlichkeit in der Nutzung generativer KI entstehen, da es oft unklar ist, wer für durch die KI generierte Inhalte verantwortlich ist und Fragen des Urheberrechts bisher teilweise ungeklärt sind. Die Autonomie und Unvorhersehbarkeit von KI-Ergebnissen stellen zusätzliche Herausforderungen dar, die eine Anpassung bestehender Gesetze erfordern können. Es fällt auf, dass größere Medienhäuser im Vergleich zu kleineren Medienunternehmen zurückhaltender bei der Nutzung von künstlich erzeugten Inhalten sind.[61] Dennoch eröffnen sich für Redaktionen vielfältige Möglichkeiten für die Unterstützung bei repetitiven Aufgaben durch den Einsatz von generativer KI. Einige Unternehmen zeigen bereits die Offenheit und Motivation, diese Technologien zu nutzen, insbesondere bei Aufgaben wie der Transkriptionsarbeit von Audioaufzeichnungen, der Übersetzung von Artikeln und der Audiowiedergabe von Artikeln. Die Norddeutsche NOZ/mh:n Mediengruppe setzt beispielsweise seit 2020 synthetische Audio-Lösungen ein, um Artikel online durch eine synthetisierte, durch KI erstellte Vorlesefassung zu ergänzen. Mehr als 600 Texte wurden als Audioinhalte aufbereitet, wobei automatisierte Untertitel eine breitere Zielgruppe mit unterschiedlichen Konsumpräferenzen bedienen. Darüber hinaus werden generative KI-Technologien wie Speech-to-Text für die Transkription von Audioinhalten und die Moderation von Kommentaren eingesetzt. Diese Anwendungen verdeutlichen den vielfältigen Nutzen von generativer KI in der Redaktionsarbeit und bei der Gestaltung von Medieninhalten.

Nach einem multimedialen Überblick (Kapitel 3.2.1) wird zuerst das Thema Personalisierung und Regionalisierung (Kapitel 3.2.2) und anschließend der Themenkomplex Automatisierung (Kapitel 3.2.2) diskutiert.

3.2.1 Multimedialer Überblick

In diesem Kapitel wird ein Überblick über die multimedialen Möglichkeiten generativer KI gegeben. Unternehmen nutzen diese Technologie vor allem zum Generieren von Text, Bild, Video und Audio. In den folgenden Abschnitten wird aufgezeigt, welche Anwendungen in diesen Bereichen bereits verbreitet sind.

61 Vgl. Hanley/Durumeric 2023

Text-Beiträge
Eine relevante Veränderung durch den vermehrten Einsatz von KI ist bei der Generierung und Moderation von Texten zu beobachten. Insbesondere in den Bereichen Börsen- und Sportnachrichten sowie Wetter- und Verkehrsberichten sind synthetische Beiträge auf dem Vormarsch.[62]

Die Anwendung von KI führt zu Zeitersparnissen bei repetitiven, wenig kreativen Tätigkeiten.[63] Dies ermöglicht eine effizientere Ressourcennutzung und eröffnet Raum für redaktionelle Tätigkeiten höherer Komplexität. Beispielhaft zeigt sich dieser Paradigmenwechsel bei Bloomberg News, wo bereits ein Drittel der redaktionellen Inhalte mithilfe von KI, insbesondere Natural Language Generation (NLG), erzeugt wird.

NLG, wie sie beispielsweise im *GPT-3 Modell* implementiert ist, findet nicht nur in der Content-Erstellung Anwendung, sondern auch in der Moderation von Kommentaren. Die Regulation von Kommentarinhalten, um Hate Speech zu vermeiden, wird durch KI-Modelle wie *Conversario* (u. a. verwendet von der Osnabrücker Zeitung und dem Bayerischen Rundfunk) und *ModBot* (eingesetzt von der Washington Post) optimiert. Diese Modelle ermöglichen nicht nur eine automatisierte Analyse von Kommentaren, sondern auch das Sortieren und Priorisieren derselben.

Bild- und Video-Beiträge
Synthetische Bilder und Videos sind digitale visuelle Inhalte, die mithilfe von KI erstellt werden. Diese Inhalte können Fotos, Bilder, Animationen und Videoclips umfassen, die von KI-Algorithmen generiert werden.

Text-to-Image Models und KI-gestützte Bildbearbeitung
Text-to-Image Models sind KI-Systeme, die in der Lage sind, aus beschreibendem Text digitale Bilder zu generieren. Diese Modelle verwenden neuronale Netze, um Textbeschreibungen in visuelle Darstellungen umzuwandeln. Sie sind besonders nützlich in Bereichen wie Grafikdesign, kreativer Content-Produktion und visueller Kommunikation, da sie die kreative Generierung visueller Inhalte erheblich erleichtern. KI-gestützte Bildbearbeitung hingegen bezieht sich auf den Einsatz von KI, um die Bearbeitung von digitalen Bildern zu automatisieren oder zu optimieren. Diese Technologie kann verwendet werden, um Bilder zu verbessern, Filter anzuwenden, Objekte in Bildern zu erkennen und zu verändern, oder um andere visuelle Effekte zu erzeugen. KI-Tools wie *Adobe Sensei* haben die Bildbearbeitung schneller und effizienter gemacht, indem sie wiederkehrende Aufgaben automatisieren und gleichzeitig die kreative Kontrolle in den Händen der Anwender:innen belassen.

Text-to-Video – AI Video Generation
Die Grundidee hinter der Text-to-Video-Technologie besteht darin, einen effizienten Mechanismus zu schaffen, um Textinformationen automatisiert in visuelle Formate zu übertragen. Dies ermöglicht eine erhebliche Zeitersparnis und neue kreative Möglichkeiten. Unternehmen können somit schnell ansprechende Videos für ihre Produkte oder Dienstleistungen erstellen, ohne aufwendige Produktionsprozesse durchlaufen zu müssen.

In der Werbe- und Marketingbranche hat die Text-to-Video-Technologie einen revolutionären Einfluss. Marken können mühelos personalisierte Werbekampagnen erstellen, indem sie einfachen Text in auffällige visuelle Inhalte umwandeln. Dies ermöglicht eine zielgerichtete Ansprache verschiedener

62 Vgl. Goldhammer et al. 2019
63 Vgl. Heesen 2022

Zielgruppen und eine potenziell bessere Kundenbindung. Die dynamischen Videos, die aus Text generiert werden, können für Social-Media-Plattformen, Webseiten-Banner, E-Mail-Marketing und andere Kanäle verwendet werden, um die Markenpräsenz zu stärken. Synthesia[64] ist eins der Start-ups, das die Videoproduktion durch KI-generierte Avatare umsetzt. Dieses Unternehmen agiert an der Schnittstelle zwischen KI und visueller Kommunikation. Die Technologie ermöglicht authentisch klingende KI-Stimmen in über 120 Sprachen und integriert mehr als 140 KI-Avatare in Videos. Die Videobearbeitung gestaltet sich unkompliziert und ist für alle Personen, auch ohne Erfahrung mit spezieller Software, zugänglich. Synthesia betont dabei, dass ihre Technologie Unternehmen unterstützt, wettbewerbsfähig zu bleiben, indem sie in interaktive Dialoge (via KI-Chatbots) mit ihrer Kundschaft eintreten können. Auf diese Weise können sie Bedürfnisse präzise ermitteln und personalisierte Inhalte bereitstellen.

Im Bildungsbereich entfaltet die Text-to-Video-Technologie eine facettenreiche Anwendbarkeit. Unternehmen können anspruchsvolle Konzepte durch die Erstellung animierter Videos aus Texten einfach veranschaulichen, wodurch der Lernprozess nicht nur interaktiver, sondern auch ansprechender wird. Auch Schulungsmaterialien können einfacher entwickelt werden, indem Unternehmen unkompliziert informative Videos für ihre Mitarbeiter:innen erstellen können. Ein Unternehmen, das diese Technologie zur Verfügung stellt, ist Hour One[65], ein Start-up mit Sitz in Tel Aviv, das diese Technologie in Online-Lernvideos, Geschäftspräsentationen, Nachrichtenberichte und Werbung integriert. Der Erfolg von Hour One manifestiert sich bereits durch die Bereitstellung von Videos für multinationale Unternehmen im Gesundheitswesen sowie für Bildungsunternehmen. Darüber hinaus fungieren ihre Kreationen auch als Grundlage für Nachrichtenaktualisierungen für eine Krypto-Webseite und für Fußballberichte für ein deutsches Fernsehnetzwerk. Hour One zeichnet sich durch die rasche Erstellung professioneller KI-Trainingsvideos aus, indem sie jeden Text in ein hochwertiges, von einer Moderator:in geführtes Video umwandeln, ohne dass dabei umfassende Bearbeitungs- oder Designkenntnisse erforderlich sind. Zeit und Ressourcen werden eingespart, indem mithilfe einer benutzerfreundlichen Plug-and-Play-Schnittstelle Videos mit hyperrealistischen KI-Moderator:innen erstellt werden können. Die chinesische Nachrichtenagentur Xinhua nutzt beispielsweise einen video-generierten Nachrichtensprecher basierend auf einem KI-Avatar.[66] Die BBC experimentiert mit synthetisch erstellten Wettervorhersagen.[67] ProSieben ist mit automatisierten Videos bereits sehr erfolgreich. Ursprünglich für die Darstellung aktueller Corona-Zahlen verwendet, erstreckt sich ihre Anwendung mittlerweile auf Unfallmeldungen, Polizeieinsätze, Rettungsaktionen, Kriminalfälle und Sportergebnisse.[68] Dies unterstreicht den vielseitigen Einsatz automatisierter Videoinhalte und ihre zunehmende Bedeutung in verschiedenen Bereichen. Die Fortschritte in der Entwicklung von Text-to-Video-Technologien, virtuellen Avataren und automatisierter Spracherzeugung haben zweifellos positive Anwendungen, doch es sind auch ernsthafte Bedenken zu möglichem Missbrauch aufgekommen. Eine Kombination dieser Werkzeuge birgt das Potenzial, betrügerischen oder propagandistischen Aktivitäten zusätzlichen Auftrieb zu verleihen. Vor den sich intensivierenden Risiken, die durch die Verschmelzung von Deepfakes, virtuellen Avataren und automatisierter Spracherzeugung entstehen, wird von Wissenschaftler:innen eindringlich gewarnt.[69] In Anbetracht dieser Herausforderungen wird auf die unmittelbare Notwendigkeit hingewiesen, angemessene Sicherheitsmaßnahmen und Richtlinien zu implementieren. Das übergeordnete Ziel besteht darin, sicherzustellen, dass diese Technologie verantwortungsbewusst genutzt wird und nicht für Fälschungen oder Desinformationen missbraucht wird.

64 https://www.synthesia.io/home
65 https://hourone.ai/
66 Vgl. Techvangart 2023
67 Vgl. Rowlatt 2023
68 Vgl. Galileo 2023
69 Kietzmann et al. 2020

Audiobeiträge
Synthetische Audiobeiträge sind künstlich generierte Tonsequenzen, die mithilfe von KI erstellt werden. Diese Beiträge können Sprache, Musik oder Soundeffekte umfassen. Synthetische Audiobeiträge sind das Ergebnis von KI-Algorithmen, die Text in menschenähnliche Stimmen umwandeln oder komplexe musikalische Kompositionen erstellen können.

In den Anfangsphasen der elektronischen Sprachsynthese klangen die Ergebnisse noch sehr roboterhaft und waren teilweise schwer verständlich. Seit Einführung der generativen KI hat sich die Qualität erheblich verbessert, sodass es mitunter herausfordernd ist, diese von menschlichen Sprecher:innen zu unterscheiden. Text-to-Speech-Technologie ermöglicht die Umwandlung von geschriebenem Text in gesprochene Sprache. Populäre Beispiele sind Sprachassistenten wie *Amazon Alexa* oder *Google Assistant*. Diese KI-gesteuerten Systeme verwenden synthetische Stimmen, um Antworten auf Anfragen von Nutzer:innen zu generieren. Ein weiteres Beispiel ist die Vorlesefunktion in E-Book-Readern, die Texte in gesprochene Worte umwandelt und somit unter anderem Menschen mit Sehbehinderungen den Zugang zu Büchern und anderen geschriebenen Inhalten erleichtert.

Im Gegensatz zur Text-to-Speech-Technologie konzentriert sich die Speech-to-Text-Technologie auf die erweiterte Spracherkennung. Speech-to-Text ist eine KI-gesteuerte Technologie, die gesprochene Sprache von einer analogen in eine digitale Form übersetzt und diese dann in Text umwandelt. Obwohl die automatische Spracherkennung bereits vor langer Zeit entwickelt wurde, fand sie lange Zeit kaum Anwendung. Inzwischen haben jedoch Fortschritte in der Technologie, insbesondere durch den Einsatz von KI, dazu geführt, dass die Audiotranskription erhebliche Verbesserungen erfahren hat. Dies ermöglicht schnelle und präzise Ergebnisse. Bekannte Anwendungen wie TikTok, Spotify und Zoom haben diese Technologie in ihre mobilen Apps integriert. Die Vorteile dieser Technologie liegen in der verbesserten Zugänglichkeit von multimedialen Inhalten sowie der Steigerung der Effizienz und Barrierefreiheit in der Produktion von audiovisuellen Medienbeiträgen. Im Bereich der audiovisuellen Medien wird bereits versucht, durch die Aufzeichnung der Lippenbewegungen des Sprechenden mithilfe einer Videokamera und deren Kombination mit akustischer Erkennung, automatische Untertitel bei Aufnahmen hinzuzufügen. Trotz der Anpassungsfähigkeit von Spracherkennungsprogrammen an Hochsprachen stoßen sie häufig an ihre Grenzen, insbesondere wenn es um Dialekte und Soziolekte geht. Im Gegensatz zu Menschen, die sich schnell an unterschiedliche Mundarten anpassen können, erfordert es für Spracherkennungssoftware aufwändige Prozesse, um mit Dialekten umzugehen.[70]

3.2.2 Personalisierung und Regionalisierung

Die Medienlandschaft erfährt maßgebliche Veränderungen durch die Fortschritte in generativer KI, insbesondere im Bereich der Personalisierung und Regionalisierung. Innovative Technologien ermöglichen eine präzise Anpassung von Inhalten an individuelle Präferenzen und lokale Gegebenheiten, wodurch ein bedeutsamer Schritt in Richtung einer verbesserten und individualisierten Nutzererfahrung getan wird. Vergangene Erfahrungen haben jedoch auch gezeigt, dass die Personalisierung von Inhalten durch algorithmische Empfehlungssysteme auch Herausforderungen mit sich bringt. Eine exzessive Anpassung an individuelle Vorlieben birgt das Risiko der Entstehung von Filterblasen, in denen Nutzer:innen nur Informationen und Standpunkte präsentiert werden, die ihren bestehenden Überzeugungen entsprechen. Dies könnte dazu führen, dass Nutzer:innen in einer begrenzten Informationsumgebung verweilen, die Diversität ausschließt. Auf den meisten

70 Vgl. Schoenert 2012

Plattformen können Nutzer:innen nicht direkt entscheiden, welche Informationen ihnen präsentiert werden. Ihre indirekte Beeinflussung erfolgt stattdessen durch bewusste Entscheidungen bei Interaktionen, wie der gezielten Suche nach Inhalten, dem Abonnieren oder Abbestellen von Seiten und dem Bewerten von Beiträgen. Diese Handlungen werden von Algorithmen berücksichtigt, um personalisierte Empfehlungen zu generieren. Personalisierung und Regionalisierung können viele positive Auswirkungen haben, insbesondere wenn die Inhalte konstant bleiben, sich aber an die spezifischen Bedürfnisse der Nutzer:innen in Bezug auf Sprache, Komplexität oder Länge anpassen. Dies trägt zu einer verbesserten Nutzererfahrung, effektiveren Informationsübermittlung und einer stärkeren Einbindung der Zielgruppe bei. Die gezielte Anpassung an regionale Besonderheiten fördert zudem kulturelle Vielfalt und lokale Identifikation. Trotz dieser positiven Aspekte sind jedoch Bedenken hinsichtlich des Datenschutzes und ethischer Fragestellungen zu berücksichtigen. Der Einsatz sensibler Daten für personalisierte Inhalte birgt das Risiko von Privatsphärenverletzungen und erfordert eine kritische Betrachtung.

Personalisierte Werbung
Aktuell zeichnet sich ab, dass Technologiegiganten wie Alphabet und Meta die Integration generativer KI in ihre Werbeanzeigenplattformen planen.[71] Generative KI wird derzeit als Instrument für Werbetreibende beworben, um visuell ansprechende Elemente zu erstellen.[72] Eine mögliche Weiterentwicklung besteht darin, dass personalisierte Werbung, unterstützt durch generative KI, direkt von den Plattformen anstelle von Werbetreibenden basierend auf deren Vorgaben generiert wird. Ein bemerkenswertes Beispiel ist Carvana, ein Gebrauchtwagenhändler, der eine umfassende personalisierte Werbekampagne mit generativer KI durchgeführt hat und 1,3 Millionen individuelle Videos für Kunden erstellt hat, inklusive KI-generierter Synchronsprecherstimmen, Animationen des gekauften Fahrzeugmodells und relevanter Ereignisse zum Kaufzeitpunkt.[73] Die Qualität der Videos kann noch nicht mit professionellen Werbevideos konkurrieren, es wird jedoch erwartet, dass sich die Unterschiede in kurzer Zeit verringern werden. In einem anderen Szenario schuf ein Fotograf einen KI-generierten Influencer basierend auf einfachen Illustrationen, was die Erzeugung vieler Bilder ermöglicht, die wie derselbe Charakter aussehen.[74] Diese KI-generierten virtuellen Persönlichkeiten könnten in Zukunft für personalisierte Werbung genutzt werden, die auf den individuellen Vorlieben der Nutzer:innen basiert. Trotz der offensichtlichen Vorteile generativer KI in der personalisierten Werbung sind Bedenken hinsichtlich des Datenschutzes und Manipulation zu berücksichtigen. Während personalisierte Anzeigen bereits auf Nutzerdaten für eine gezielte Ausrichtung basieren, birgt die Möglichkeit, dass Plattformbetreiber personalisierte Anzeigeninhalte generieren, das Risiko der Verwendung sensibler Informationen zur Manipulation von Nutzer:innen.

AI Audio-Dubbing and Voice Cloning
Die Praxis des Audio-Dubbings, bei der bestehende Audioinhalte durch neue Tonspuren ersetzt oder erweitert werden, findet in der Filmproduktion, Fernsehsendungen und anderen audiovisuellen Medienanwendungen breite Anwendung. Konventionelles Audio-Dubbing strebt an, Originalsprachenaufnahmen zu übersetzen, den Ton zu optimieren oder anzupassen sowie die ursprünglichen Aufnahmen durch lokalisierte Versionen zu ersetzen. Die Einführung von KI im Bereich des Audio-Dubbings ermöglicht hochwertige Echtzeit-Synchronisationen in verschiedenen Sprachen, was bis zum Voice Cloning reicht. Voice Cloning ermöglicht die Wiederverwendung des originalen Stimmklangs zur Generierung neuer Inhalte. Ein führendes Unternehmen auf diesem Gebiet ist ElevenLabs,

71 Vgl. Mehta 2023a; Murphy/Criddle 2023
72 Vgl. Mehta 2023b
73 Vgl. Thwaites 2023
74 Vgl. Growcoot 2023; Zhang/Agrawala 2023

das mit Hilfe seiner KI-Kompetenzen Klassiker wie *Dinner for One* in zahlreiche Sprachen übersetzt und dabei die Originalstimmen der Darsteller:innen klonen kann.[75] Diese innovative Herangehensweise erstreckt sich auch auf die Wiederbelebung weiterer Klassiker, wie im Falle des neuen Pumucklfilms, bei dem auf RTL+ zwischen den beiden Stimmversionen von Hans Clarin und Maximilian Schafroth gewählt werden kann.[76] Die Vorteile von AI-Dubbing liegen in der Möglichkeit, Audioinhalte an verschiedene Zielgruppen oder Regionen anzupassen, was eine erhöhte Personalisierung ermöglicht und Informationen einem breiteren Publikum zugänglich macht. Für langjährige Fans von Pumuckl mag die Möglichkeit, den Kobold mit der vertrauten Stimme zu hören, als Bereicherung empfunden werden. Allerdings werfen die Einsatzmöglichkeiten von Voice Cloning ethische Fragen auf, insbesondere hinsichtlich der Schwierigkeiten für Sprecher:innen, sich zu etablieren, wenn ihre stimmliche Arbeit ausschließlich dazu dient, die Stimme verstorbener Schauspieler:innen zu reproduzieren. Weitere Nachteile bestehen darin, dass die Voice Cloning-Technologie die Verbreitung von Fälschungen begünstigen kann, was zu Manipulation und Desinformation in audiovisuellen Medien beitragen kann. Realistisch wirkende Audiomanipulationen, nutzen Technologien wie die Lippen-Synchronisation, um Mundbewegungen von Schauspieler:innen in der Originalversion zu analysieren und synthetische Lippenbewegungen für die Fremdsprachenversion zu erzeugen. Erste Demonstrationen dieser synthetischen Medienproduktion zeigen lippensynchrone Filmausschnitte von einem japanisch sprechenden Tom Hanks und einem deutschsprachigen Robert de Niro und geben einen Ausblick auf die fortlaufende Entwicklung und Forschung im Bereich des AI-Dubbings.[77]

75 https://elevenlabs.io/dubbin
76 Vgl. Angrick 2023
77 Vgl. Flawless AI 2023

Veränderung der Medienlandschaft durch KI – Chancen und Risiken

Fallbeispiel: KI im Radio

Entwicklung
Das Medium Radio hat in der Transformation der Medienlandschaft verschiedene Entwicklungsstufen durchlaufen. Von den traditionellen UKW-Frequenzen (FM) bis hin zu Radio-Streams im Internet hat die technologische Entwicklung das Radio in die digitale Ära geführt. Dies hat nicht nur eine Ausweitung der Übertragungswege mit sich gebracht, sondern auch Veränderungen in der Art und Weise, wie Inhalte erstellt und präsentiert werden. Ein Merkmal dieser Veränderungen ist die fortschreitende Integration von KI in den Radiobetrieb. Während Radio-Streams im Internet bereits eine erhebliche Diversifizierung und Globalisierung der Musik und Inhalte ermöglichen, könnte sich das Medium Radio in Richtung einer engeren Verbindung mit KI entwickeln. KI-Technologien, die in der Lage sind, Inhalte zu generieren, kuratieren und moderieren, eröffnen neue Horizonte für Radiosender und Hörer:innen gleichermaßen. Diese Entwicklung verspricht eine noch nie dagewesene Personalisierung. KI kann nicht nur musikalische Vorlieben der Hörer:innen analysieren und passende Songs auswählen, sondern auch synthetische Audio-Beiträge erstellen, die nahtlos in den Radiosendebetrieb integriert werden könnten.

Radiosendungen und KI – Wie hoch ist der Einfluss generativer KI bereits?
Einige Radiosender experimentieren mit der Möglichkeit, generativen KI-Systemen die Verantwortung für komplette Sendungen zu übertragen. Hierbei verfolgen sie zwei unterschiedliche Ansätze: Zum einen entwickeln sie eigene künstliche Charaktere, wie es beispielsweise Antenne Deutschland mit dem Radiostream *kAI* getan hat. Zum anderen versuchen einige Sender, etablierte Radiopersönlichkeiten durch KI zu ersetzen, indem sie bestehende Daten verwenden, um völlig neue Sendungsinhalte zu generieren. Die Ergebnisse dieser Experimente variieren und die Radiosender ziehen unterschiedliche Schlussfolgerungen aus diesen Erfahrungen.

Die Sendergruppe Absolut Radio der Antenne Deutschland GmbH & Co. KG hat mithilfe der Text-to-Speech-Technologie auf dem *Stream Absolut Radio AI* einen Radiosender gestartet, der vollständig von der KI kAI moderiert wird.[78] Dieser Stream ist über die Absolut Radio App, die Webseite und gängige Streaming-Portale verfügbar. kAI fungiert nicht nur als Moderator, sondern hat auch die Aufgabe, Menschen behutsam an das Thema KI heranzuführen und deren Anwendungen und Vorteile zu erläutern. Angesprochen wird die Zielgruppe von 14- bis 49-Jährigen. Antenne Deutschland betrachtet dies als bedeutsamen Schritt für die deutsche Radiolandschaft. Laut Mirko Drenger, CEO der Antenne Deutschland GmbH & Co. KG, wird durch die Einführung von KI in der Radiomoderation niemand seinen Arbeitsplatz verlieren, sondern die Arbeitsplätze werden sich verändern und effizienter gestaltet. Die KI soll kontinuierlich weiterentwickelt werden, um qualitativ hochwertige Programme sicherzustellen. Darüber hinaus werden KI-generierte Musikstücke in das Programm integriert. Das Ziel besteht darin, den Stream mit weiteren Erkenntnissen zur KI auszubauen und neue Möglichkeiten in der Programmgestaltung zu erkunden.

78 Vgl. Antenne Deutschland 2023

Ein weiteres Experiment wurde im Rahmen eines Forschungsprojekts des WDR *Innovation Hub* durchgeführt. Beiträge der WDR 2-Moderatorin Steffi Neu wurden mithilfe von KI synthetisch hergestellt. Auch der Radiosender SWR3 führte ein ähnliches Experiment durch, bei dem KI einen ganzen Tag lang Beiträge moderierte. Das KI-Stimmenmodell wurde mittels Voice Cloning auf Basis von Tonaufnahmen von Volker Janitz, einem der Moderatoren des Senders, trainiert. Diese Experimente illustrieren, welche Fragen man sich zur Zukunft der Radiomoderation und den Chancen und Risiken synthetischer Medien stellen kann.

Die Integration von KI-Stimmen markiert einen Trend in der Radiobranche und trägt zur Transformation dieses Sektors bei. Dieser Ansatz bietet eine Vielzahl von Vorzügen, darunter die kontinuierliche Content-Produktion, da KI-Moderatoren ununterbrochen einsatzbereit sind, was insbesondere für den Nachtdienst und Eilmeldungen von Bedeutung ist. Darüber hinaus sind KI-Stimmen in der Lage, Inhalte in verschiedenen Sprachen und Dialekten bereitzustellen, die von menschlichen Moderator:innen eventuell nicht abgedeckt werden, was die Reichweite der Sender erweitern kann. Zudem bieten synthetische Stimmen eine effiziente Lösung für Radiospots.

Werbung im Radio – RadioGong AdMaker

Die Einführung des *RadioAdMaker*[79] hat eine bedeutsame Veränderung in der Radiowerbung herbeigeführt, indem er die bislang zeit- und kostenintensive Herstellung von Werbespots neu gestaltet hat. Mit der Verwendung von generativen KI-Technologien haben Werbetreibende die Möglichkeit, Radiowerbespots in kürzester Zeit zu erstellen und zu buchen. Radiosender können so kosteneffizient Werbung schalten und besser mit etablierten Größen wie Facebook und Spotify konkurrieren. Bei der Spot-Produktion können Nutzer:innen aus einer Auswahl an Stimmen und Musikstücken wählen, wobei der Preis je nach gewünschter Reichweite variiert. Die Kosten für KI-generierte Spots sind im Vergleich zu herkömmlich produzierten Werbespots gering.

Derzeit dürfen KI-generierte Spots noch nicht im Hauptprogramm von Radiosendern ausgestrahlt werden, da die Qualität der Sprachsynthese noch nicht das gewünschte Niveau erreicht hat. Stattdessen werden sie als Pre-Streams für das Webradio verwendet, mit der Möglichkeit, in absehbarer Zeit in das Hauptprogramm integriert zu werden. Der RadioAdMaker unterstützt momentan ausschließlich Hochdeutsch, allerdings sind zukünftige Erweiterungen geplant, um auch Dialekte zu integrieren. Besonders Unternehmen, die zuvor nicht über die Ressourcen für die Produktion von Radiowerbespots verfügten, können von dieser Innovation profitieren.

KI-generierte Musik im Radio

Die Verwendung von KI im Musiksektor, exemplarisch durch das Programm *Music-Gen* von Meta illustriert, hat in den letzten Jahren enorme Fortschritte gemacht. Diese Technologie ermöglicht es, maßgeschneiderte Musikstücke zu generieren und bietet Radiosendern die Möglichkeit, sich in einem hart umkämpften Markt zu differenzieren. Laut Dr. Esther Fee Feichtner, Leiterin des Digitalisierungskollegs Artificial Intelligence in Culture and Arts (AICA), können Radiosender durch den Einsatz von KI-generierter Musik einen Wettbewerbsvorteil erzielen. Eine zentrale Chance bestehe darin, auf individuelle Musik zu setzen und personalisierte Inhalte für Hörer:innen zu schaffen.

79 https://radioadmaker.de/

Traditionell konzentrierte sich Radio auf Massenmusik. KI ermöglicht eine personalisierte Musikauswahl, die die emotionale Bindung vertiefen und damit mehr Loyalität bei den Hörer:innen erzeugen kann. Allerdings gibt es Bedenken, dass KI-Systeme uninspirierte Durchschnittsmusik produzieren könnten, wenn sie sich zu stark auf etablierte Muster verlassen. Doch KI kann auch als Werkzeug dienen, um kreative Neukompositionen zu fördern und die Musiklandschaft aufregender und vielfältiger zu gestalten. Zusätzlich ermöglicht KI die Restaurierung historischer Aufnahmen und kulturellen Erbes, was Radiosendern die Präsentation seltener und kulturell bedeutsamer Musikstücke in höchster Qualität ermöglicht. Die Integration von KI-generierter Musik im Radio bietet also vielfältige Perspektiven und Chancen. Radiosender müssen sicherstellen, dass KI nicht die kulturelle Vielfalt reduziert, sondern als Werkzeug zur Förderung von Innovation und Bereicherung genutzt wird. Die verantwortungsvolle Nutzung dieser Technologie kann die Musiklandschaft inspirieren und erweitern.[80]

In Bezug auf das Urheberrecht ist es von entscheidender Bedeutung, dass bei der Verwendung von KI-generierten Musikstücken im Radio die entsprechenden rechtlichen Rahmenbedingungen und Lizenzvereinbarungen eingehalten werden. Die Anerkennung und angemessene Vergütung der Künstler:innen und Rechteinhaber:innen müssen gewährleistet sein, um die nachhaltige Entwicklung der Musikindustrie zu fördern und einen fairen Umgang mit den kreativen Schöpfenden zu gewährleisten.

Wie sieht die Zukunft des Radios aus?

KI und digitale Übertragung eröffnen dem Radio neue Möglichkeiten, die über die traditionelle Radiolandschaft hinausgehen. Die Zukunft des Radios wird lebhaft diskutiert, und obwohl einige das Ende dieses Mediums vorhergesagt haben, bleibt das Radio nach wie vor ein wesentlicher Bestandteil des Alltags vieler Menschen.

Mit den Fortschritten im Bereich der KI und der digitalen Übertragung ist es möglich, personalisierte Radioprogramme anzubieten, die sowohl von menschlichen Moderator:innen als auch von KI-Systemen präsentiert werden können. Vollautomatische KI-Radiosysteme wie Radio-GPT befinden sich bereits in der Testphase und könnten den Bedarf an menschlichen Moderator:innen weitgehend überflüssig machen, insbesondere bei der Musikauswahl.

Dennoch besteht die Herausforderung darin, dass das Radio in einer Zeit, in der Podcasts und On-Demand-Inhalte immer beliebter werden, relevant bleibt. Ein Ansatz für die Zukunft könnte sein, eine gemeinsame Plattform anzubieten, auf der Hörer:innen aus verschiedenen Anbietern auswählen können. Die Plattform kann personalisierte Programme basierend auf den Vorlieben der Hörer:innen anbieten und sich kontinuierlich weiterentwickeln, um deren Interessen besser zu verstehen.

KI könnte dabei eine wichtige Rolle spielen, indem sie Inhalte kuratiert und personalisierte Empfehlungen gibt. Dennoch sollte das Radio seinen besonderen Wert, nämlich die Verbindung zwischen Hörer:innen und Moderator:innen, aufrechterhalten. Live-Radio wird daher voraussichtlich weiterhin eine Rolle spielen.[81]

80 Vgl. Haase 202
81 Vgl. Beck 2023

Fazit – KI im Medium Radio
Die Nutzung von KI im Medium Radio bietet eine Vielzahl von Potenzialen. KI ermöglicht es, aus menschlichen Sprachbausteinen neue Beiträge zu generieren. Radiosender können so Inhalte schneller und effizienter erstellen, indem sie auf vorherige Aufzeichnungen und Datenbanken zurückgreifen, um relevante und ansprechende Beiträge zu gestalten. Dies trägt dazu bei, die Produktionskosten zu senken und die Aktualität der Inhalte zu gewährleisten.

KI-Agenten
Die Integration generativer KI in die Medienlandschaft erreicht eine neue Dimension durch virtuelle KI-Agenten. KI-Agenten sind Computerprogramme oder Systeme, die mithilfe von KI Aufgaben ausführen können, die normalerweise menschliche Intelligenz erfordern. Sie sind in der Lage, Informationen zu verarbeiten, zu lernen, Muster zu erkennen und Entscheidungen zu treffen. KI-Agenten können in verschiedenen Formen auftreten, darunter Chatbots, digitale Assistenten, virtuelle Avatare und andere Arten von automatisierten Systemen. Ihr Zweck kann von einfachen Aufgaben wie dem Beantworten von Fragen oder dem Ausführen von Befehlen bis hin zu komplexeren Aufgaben wie der Personalisierung von Benutzererfahrungen, der Analyse großer Datensätze oder der Durchführung autonomer Handlungen reichen. KI-Agenten könnten insbesondere für Unternehmen von Interesse sein, die ihre Online-Benutzererfahrung verbessern möchten. Da KI-Agenten typischerweise unsichtbare und nicht-physische Entitäten sind, die im Hintergrund agieren und keine menschenähnliche Form besitzen, hat die Einführung von Digital Humans in Kombination mit KI-Agenten an Bedeutung gewonnen. Digital Humans sind virtuelle Darstellungen von menschenähnlichen Figuren oder Avataren, die mit fortschrittlichen Technologien wie Computergrafik, Animation und, in einigen Fällen, KI, erstellt werden (siehe auch Kapitel 3.2.1). Die Integration von Digital Humans könnte die Online-Benutzererfahrung weiter verbessern, indem sie eine menschenähnliche Interaktion ermöglichen. Im Gegensatz dazu sind KI-Agenten eher auf Hintergrundoperationen spezialisiert und bieten weniger visuelle und interaktive Elemente.

Es gibt bereits einige Unternehmen, die KI-Agenten mit menschenähnlichen Avataren nutzen. D-ID[82] ermöglicht beispielsweise die Verbesserung von Kundenerlebnissen durch digitale Vertreter:innen, die interaktive KI-Avatare in Webseiten einbinden und maßgeschneiderte Tutor:innen für Online-Kurse bereitstellen. Der Übergang von grafischen Benutzeroberflächen zu natürlichen Benutzerschnittstellen, die Gesten und Sprache nutzen, stellt eine qualitative Veränderung der Kundeninteraktion dar. Diese einfacheren und intuitiveren Bedienmöglichkeiten fördern eine nutzerfreundliche Erfahrung, besonders auf mobilen Geräten, und machen das Erlernen der Bedienung leichter. Uneeqs *Digital Humans*[83] eröffnen innovative Perspektiven im Bereich dialogorientierter Kommunikation, indem sie sich darauf konzentrieren, während des Besuchs der Webseite den Nutzer:innen individuell angepasste Informationen bereitzustellen. Im Gegensatz zu monologischen Inhalten setzen sie auf einen interaktiven Dialog, der durch fundiertes Branchenwissen unterstützt wird. Dies ermöglicht eine präzise Anpassung an die spezifischen Anforderungen der Nutzer, wodurch personalisierte KI immersive Erlebnisse schafft, verglichen zu herkömmlichen Webseiten.

Jedoch sind auch bei diesen Anwendungen Aspekte hinsichtlich des Datenschutzes sowie ethischer Überlegungen zu berücksichtigen. Die Verwendung personalisierter KI, insbesondere in einem interaktiven Kontext, birgt potenzielle Risiken bezüglich des Schutzes der Daten von Nutzer:innen und der Wahrung ethischer Prinzipien. Es ist von entscheidender Bedeutung, dass Entwickler und

82 https://www.d-id.com
83 https://www.digitalhumans.com/

Betreiber solcher Systeme sicherstellen, dass Datenschutzrichtlinien strikt eingehalten werden und die Interaktionen ethisch verantwortungsbewusst gestaltet sind, um das Vertrauen der Nutzer:innen zu wahren und etwaige negative Auswirkungen zu minimieren.

3.2.3 Automatisierung

KI ermöglicht es, zunehmend realistische Texte, Bilder und sogar Stimmen zu generieren, was weitreichende Auswirkungen auf automatisierte Prozesse hat. Durch die Fähigkeit, komplexe Muster zu erkennen und kreativ zu agieren, verändert generative KI die Art und Weise, wie Aufgaben automatisiert werden. Die Möglichkeiten reichen von der personalisierten Content-Erstellung bis hin zur Simulation menschenähnlicher Interaktionen. Der Einsatz von generativer KI in der Automatisierung verspricht nicht nur Zeit- und Ressourceneinsparungen, sondern eröffnet auch neue Horizonte für kreative Anwendungen in verschiedenen Branchen.

Voice Cloning zur Automatisierung
Die Fähigkeit zur synthetischen Stimmerzeugung, auch als Voice Cloning bekannt, bezieht sich auf die künstliche Nachbildung der Stimme einer Person. In Kapitel 3.2. wurde bereits eingehend auf die Thematik des Voice Cloning eingegangen. Die Voice Cloning Technologie hat sich stetig weiterentwickelt und erzielt mittlerweile äußerst präzise Ergebnisse.[84]

Ein Beispiel für den Einsatz von Voice Cloning und die damit verbundenen Automatisierungseffekte liefert Spotify. Das Unternehmen hat ein Tool entwickelt, das die von OpenAI veröffentlichte Sprachgenerierungstechnologie integriert. Zum Beispiel kann eine Podcast-Episode, die ursprünglich auf Englisch aufgenommen wurde, nun automatisiert in andere Sprachen übersetzt werden, wobei die charakteristischen Sprachmerkmale der Originalsprecher:in erhalten bleiben. Im Rahmen dieses Projekts hat Spotify eng mit verschiedenen Podcaster:innen zusammengearbeitet, um KI-gesteuerte Sprachübersetzungen in andere Sprachen zu generieren, darunter Spanisch, Französisch und Deutsch. Diese Automatisierung ermöglicht nicht nur eine breitere internationale Verfügbarkeit von Inhalten, sondern eröffnet auch die Möglichkeit für Hörer:innen weltweit, neue Podcaster:innen zu entdecken.[85]

Dies ist nur eine exemplarische Anwendung von Voice Cloning, mit der fortschreitenden Entwicklung der Technologie werden weitere entstehen. Ein ethischer Umgang bei der Synthetisierung von Stimmen muss dabei immer berücksichtigt werden. Eine Standardisierung des Genehmigungsprozesses wäre eine Möglichkeit, um die Stimmen aller zu schützen und sicherzustellen, dass jede Person die vollständige Kontrolle darüber hat, wie die eigene Stimme verwendet wird.[86]

Prosoziale Chatbots
In der Ära der digitalen Kommunikation und sozialen Medien stehen Plattformen und Anwender:innen vor der Herausforderung, mit einer Vielzahl von Inhalten und Interaktionen umzugehen. In diesem Kontext gewinnen prosoziale Chatbots zunehmend an Bedeutung. Diese KIs, die darauf ausgelegt sind, soziale Interaktionen zu unterstützen und zu verbessern, bieten ein vielversprechendes Potenzial für soziale Medien. Desinformation und Hassreden stellen nach wie vor große Herausforderungen in sozialen Medien dar. Während Beiträge, die gegen die Content-Richtlinien

84 Vgl. Moh 2023
85 Spotify 2023
86 Vgl. Baker 2023

der Plattformen verstoßen, entfernt werden, bleiben oft Grauzonen oder Randbemerkungen, die nicht entfernt werden. Eine Möglichkeit, auf solche Beiträge zu reagieren, ist die Konfliktmediation. Es gibt verschiedene bekannte Techniken, um das Konfliktpotenzial zwischen Menschen, die unterschiedlicher Meinung sind, zu reduzieren. Ein Online-Experiment ergab, dass die Verwendung von Chatbots zur Neuformulierung von Nachrichten in einer Diskussion über Waffenkontrolle die politische Spaltung verringerte: Die Menschen fühlten sich verstandener und respektierter, wenn sie Nachrichten erhielten, die mit Hilfe von Chatbots umformuliert wurden.[87]

Auch im Bereich der Beitragsmoderation wird generative KI verwendet. Wenn ein Beitrag in sozialen Medien gegen die Regeln verstößt, wird er oft mit wenig oder keiner Erklärung entfernt. Einige Plattformen ermöglichen Benutzer:innen, gegen diese Entscheidungen Einspruch zu erheben, aber dieser Prozess kann lange dauern und ist für die Plattform kostspielig. Ein Chatbot könnte erklären, warum ein Beitrag gegen die Regeln verstößt, sogar vor der Veröffentlichung, und so den Erstellenden die Möglichkeit geben, Beiträge entsprechend anzupassen. Ein Dialog mit dem Chatbot würde den Benutzer:innen dabei helfen, die Content-Richtlinien besser zu verstehen.

Ein weiterer Anwendungsfall für prosoziale Chatbots besteht darin, faktische Fragen zu beantworten. Wenn zwei Personen eine faktische Frage debattieren, etwa die Energiekosten von Elektro- und Benzin-Fahrzeugen, könnte ein Bot die Forschung zu diesem Thema finden und zusammenfassen. Dies ähnelt den Werkzeugen, die bereits auf Plattformen wie Skype und Slack zu finden sind, die beide OpenAI's Sprachmodelle verwenden, um die Fragen der Benutzer:innen zu beantworten.[88] Obwohl diese Werkzeuge auf persönliche Gespräche oder Unternehmensanfragen ausgerichtet sind („An wen in der Personalabteilung sollte ich mich wegen eines arbeitsbezogenen Problems wenden?"), könnten sie genauso gut auf Social-Media-Plattformen übertragen werden.

Social-Media-Unternehmen forschen darüber hinaus auch an einer Schreibunterstützung für Benutzer:innen mit Chatbots. LinkedIn etwa hat KI-basierte Artikelvorschläge, Benutzerprofile und Jobbeschreibungen integriert.[89] Der X (vormals Twitter) Konkurrent Koo hat eine Funktion zur Erstellung von Beiträgen mithilfe von ChatGPT eingeführt. Solche Werkzeuge können insbesondere für Personen nützlich sein, die in Sprachen schreiben, die sie nicht sicher beherrschen, gleichzeitig erleichtern sie aber auch die Erstellung von Spam und könnten zur Verbreitung von minderwertigen Inhalten auf Plattformen führen.

Überblick Anbieter

In Kapitel 3.1 und 3.2 wurden zahlreiche Anwendungen von KI im Medienbereich sowie die damit verbundenen Chancen und Risiken dargestellt. Tabelle 1 gibt einen Überblick darüber welche Systeme und Anbieter derzeit auf dem Markt sind (Stand Januar 2024). Neben den etablierten Unternehmen, wie etwa Microsoft und Google, gibt es auch zahlreiche Start-ups, die sich in Marktnischen platzieren.

87 Vgl. Argyle et al. 2023
88 Vgl. Lardinois 2023; Salesforce 2023
89 Vgl. Clark 2023; Sato 2023

Auswahl bestehender Systeme auf Basis generativer KI		
Anwendung	**System**	**Anbieter**
Textgenerierung	ChatGPT Bard Claude ChatFlash LLaMa	OpenAI Google Anthropic neuroflash Meta
Text-to-Image	Midjourney Stable Diffusion DALL-E Adobe Firefly	Midjourney Stability AI OpenAI Adobe
Text-to-Video	Synthesia Hour One ModelScope Jasper	Synthesia Hour One AliBaba Jasper
KI-Agents	UneeQ D-ID	UneeQ D-ID
Audiogenerierung	VALL-E ElevenLabs CloneDub LyreDub Resemble	Microsoft ElevenLabs CloneDub descript Resemble
Prosoziale Chatbots	Captionit.ai	Captionit.ai

Tabelle 1: Systeme basierend auf generativer KI mit jeweiligen Anbietern. Quelle: Eigene Darstellung

3.3 Risiko Desinformation

Wie zu Beginn des Kapitels dargestellt, hat sich die Medienlandschaft infolge der Integration von Künstlicher Intelligenz (KI) nachhaltig gewandelt. In den vergangenen Jahren hat eine signifikante Veränderung im Medienkonsum stattgefunden. Im EU-Durchschnitt verzeichnete die Nutzung sozialer Medien einen stetigen Anstieg, während die Präferenz für gedruckte Zeitungen rückläufig ist.[90] Besondere Aufmerksamkeit gilt dabei dem Thema Desinformation, da diese unter anderem die für demokratische Gesellschaften wichtige Meinungsbildung im Umfeld von demokratischen Wahlen beeinflussen kann. Diese Bedrohung ist aufgrund fehlender Nachrichtenintermediäre insbesondere auf digitalen Plattformen ausgeprägt. KI erweist sich nicht nur als Instrument zur erleichterten und schnelleren Erstellung überzeugender Desinformation, sondern auch zur beschleunigten und weitreichenden Verbreitung derselben.

90 Vgl. European Commission 2023b

Das nachfolgende Kapitel erörtert eingehend das Themenfeld der Desinformation. Nach einer Definition der relevanten Begrifflichkeiten (Kapitel 3.3.1) werden die Mechanismen der Erstellung (Kapitel 3.3.2) sowie der Verbreitung von Desinformation (Kapitel 3.3.3) dargestellt und erste Erkenntnisse über die Effekte von Desinformation eingeordnet (Kapitel 3.3.4).

3.3.1 Definition

Die Termini Misinformation und Desinformation sind eng mit dem Begriff Fake News assoziiert, wobei insbesondere Donald Trump maßgeblichen Einfluss auf die Verbreitung letzteren Begriffs in der medialen Berichterstattung genommen hat.[91] Diese Begriffe sollten jedoch klar unterschieden werden. Fake News stellt ein Schlagwort dar, das aktuell, wie im Fall von Donald Trump, häufig dazu verwendet wird, die Legitimität und Wahrhaftigkeit seriöser Medienquellen in Frage zu stellen.[92] Diese Verwendung ist als politisch geprägter Kampfbegriff zu verstehen, dessen Anwendung auf diese spezifische Bedeutung beschränkt werden sollte.[93]

Im Gegensatz dazu herrscht weitgehender Konsens in der Definition der Begriffe Desinformation und Misinformation dahingehend, dass die vermittelte Information nicht der Wahrheit entspricht, wenngleich der wissenschaftliche Diskurs zur präzisen Definition beider Begriffe noch nicht abgeschlossen ist.[94] Ein entscheidender Aspekt bei der Differenzierung liegt auf der Intention der Informationsgeber:in. In diesem Gutachten sprechen wir bei unbeabsichtigter Verbreitung falscher Informationen von Misinformation, während bei einer bewussten Täuschungsabsicht der Begriff Desinformation Anwendung findet.[95]

Die vorliegendwe Definition erscheint bei oberflächlicher Betrachtung plausibel, weist jedoch inhärente Probleme auf. Die Intention der Informationsgeber:in kann nicht immer zweifelsfrei ermittelt werden. Es besteht die Möglichkeit, dass fehlerhafte Informationen beispielsweise in sozialen Netzwerken unwissentlich weiterverbreitet werden, ohne dass die betreffende Person sich der Falschheit bewusst ist, selbst wenn der oder die Urheber:in dieser Falschinformationen diese absichtlich erstellt hat. Eine zusätzliche Fragestellung in diesem Zusammenhang betrifft die Differenzierung zwischen Wahrheit und Unwahrheit, da dies häufig nicht objektiv beantwortet werden kann.[96]

Insbesondere im Kontext von Verschwörungstheorien sind die Anhänger:innen in der Regel von ihren Überzeugungen geprägt, wodurch subjektive Wahrheiten entstehen.[97] Des Weiteren fällt Satire in den Bereich der Fehlinformation, wobei die Absicht hinter der Falschinformation in der Regel auf einer ironischen Überzeichnung beruht und nicht auf einer unbeabsichtigten politischen Einflussnahme.[98]

Eine von den Medienanstalten veröffentlichte Studie klassifiziert Falschinformationen anhand zweier Definitionsebenen, der Faktizität und der Absicht des Senders (siehe Abbildung 1). Die Darstellung in der Abbildung unterteilt die Kategorien in vier Quadranten:

91 Vgl. Woodward 2020
92 Vgl. Egelhofer et al. 2020; Möller et al. 2020
93 Vgl. Wardle/Derakhshan 2017; Dreyer et al. 2021
94 Vgl. Möller et al. 2020
95 Vgl. Tandoc et al. 2018; Wardle 2017; Möller et al. 2020
96 Vgl. Stapf 2021
97 Vgl. Fallis 2015; Dreyer et al. 2021
98 Vgl. Zimmermann/Kohring 2018

Quadrant I (Misinformation): Geringfügige Abweichung von der Wahrheit, ohne bewusste Täuschungsabsicht. Dies umfasst ungenaue und unbeabsichtigt dekontextualisierte Berichterstattung.

Quadrant II (Misinformation): Starke Abweichung von der Wahrheit, ohne bewusste Täuschungsabsicht. Hierzu gehören unbeabsichtigt irreführende Inhalte.

Quadrant III (Desinformation): Geringfügige Abweichung von der Wahrheit, mit bewusster Täuschungsabsicht. Dies schließt bewusste Dekontextualisierung realer Informationen und unauthentischen, irreführenden Pseudojournalismus ein.

Quadrant IV (Desinformation): Starke Abweichung von der Wahrheit, mit bewusster Täuschungsabsicht. Hierzu zählen bewusste Falschinformationen, manipulative (politische) Werbung und Propaganda.

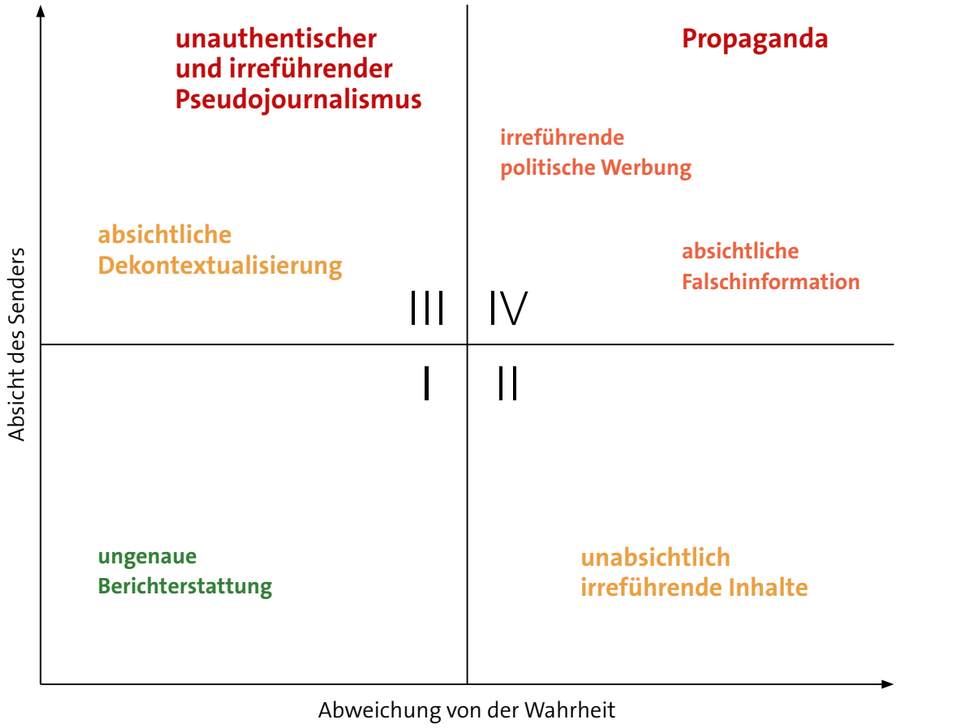

Abbildung 1: Verschiedene Arten der Mis- und Desinformation. Quelle: Möller et al. 2020, S. 13

Die Farbgebung in der Abbildung reflektiert das gesellschaftliche Risiko, das durch die vorliegenden Formen von Mis- bzw. Desinformation entsteht. Grüne Kennzeichnungen repräsentieren ein geringes Risiko, wohingegen mit zunehmender Schriftgröße und Intensität des Rottons auch das Risiko für die Gesellschaft steigt.

Unpräzise Berichterstattung weist üblicherweise einen vergleichsweise hohen Grad an Faktentreue auf, die Verfasser:innen verfolgen in der Regel keine expliziten Ziele damit. Eine ebenso unbeabsichtigte, jedoch ausgeprägtere Abweichung von der Wahrheit wird als unbeabsichtigt irreführender Inhalt bezeichnet. Misinformation wird nicht absichtlich und gezielt verbreitet, was ihre Bekämpfung unkomplizierter gestaltet.[99] Im einfachsten Fall können dabei kleine Fehler vorliegen.

Im Gegensatz dazu verfolgt Desinformation das grundlegende Ziel, absichtlich zu täuschen und gezielt Wahlen, die öffentliche Meinung sowie den öffentlichen Diskurs zu manipulieren. Dies trifft auf absichtlich verbreitete Falschinformationen und Propaganda zu, wobei die Hauptakteure letztgenannter Regierungen und andere Akteure in Machtpositionen sind. Absichtliche Dekontextualisierung bewegt sich zwar in relativer Nähe zur Wahrheit, jedoch werden durch die bewusst fehlerhafte Einordnung gezielt falsche Inhalte vermittelt. Unauthentischer und irreführender Pseudojournalismus weist in der Regel bis zu einem gewissen Grad einen Bezug zur Wahrheit auf. Aufgrund der journalistischen Darstellung und der dazugehörigen Sprache fällt es den Rezipient:innen jedoch besonders schwer, Abweichungen von der Faktizität zu erkennen. Irreführende politische Werbung wird primär in Wahlkämpfen eingesetzt.[100] Tabelle 2 gibt eine Übersicht über die Erscheinungsformen von Desinformation.

Für eine eingehendere Analyse zum Thema Desinformation und Misinformation wird auf das Gutachten „Typen von Desinformation und Misinformation" der Medienanstalten verwiesen.[101] Das vorliegende Gutachten fokussiert an dieser Stelle darauf, welche Auswirkungen sich in den Bereichen Desinformation durch die Anwendung von KI ergeben. Obwohl die Möglichkeit besteht, dass durch Vorschlagsalgorithmen Misinformation schneller und weiter verbreitet wird, als dies ohne KI der Fall wäre, ist das hiermit verbundene Potenzial vergleichsweise gering. Aus diesem Grund wird im weiteren Verlauf keine detaillierte Analyse von Misinformation durchgeführt.

99 Vgl. Thorson 2016
100 Vgl. Möller et al. 2020
101 Vgl. Möller et al. 2020

Typen von Desinformation					
	Dekontextualisierung	**Falschinformation**	**Manipulative (politische) Werbung**	**Pseudojournalismus**	**Propaganda**
Abweichung von der Faktizität	gering	hoch	verschieden	eher gering	hoch
Typische Intention	manipulatives Narrativ verbreiten, das eine politische Ideologie stützt; ökonomisch (Clickbait)	ökonomisch; ideologische (De-)mobilisierung	politische Mobilisierung	ökonomisch; ideologische (De-)Mobilisierung	geopolitische und ideologische (De-)Mobilisierung
Typische Sender	politische Akteure, Medien, alternative Medien	Internetbetrüger, Verschwörungstheoretiker, alternative Medien	politische Akteure, NGOs	Internetbetrüger, alternative Medienmacher	Staatsregierungen und (internationale) Organisationen
Typische Verbreitung	weite Verbreitung: kommt in allen Medien vor, häufig in alternativen Medien zu finden, weitere Verbreitung durch Nutzer	eingeschränkte Verbreitung: häufig über Soziale Medien, manchmal unterstützt durch koordiniertes unauthentisches Verhalten	bezahlte Verbreitung: häufig in Sozialen Medien, aber auch über andere Wahlkampfkanäle	eingeschränkte Verbreitung: in eigenen Onlinemedien oder über Soziale Netzwerke, weitere Verbreitung durch Nutzer	professionelle Verbreitung: über alle Kommunikationskanäle einschließlich eigener Medienorganisationen und mithilfe von koordiniertem unauthentischen Verhalten
Typische Risiken für das Individuum	kognitiv, emotional, (politische) Fehlentscheidungen (in unterschiedlichem Ausmaß)				
Typische Risiken für die Gesellschaft	misinformierte Wählerschaft, polarisierend	misinformierte Wählerschaft, spaltend, demokratiegefährdend	polarisierend	misinformierte Wählerschaft, polarisierend, spaltend	geopolitisch, spaltend, demokratiegefährdend

Erklärung: politisch = taktische, kurzfristige Wahlbeeinflussung, ideologisch = strategische, langfristige Beeinflussung, polarisierend = gesamtgesellschaftlich, spaltend = Segmente betreffend, demokratiegefährdend = hochgradig manipulativ

Tabelle 2: Typen von Desinformation. Quelle: Die Medienanstalten 2021a, S. 15

3.3.2 Künstliche Intelligenz zur Erstellung von Desinformation

Das Schadenspotenzial von Desinformation ist für die Gesellschaft als Ganzes sowie für einzelne Individuen insgesamt erheblich. Insbesondere, wenn sie mit bösartiger Absicht eingesetzt wird, ermöglicht KI eine massive Steigerung des Potenzials durch Desinformation. KI bietet vielfältige Möglichkeiten in diesem Kontext: Sie kann dazu beitragen, überzeugendere Desinformationen einfacher und schneller zu erstellen, wie das Fallbeispiel Countercloud (siehe Infobox) eindrücklich zeigt. Deepfakes, die mittels KI erstellt werden, sind mittlerweile für ungeübte Mediennutzer:innen

nur schwer von Originalen zu unterscheiden. Zugleich sinken die Zugangshürden für die Nutzung solcher Technologien, sodass eine wachsende Anzahl von Menschen die Fähigkeit hat, Deepfakes zu erstellen.

Im Folgenden werden die Möglichkeiten analysiert, die KI bietet, um Desinformation zu erstellen.

Bewusste Dekontextualisierung
Bei der bewussten Dekontextualisierung wird eine grundsätzlich korrekte Information absichtlich in einen falschen Kontext gestellt, um eine andere Bedeutung zu vermitteln (siehe Tabelle 2). Ein Beispiel hierfür ist die Manipulation von Bildern bekannter Politiker:innen durch Hinzufügen eines falschen Datums und einer anderen Bildunterschrift, um eine irreführende Verbindung herzustellen. Die bewusste Dekontextualisierung kann ein erhebliches Risiko für die demokratische Gesellschaftsordnung darstellen. Für Faktenchecker:innen stellt diese Form der Desinformation gegenwärtig eine größere Herausforderung dar als Deepfakes.[102] Da jedoch im Wesentlichen keine neuen Inhalte generiert werden, sind die Anwendungsmöglichkeiten von KI für die Erstellung begrenzt. Zwar können mithilfe von Sprachmodellen prägnantere Bildunterschriften gefunden werden, die zentrale kognitive Aufgabe, Inhalte in einen neuen – wenn auch falschen – sinnvollen Kontext zu setzen, bleibt allerdings bei der Person, die die Desinformation erstellt.

Schriftliche Desinformation
Insbesondere geschriebene Texte sind für Anwender:innen schwer als das Produkt von KI zu erkennen, was weitreichende Potenziale im Bereich der Desinformation mit sich bringt. Darüber hinaus trägt die Anwendung eines journalistischen Schreibstils dazu bei, die Glaubwürdigkeit von Texten erheblich zu steigern.[103] Mittels KI können bereits jetzt Texte verfasst werden, die sich dem journalistischen Sprachgebrauch stark annähern, ohne die bisher erforderlichen sprachlichen Fähigkeiten zu besitzen. Zudem können Texte auch gezielt in eine zielgruppengerechte Sprache übertragen werden, was eine Verbreitung von Inhalten an spezifische Gruppen erleichtert. Dies ermöglicht es, die generierten Inhalte an die individuellen Interessen, Überzeugungen oder Vorurteile der Nutzer:innen anzupassen, um eine größere Wirkung zu erzielen.

Bei der Nutzung von generativer KI zur Texterstellung besteht derzeit noch eine erhebliche Schwäche: Die Inhalte werden nicht auf ihre Faktizität überprüft. Die KI halluziniert bisweilen, das heißt, sie erfindet Informationen ohne Bezug zur Realität. Bei der Erzeugung von Desinformationen ist diese Schwäche allerdings von geringer Bedeutung und kann sogar als Stärke fungieren, solange die Rezipient:innen die Inhalte als glaubwürdig empfinden.[104] Diese können gefälschte Nachrichten, manipulierte Zitate oder falsche Behauptungen beinhalten, die darauf abzielen, eine bestimmte Agenda zu fördern oder die Meinung der Leser:innen zu beeinflussen.

Die Möglichkeiten, KI zur Erstellung von Desinformation zu nutzen, gehen jedoch noch weiter. Der Prozess, Texte und Bilder zu generieren, kann ebenfalls automatisiert werden. Die Entwickler des Konzepts Countercloud (siehe Infobox) zeigen eindrucksvoll, dass sich vollständige Nachrichtenportale mit Inhalten füllen lassen, die vollautomatisch durch KI produziert wurden und alle einem vorgegebenen Narrativ folgen.

102 Vgl. Weikmann/Lecheler 2023
103 Vgl. Holzer/Sengl 2020
104 Vgl. Marcus 2020

Veränderung der Medienlandschaft durch KI – Chancen und Risiken

Fallbeispiel Countercloud

Das unveröffentlichte Proof-of-Concept Countercloud, erstellt von unbekannten Urhebern unter dem Pseudonym Nea Paw, verdeutlicht eindrücklich die Möglichkeiten, die sich durch KI im Bereich der Desinformation ergeben.[105] Trotz stark begrenzter finanzieller Ressourcen und der Entwicklung durch lediglich zwei Personen lassen sich die Wirkpotenziale erahnen, die mit solchen Werkzeugen von ressourcenstarken Akteuren erreicht werden können, die beabsichtigen, mittels Desinformation politischen Einfluss zu nehmen.

Es handelt sich hierbei um ein Konzept zur vollautomatischen Erstellung von Desinformation. Dabei werden bestimmte Narrative festgelegt, die unterstützt oder opponiert werden sollen. Auf Basis dieser vordefinierten Weltanschauung kann ein umfassendes Ökosystem automatisch generierter Desinformationen geschaffen werden, die alle den vorgegebenen Narrativen folgen: von Nachrichtenseiten über Kommentare bis hin zu X (vormals Twitter)-Accounts.

Als Ausgangspunkt dienen die Inhalte bestehender Nachrichtenportale und Tweets. Diese werden in generative Sprachmodelle geladen, und darauf aufbauend werden vollautomatisch Gegenartikel bzw. -tweets verfasst. Diese generierten Inhalte folgen den vorgegebenen Narrativen, verwenden gefälschte Sachverhalte sowie historische Ereignisse, um eine starke Wirkung auf die Leser:innen zu erzielen. Gleichzeitig wird die Faktizität des ursprünglichen Inhalts in Frage gestellt. Es besteht darüber hinaus die Möglichkeit, bereits existierende Verschwörungstheorien in die verfassten Texte einzubetten. Die generierten Inhalte können präzise angepasst und in ein Gesamtkonstrukt integriert werden. Nachrichtenseiten und X (vormals Twitter)-Accounts, die einen seriösen Eindruck vermitteln sollen, verzichten auf Verschwörungstheorien, während diese vermehrt bei unseriösen Nachrichtenportalen, Kommentaren und dezidierten X (vormals Twitter)-Accounts genutzt werden.

Soweit möglich, werden Abbildungen aus den ursprünglichen Artikeln verwendet und bewusst dekontextualisiert. Falls dies nicht realisierbar ist, beispielsweise wenn der Text fest im Bild eingebettet ist, werden passende Bilder mithilfe von KI automatisiert generiert. Zur Steigerung der Realitätsnähe dieser Täuschung werden fingierte Autorenprofile erstellt, die Informationen wie Namen, Biografien und gefälschte Portraitfotos umfassen. Die Biografien dieser Autor:innen werden an die von ihnen verfassten Inhalte angepasst, um einen konsistenten Eindruck hinsichtlich Ort, Sprache, Interessen und politischer Ansichten zu erzeugen. Ein weiterer Ansatz zur Verbesserung der Täuschung besteht darin, einzelne Artikel mittels Text-to-Speech zu vertonen.

Zur Erschwerung der Aufdeckung weisen die Inhalte eine breite Spanne an sprachlichem Niveau auf, die jeweils an das Medium angepasst ist. Außerdem werden zufallsgenerierte Veränderung eingebaut, wie etwa die Länge, die Absatzanzahl oder die Satzlänge. Zugleich variiert die Vehemenz der Gegenargumentation. Ein wesentlicher Beitrag zur Tarnung wird durch das implementierte Gatekeeper-Modul (siehe Abbildung 2) geleistet.

105 https://countercloud.io/?page_id=307

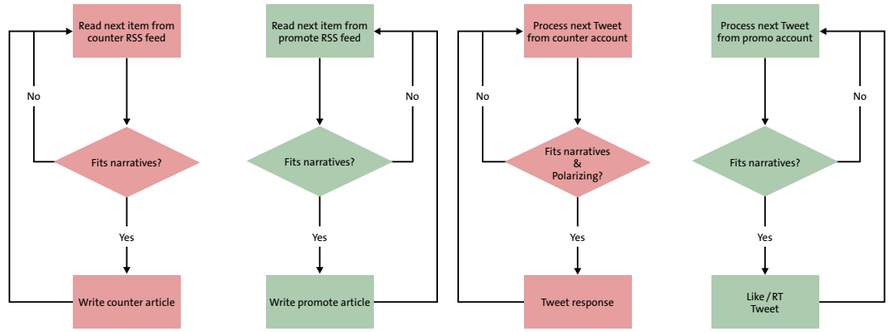

Abbildung 2: Flussdiagramm des Gatekeeper-Moduls. Quelle: https://countercloud.io/?page_id=307

Dieses Modul gewährleistet, dass nicht unter jeden Artikel und jeden Tweet automatisch eine Gegenantwort verfasst wird. Gegenargumentationen werden nur dann generiert, wenn eine inhaltlich relevante und sinnvolle Gegenposition vorliegt. Um eine Enttarnung zu verhindern, werden leicht widerlegbare Inhalte, wie beispielsweise das Ergebnis eines Fußballspiels, bewusst nicht in Frage gestellt. Zudem erfolgt eine selektive Anwendung von Kommentaren, indem nicht unter jeden Artikel kommentiert wird; bei einigen Artikeln mehr, bei anderen weniger oder überhaupt nicht. Darüber hinaus repräsentieren die Kommentare unterschiedliche Meinungen, wobei einige Kommentare den Inhalten sogar widersprechen. Diese Bausteine werden moderat eingesetzt, und nicht alle Artikel weisen Bilder, Sprachausgaben oder Kommentare auf.

Das Proof-of-Concept stützte sich primär auf Chat-GPT und war somit den vom Entwicklerteam gesetzten Beschränkungen unterworfen. Nichtsdestotrotz wurde ebenfalls demonstriert, dass dasselbe Konzept auf Open-Source-Sprachmodellen basieren kann. Diese Modelle unterliegen nicht den gleichen Einschränkungen und können beispielsweise auch Hassbotschaften und Hetze mithilfe derselben Methodik verbreiten.

Obgleich noch einige Herausforderungen existieren, lassen sich diese bei umfangreicheren Projekten mit ausreichenden Ressourcen leicht beheben. Eine Möglichkeit besteht darin, die automatisch erstellten Artikel, Tweets und Kommentare kurz von Redakteur:innen gegenlesen zu lassen, um grobe Fehler zu vermeiden. Die Urheber schätzen, dass rund 90 Prozent der vollautomatisch produzierten Inhalte überzeugend sind und dass durch minimale menschliche Interaktion sowie zusätzliche Feinjustierungen der gesamte produzierte Inhalt überzeugend gestaltet werden kann. Ressourcenstarke Akteure, wie beispielsweise russische Trollfabriken, könnten diese Instrumente nutzen, um die Menge des produzierten Propagandamaterials sowie die Reichweite massiv zu erhöhen.

Die Urheber:innen schätzen, dass mit etwa 4000 Dollar pro Monat Gegenpositionen zu etwa 40 Pressekanälen, 40 X (vormals Twitter)-Accounts und 200 Artikeln pro Tag erstellt werden können. Das Konzept kann neben X (vormals Twitter) auch auf andere soziale Medien übertragen werden. Hierdurch besteht die Möglichkeit, ein ausgedehntes Netzwerk an zielgruppenspezifischen Medien zu schaffen, die alle mehr oder weniger die gleichen Narrative bedienen und ernstzunehmende Zweifel an den Veröffentlichungen seriöser Nachrichtenportale säen könnten.

Deepfakes

Die Manipulation sowie Fälschung von Bildmaterial stellt ein bekanntes Phänomen dar und ist keineswegs als neuartig zu betrachten. Exemplarisch hierfür kann Abbildung 3 herangezogen werden, welche das berühmte Foto von Stalin vor dem Bolschoi-Theater in Moskau zeigt, das nachträglich einer Retusche unterzogen wurde, bei der Trotzki und Kamenew entfernt wurden. Mit dem Voranschreiten technologischer Entwicklungen verzeichnen Fälschungen nicht nur eine Verbesserung in der Qualität, sondern es wird ebenso eine Abnahme der Zugangshürden verzeichnet, die es ermöglicht, derartige Manipulationen zu generieren. Die jüngste Entwicklungsstufe von Fälschungspraktiken wird mit dem Terminus Deepfakes zusammengefasst.[106]

Abbildung 3: Lenin vor dem Bolschoi-Theater in Moskau. Das Original vorher (links) und die Bearbeitung nachher. Die Figuren Trotzki und Kamenew wurden nachträglich aus dem Foto entfernt. Quelle: Muscionico 2022

106 Vgl. Die Medienanstalten 2021a

Abbildung 4: Deepfake von Papst Franziskus. Quelle: Reddit / Midjourney https://www.reddit.com/r/midjourney/comments/120vhdc/the_pope_drip/?rdt=53204

Abbildung 4 zeigt ein Deepfake von Papst Franziskus in einer Daunenjacke, welches in einem Internetforum veröffentlicht wurde. Die Urheberschaft dieser Manipulation bleibt unbekannt, wobei die fortschreitenden KI-Technologien die Produktion solcher Fälschungen selbst für Personen ohne professionelle Expertise möglich macht. Die Erstellung des Bilds von Papst Franziskus erfolgte mittels der Software Midjourney[107], einer KI-Anwendung, die speziell für die Schaffung von Kunstwerken unter Einsatz von KI entwickelt wurde. Obwohl die konkrete Abbildung als Fälschung gekennzeichnet wurde, nahmen sie dennoch einige Nutzer:innen des betreffenden Forums als authentisch war. Die Nutzung derartiger Technologien für die gezielte Erzeugung von Desinformation kann eine ernstzunehmende Bedrohung für die öffentliche Wahrnehmung von Informationen sowie für den öffentlichen politischen Diskurs darstellen. In diesem Kontext ist anzunehmen, dass die Potenziale zur Erstellung hochwertiger Fälschungen in den kommenden Jahren sowohl in Bezug auf die Benutzerfreundlichkeit als auch hinsichtlich der Qualität zunehmen werden.

Abbildung 5 bietet eine umfassende Darstellung der gegenwärtigen Methoden zur Fälschungserstellung. Hierbei erfolgt eine Differenzierung zwischen hochwertigen und aufwendigen Deepfakes sowie sogenannten Cheap Fakes. Eine bekannte Variante der Deepfakes ist das Face-Swap, welches den Austausch des Gesichts einer Quellperson mit dem Gesicht einer Zielperson in Bildern oder Videos bezeichnet. Diese Technik ermöglicht die Simulation von Aussagen und Handlungen, die von den betroffenen Personen nie geäußert oder vollzogen wurden. Obschon in der öffentlichen Diskussion um Deepfakes visuelle Fälschungen oft im Vordergrund stehen, sollten Audio-Manipulationen und Sprachsynthese nicht vernachlässigt werden, da Sprache einen bedeutenden Träger von Informationen darstellt und somit diverse Angriffspunkte bietet.[108] Derzeit stellt schriftliche Desinformation die größte Bedrohung dar, da die Identifikation von synthetisch produzierten Texten deutlich anspruchsvoller ist als die von audiovisuellen Inhalten.[109]

Insbesondere im linken Bereich des Spektrums, wie in Abbildung 5 dargestellt, manifestieren sich hochwertige Fälschungen, für deren Umsetzung die Einstiegshürden gegenwärtig noch vergleichsweise hoch sind. In der jüngsten Vergangenheit gab es jedoch mehrere Start-Up-Gründungen, die das Ziel verfolgen, diese Technologien leichter zugänglich zu machen. Beispiele hierfür sind Unternehmen wie Hour One und Synthesia (siehe Tabelle 1). Obwohl diese Unternehmen betonen, die Einsatzmöglichkeiten synthetischer Avatare im Rahmen legaler Anwendungen in Einklang mit rechtlichen Vorgaben zu entwickeln, sind im Internet bereits

107 https://www.midjourney.com/
108 Vgl. Die Medienanstalten 2021a
109 Vgl. DiResta 2020

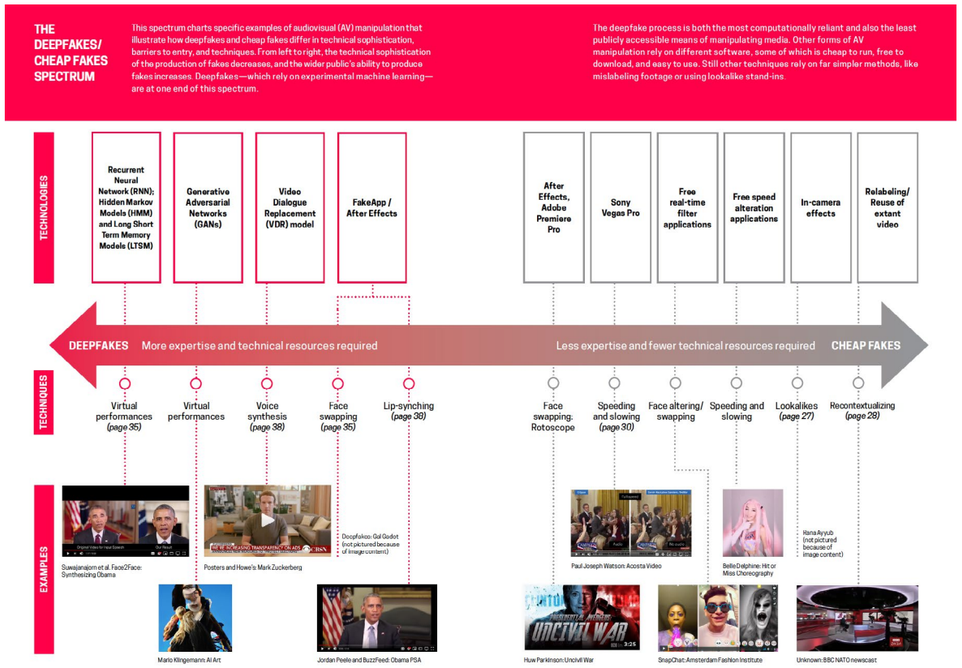

Abbildung 5: Das Spektrum zwischen Deepfakes und Cheap Fakes. Quelle: https://datasociety.net/library/deepfakes-and-cheap-fakes/

Obwohl Deepfakes bereits erfolgreich genutzt wurden, beispielsweise für Banküberfälle[111] und Betrug[112], befinden sie sich noch nicht in einem Reifestadium, das eine vergleichsweise einfache Enttarnung verhindert. Insbesondere automatisch erzeugte synthetische Stimmen und Avatare lassen sich noch leicht identifizieren.[113] Die Erstellung hochwertiger Deepfakes bleibt gegenwärtig komplex, jedoch haben einige Hollywood-Filme bereits die Potenziale illustriert.[114] Das Beispiel von Franziska Giffey, die während eines Videotelefonats mit dem vermeintlichen Vitali Klitschko einem unbekannten Betrüger aufgesessen ist, veranschaulicht lediglich die Spitze des Eisbergs an Möglichkeiten, die Akteure mit bösartigen Absichten durch diese Technologien erreichen können.[115] Eine denkbare Anwendung könnte beispielsweise die vorgetäuschte Ermordung eines Staatsoberhaupts während einer auf ähnliche Weise manipulierten Videoschalte sein, mit dem Ziel Unruhen anzustiften.[116] Sam Gregory, Geschäftsführer der Menschenrechtsorganisation Witness, erkennt in der Kombination verschiedener Ansätze eine erhebliche Bedrohung. Durch den gemeinsamen Einsatz von Deepfakes, virtuellen Avataren, automatisierter Sprachgenerierung und Sprachmodellen können schnell und kostengünstig synthetische Figuren geschaffen werden, die nicht real existieren, aber in der Lage sind, Narrative und Desinformation in sämtlichen Sprachen weltweit zu verbreiten. Gleichzeitig ermöglichen diese Technologien die Erstellung täuschend echter gefälschter Videos von Prominenten und Politiker:innen.

110 Vgl. Satariano / Mozur 2023
111 Vgl. Ropek 2021
112 Vgl. Brown 2019
113 Vgl. Satariano / Mozur 2023
114 Vgl. beispielsweise Giardina 2016; Giardina 2015; Reul 2022
115 Vgl. Simmons et al. 2022
116 Vgl. Horvitz 2022

Liar's Dividend

Die strategische Verwendung von Deepfakes zur Diskreditierung authentischer Inhalte stellt eine besonders simple Strategie zur Manipulation der öffentlichen Meinung dar. Durch die bloße Möglichkeit der Erstellung von Deepfakes kann die Authentizität von Informationen begründet in Frage gestellt werden.[117] In diesem Kontext findet häufig das Konzept der Liar's Dividend Anwendung. Das Konzept bezeichnet eine Situation, in der eine Person, die falsche oder irreführende Informationen verbreitet (beispielsweise, dass ein Nachrichteninhalt ein Deepfake ist), davon profitiert, selbst wenn ihre Behauptungen später als falsche aufgedeckt werden. Dieser Nutzen kann sich in verschiedenen Formen, darunter die Beeinflussung der öffentlichen Meinung, die Manipulation von Ereignissen sowie die Schaffung von Unsicherheit und Verwirrung manifestieren. Mechanismen, die die Liar's Dividend antreiben, sind vielfältig. Einerseits trägt die anfängliche Desinformation dazu bei, Zweifel zu säen und das Vertrauen in etablierte Institutionen zu untergraben. Selbst nach Präsentation von Fakten zur Widerlegung können Skepsis und Ablehnung persistieren, was zu einer anhaltenden Spaltung in der öffentlichen Meinung führen kann. Ein weiterer Mechanismus ist die Geschwindigkeit der Desinformationsverbreitung im Vergleich zur häufig langsameren Korrektur oder Aufklärung. Dieser zeitliche Unterschied ermöglicht den Akteuren hinter den falschen Informationen, bereits vor der Klärung erheblichen Einfluss auszuüben und politische Ziele zu erreichen.[118]

Synthetische Geschichtsschreibung

Eric Horvitz beschreibt einen Ansatz, wie der gezielte Einsatz von Fälschungen dazu genutzt werden kann, nicht nur punktuell Desinformation zu verbreiten, sondern umfassende synthetische Geschichtsschreibung zu betreiben. Dieser Ansatz beinhaltet die Ergänzung von realen Weltgeschehnissen durch erdachte und gefälschte Ereignisse, um das Narrativ gezielt in eine bestimmte Richtung zu lenken. Dies schließt nicht nur die Verwendung von Deepfakes ein, sondern umfasst auch die gezielte Inszenierung von Ereignissen, wie beispielsweise die orchestrierte Darstellung von Protesten.[119] Ziel dabei ist es, einem spezifischen Narrativ zusätzliche Bedeutung und Überzeugungskraft zu verleihen.

Es erscheint plausibel, dass ein konstruiertes (einfaches) Narrativ eine höhere Glaubwürdigkeit erzeugen kann als die vielschichtigen und komplexen Zusammenhänge, die der Realität zugrunde liegen. Selbst wenn dies nicht erreicht wird, kann zumindest eine alternative Erklärung für Weltgeschehnisse präsentiert werden, die kontrastierend zur Realität steht

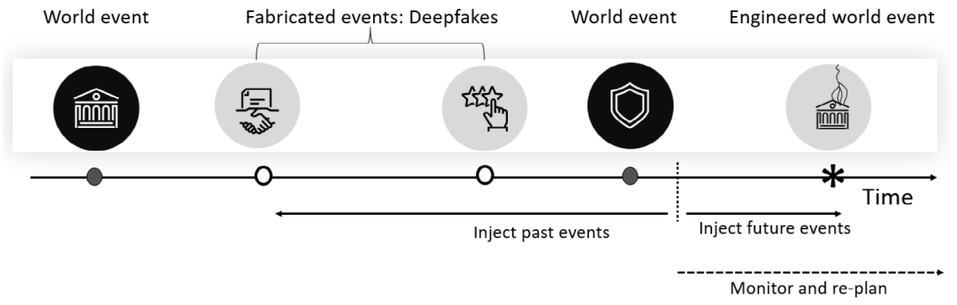

Abbildung 6: Synthetische Geschichtsschreibung. Quelle: Horvitz 2022

117 Vgl. beispielsweise Metzger 2024
118 Vgl. Ryan-Mosley 2023
119 Vgl. beispielsweise Bewarder et al. 2023

Abbildung 6 visualisiert das Konzept grafisch. Ausgehend vom aktuellen Zeitpunkt werden Deepfakes erstellt, die vortäuschen, zwischen vergangenen Ereignissen stattgefunden zu haben. Inszenierte Ereignisse dienen dazu, das konstruierte Narrativ zu stützen. Durch aktives Monitoring der Stimmung und des Verhaltens der Zielgruppe können Anpassungen an der Strategie vorgenommen werden. Zur Steigerung der Effektivität schlägt Horvitz vor, die Strategien an kleineren Gruppen von Proband:innen zu testen. Zusätzlich können weitere Methoden, wie beispielsweise das Microtargeting, integriert werden, um unterschiedlichen Zielgruppen jeweils ein eigenes, angepasstes Narrativ zu präsentieren.[120]

Um zeitlich zurückliegende Fälschungen zu erstellen, existieren zwei methodische Ansätze. Eine Möglichkeit besteht in der Manipulation von Zeitstempeln. Jedoch könnte dies zu Misstrauen führen, insbesondere wenn synthetische Ereignisse erst zu einem deutlich späteren Zeitpunkt auf etablierten Plattformen veröffentlicht werden. Eine alternative Vorgehensweise ist das Pre-Positionieren von Fälschungen. Hierbei werden Deepfakes auf unbekannten Accounts hochgeladen, um sie zu einem späteren Zeitpunkt hervorzuholen und durch soziale Medien und unseriöse Nachrichtenquellen zu verbreiten. Um flexibel auf diverse Ereignisse vorbereitet zu sein, können verschiedene Fälschungen im Voraus positioniert und passend zu den real manifestierten Geschehnissen entweder verbreitet oder entfernt werden.

In der Tat wendete Russland dieses Vorgehen im Zuge des Angriffs auf die Ukraine an. Im November 2021 wurde ein Video auf YouTube hochgeladen, in dem behauptet wurde, dass sich in der Ukraine Labore für Biowaffen befänden, die von den USA finanziert würden. Als die Invasion auf die Ukraine begann, wurde diese Geschichte von staatlichen russischen Nachrichtenkanälen wieder aufgegriffen und über soziale Medien verbreitet.[121]

3.3.3 Künstliche Intelligenz zur Verbreitung von Desinformation

Die Nutzung von KI bietet nicht nur Potenziale zur Erstellung von Desinformation, sondern auch vielfältige Möglichkeiten für deren Verbreitung, insbesondere auf sozialen Medien. Derzeit ist die Erstellung von Inhalten eine vergleichsweise geringere Hürde im Vergleich zur Herausforderung, diese Inhalte effektiv zu verbreiten.[122] Im Rahmen der Verbreitung von Desinformation können die Vorschlagsalgorithmen sozialer Plattformen gezielt ausgenutzt werden, um Desinformationen rasch an einen breiten Kreis an Rezipient:innen zu verteilen. Dies kann beispielsweise durch den gezielten Einsatz hyperaktiver Bots erfolgen, die bestimmte Meldungen hervorheben und somit den gesellschaftlichen Diskurs verzerren.[123] Microtargeting wird dabei immer häufiger eingesetzt, um gezielt (falsche) Informationen zu verbreiten. Da Menschen durch ihre Präsenz in sozialen Medien zunehmend mehr Daten preisgeben, können immer präzisere Profile erstellt werden, die für das Microtargeting genutzt werden können.

Im Folgenden werden verschiedene Wege beleuchtet, wie KI in diesem Kontext eingesetzt werden kann.

120 Vgl. Horvitz 2022
121 Vgl. Microsoft 2022 und Cercone 2022
122 Vgl. Simon et al. 2023b
123 Vgl. Serrano et al. 2019

Digitale Wahlwerbung
Soziale Medienplattformen spielen eine zunehmend bedeutende Rolle im politischen Meinungsbildungsprozess. Während der Präsidentschaftswahl in den USA im Jahr 2012 bezogen lediglich etwa 12 Prozent der US-Amerikaner:innen regelmäßig Wahlwerbung über Facebook.[124] Diese Zahl stieg bis zum Jahr 2016 auf über 60 Prozent an, wodurch Facebook auf Platz drei der am häufigsten zitierten Quellen für Informationen während der Wahlkampagne dieses Jahres gelistet wurde.[125]

In Europa beziehen 37 Prozent der Menschen ihre Nachrichten in erster Linie über soziale Medien, wobei das Vertrauen in diese Informationskanäle mit 16 Prozent nicht besonders ausgeprägt ist. Im Gegensatz dazu werden die klassischen Kanäle wie öffentlich-rechtliche Fernseh- und Radiostationen sowie Presseberichte nach wie vor als deutlich vertrauenswürdiger wahrgenommen.[126]

Interne Dokumente von Facebook, die der Securities and Exchange Commission vorliegen, offenbaren, dass Wahlwerbung hohe Umsätze für den Konzern generiert. Facebook bietet Werbetreibenden vielfältige Möglichkeiten, ihre Werbung gezielt zu verbreiten. Während des US-amerikanischen Wahlkampfs 2020 unterzog Facebook politische Werbeanzeigen keinen Faktenchecks.[127] Der Konzern Meta erlaubt weiterhin Wahlwerbung sowohl für Kandidat:innen auf öffentliche Ämter als auch von Political Action Committees (PACs), wobei die Option eingeführt wurde, dass Nutzer:innen sich proaktiv von politischer Werbung ausschließen können (Opt-out).[128] Im Gegensatz dazu führte die Plattform X (vormals Twitter) 2019 ein Verbot politischer Werbung ein, das der Eigentümer Elon Musk nun aufgehoben hat.[129] Google beschränkt die Möglichkeiten des Microtargetings für politische Parteien.[130]

Whistleblower Christopher Wylie beschreibt, dass während der Wahlkampagne von Donald Trump im Jahr 2016 ein umfassendes Ökosystem aus vermeintlich unabhängigen Webseiten und Blogs aufgebaut wurde, das gezielt an besonders empfängliche Individuen kommuniziert wurde. Wiley charakterisiert dieses System als ein „Werkzeug der psychologischen Kriegsführung".[131] Diese Werkzeuge und Methoden zur gezielten Einflussnahme durch Propaganda und Desinformation können auch ausländische Akteure nutzen. Laut einem Bericht des britischen Komitees für Digitales, Kultur, Medien und Sport hat ein unbekannter Akteur mehr als 250.000 britische Pfund für Wahlwerbung auf Facebook ausgegeben, mit dem Ziel, den Brexit voranzutreiben. Diese Kampagne erreichte über 10 Millionen britische Bürger:innen.[132]

Unter anderem aufgrund dieser Entwicklungen hat die Europäische Kommission im Vorfeld der Europawahlen 2024 ein Maßnahmenpaket zur Verteidigung der Demokratie geschnürt, um der Bedrohung durch Einflussnahme von Drittstaaten mit erhöhter Transparenz entgegenzuwirken.[133] Die Europäische Union führte 2018 erstmals einen Verhaltenskodex für Desinformation ein, der 2022 zuletzt erneuert wurde. Für die beteiligten Unternehmen, insbesondere sehr große digitale

124 Vgl. Pew Research Center 2012
125 Vgl. Pew Research Center 2014; Pew Research Center 2016; Pew Research Center 2018
126 Vgl. European Parliament 2023
127 Vgl. Lima-Strong 2021
128 Vgl. Meta 2024
129 Vgl. Tagesschau 2023
130 Vgl. Langer 2019
131 Vgl. Dachwitz/Rudl 2018
132 Vgl. UK Parliament 2019
133 Vgl. European Commission 2023a

Plattformen, soll der Kodex eine Maßnahme zur Risikominderung im Rahmen der Koregulierung durch den Digital Service Act (DSA) werden (siehe auch Kapitel 4.1).[134] Der Kodex wurde von den Betreibern aller großen sozialen Medienplattformen außer X (vormals Twitter) unterzeichnet.[135]

Social Bots
Die Vorschlagsalgorithmen digitaler Plattformen sind anfällig für Manipulationen durch hyperaktive Accounts, unabhängig davon, ob diese von Menschen oder KI gesteuert werden. Es besteht die Möglichkeit, Fake-Accounts auf sozialen Medienplattformen zu erstellen, die mit Profilbildern, Biografien und weiteren Informationen ausgestattet werden, um einen authentischen Eindruck zu vermitteln (siehe Infobox Countercloud). Diese erstellten Fake-Accounts können von sogenannten Social Bots gesteuert werden, die automatisierte Aktivitäten wie das Teilen von Inhalten, das Veröffentlichen von Beiträgen und das Liken von Beiträgen anderer Nutzer:innen durchführen. Diese Aktivitäten sollen dazu beitragen, dass die Fake-Accounts wie echte Nutzer:innen erscheinen.

Social Bots können gezielt bestimmte Inhalte verbreiten, sei es in Form von Artikeln, Videos, Bildern oder Meinungsäußerungen. Dies kann dazu dienen, spezifische politische Ansichten und Narrative zu fördern, Desinformation zu verbreiten oder Diskussionen in eine gewünschte Richtung zu lenken. Social Bots können eingesetzt werden, um die Vorschlagsalgorithmen von sozialen Medienplattformen zu manipulieren, indem sie künstlich Trends generieren, beispielsweise durch die massenhafte Verwendung bestimmter Hashtags oder Fokussierung auf bestimmte Themen. Dies kann dazu führen, dass echte Nutzer:innen auf den Trend aufmerksam werden und sich der Diskussion anschließen. Die Bots können so programmiert werden, dass sie kontinuierlich mit anderen Nutzer:innen interagieren, um ihre Glaubwürdigkeit zu erhöhen, sei es durch das Beantworten von Kommentaren, das Teilen von relevanten Inhalten oder das Vortäuschen von menschenähnlichem Verhalten.

Microtargeting
Eine Methode, Inhalte zu verbreiten, ist das sogenannte Microtargeting. Microtargeting bezeichnet die gezielte Ausspielung von Informationen an kleine Personengruppen, basierend auf spezifischen Persönlichkeitsmerkmalen oder -eigenschaften. Das Thema erlangte spätestens durch die Enthüllungen rund um die Firma Cambridge Analytica im Jahr 2015 öffentliche Aufmerksamkeit. Die besagte Firma hatte unrechtmäßigen Zugang zu Facebook-Profilen erlangt und mindestens 30 Millionen psychographische Profile erstellt. Diese Profile wurden anschließend genutzt, um Wähler:innen gezielt Wahlwerbung und Desinformation zu präsentieren, die auf die individuellen Persönlichkeitsprofile zugeschnitten war.[136] Diese Praxis fand nicht nur in den USA Anwendung, sondern auch in anderen Wahlkämpfen in der Karibik und Afrika.[137] In Deutschland wird ebenfalls allen großen Parteien vorgeworfen, Microtargeting auf der Plattform Facebook zu betreiben.[138]

Zielgruppenspezifische (Wahl-)Werbung ist zwar kein neues Phänomen, durch die von Nutzer:innen auf sozialen Medien bewusst oder unbewusst preisgegebenen Daten kann sie jedoch präziser, zielgerichteter und in einem deutlich größeren Umfang verbreitet werden als dies bei klassischen Medien möglich ist. Die Datenschutzgrundverordnung schränkt zwar die Möglichkeiten ein, sensible Daten für Microtargeting zu verwenden. Es stehen jedoch aktuell Vorwürfe im Raum, dass

134 Vgl. European Commission 2022a
135 Vgl. European Commission 2022b
136 Vgl. Dachwitz/Rudl 2018
137 Vgl. Baker et al. 2018
138 Vgl. TargetLeaks 2024

gegen diese Bestimmungen verstoßen wird. Beispielsweise wird der Europäischen Kommission selbst vorgeworfen, Microtargeting auf Basis sensibler Daten auf der Plattform X (vormals Twitter) zu betreiben.[139]

Problemtisch ist in diesem Kontext, dass mögliche Wahlversprechen und politische Standpunkte, die Parteien gegenüber spezifischen Gruppen machen, der Allgemeinheit nicht bekannt sind und diese sich möglicherweise widersprechen. Die daraus potenziell erwachsenden unterschiedlichen Erwartungen der Wählergruppen können zu Unzufriedenheit und Polarisierung führen.[140] Die Verbreitung von Desinformation durch oder über Kandidat:innen für politische Ämter oder Parteien stellt eine potenzielle Gefahr für demokratische Prozesse dar, welche durch eine an psychographische Profile maßgeschneiderte Verbreitung zusätzlich verstärkt wird.[141]

3.3.4 Effekte durch Desinformation

Da die Untersuchungen zu den Effekten von Desinformation noch vergleichsweise neu sind, gibt es noch keinen wissenschaftlichen Konsens darüber, wie groß der Einfluss von Desinformationskampagnen beispielsweise auf die Wahrnehmung gesellschaftlicher Prozesse oder auf das Wahlverhalten von Bürger:innen ist. Im Folgenden werden erste Erkenntnisse über die Effekte dargestellt und eingeordnet.

Einflussnahme auf das Wahlverhalten

Digitale Plattformen gewähren nur eingeschränkten Zugang zu ihren Daten, was die genaue Bestimmung der Effekte politischer Kampagnen auf diesen Plattformen erschwert.[142] Wahlergebnisse hängen von zahlreichen Einflussfaktoren ab, die unabhängig von der Effektivität zielgerichteter Werbekampagnen auf sozialen Plattformen wirken. Zudem sind zuverlässige Daten über das Wahlverhalten begrenzt verfügbar. Der Effekt von digitaler Wahlwerbung und von Social Bots kann daher nur punktuell nachgewiesen und die systemische Wirkung nicht abschließend geklärt werden

Die empirische Studie von Gregory Eady et al. liefert Erkenntnisse bezüglich der Verbreitung von russischer Desinformation im Kontext der US-Wahlen im Jahr 2016 auf der Plattform X (vormals Twitter). Die Untersuchung zeigt eine signifikante Konzentration der Desinformationsinhalte. Ein Prozent der Nutzer:innen, vornehmlich Republikaner:innen, machten 70 Prozent der Exposition aus. Bei der Mehrheit wurde die Kampagne von Inhalten inländischer Nachrichtenmedien und Politiker:innen in den Hintergrund gedrängt. Die Untersuchung bietet keine Anhaltspunkte für einen substanziellen Zusammenhang zwischen der Exposition gegenüber der russischen Desinformationskampagne und nachweisbaren Veränderungen in Einstellungen, Polarisierung oder Wahlverhalten.[143]

In einer Untersuchung aus dem Jahr 2012 wurde nachgewiesen, dass Appelle zur Teilnahme an Wahlen auf sozialen Plattformen eine positive Wirkung entfalten. Es zeigte sich, dass nicht nur signifikant mehr Personen, die den Aufruf erhielten, tatsächlich an der Wahl teilnahmen, sondern auch deren Freund:innen und Bekannte. Eine einzelne Nachricht auf Facebook entfaltete demnach eine sehr große Reichweite, sie konnte 340.000 Menschen zur Teilnahme an der Wahl zu motivieren.[144]

139 Vgl. Noyb 2023a; Noyb 2023b
140 Vgl. Noyb 2023c
141 Vgl. Lima-Strong 2021
142 Vgl. Redoano 2019
143 Vgl. Eady et al. 2023
144 Vgl. Bond et al. 2023

Ben Tappin et al. führten im Jahr 2023 ein Experiment durch, in dem sie nachweisen konnten, dass ihre konzipierte Microtargeting-Strategie einen signifikant stärkeren Effekt erzielte als zwei alternative Strategien. Dennoch limitieren die Autor:innen die Auswirkungen von Microtargeting erheblich. Der Erfolg der Strategie ist beispielsweise stark abhängig vom politischen Thema und Kontext. Die Studie zeigt generell, dass das Targeting anhand einzelner Attribute, wie beispielsweise der Parteizugehörigkeit, äußerst effektiv ist. Hingegen fällt der zusätzliche Vorteil von Microtargeting, das multiple Kriterien wie Alter oder ideologische Ausrichtung einbezieht, vergleichsweise gering aus. Die Autor:innen betonen zudem, dass aufgrund der Komplexität des Themas und der begrenzten Verfügbarkeit von Daten zum Wahlverhalten die Untersuchung von Microtargeting im politischen Kontext erheblich herausfordernder ist als beispielsweise im Bereich von Produktverkäufen.[145]

Michela Redoano gelangt zu dem Schluss, dass Microtargeting bei der Präsidentschaftswahl 2016 äußerst effektiv war, insbesondere wenn anhand der Kriterien Ideologie, Geschlecht und Bildungsniveau differenziert wurde.[146] Eine ähnliche Schlussfolgerung ziehen auch Federica Liberini et al. in einem Working Paper. Die Ergebnisse ihrer Untersuchung verdeutlichen, dass soziale Medien effektiv genutzt werden können, um Schlüsselgruppen von Wähler:innen zu beeinflussen. Diese Erkenntnisse liefern Evidenzen dafür, dass bedeutende politische Ereignisse wie der Brexit und die Wahl von Donald Trump weitgehend auf einen effektiven Einsatz von Datenanalysen zurückgeführt werden können. In ihrer Analyse schätzten die Autor:innen den Nettoeffekt von Kampagnen auf politische Ergebnisse. Die Resultate zeigen, dass Facebook-Werbung erfolgreich darin war, Trumps Hauptunterstützer:innen dazu zu bewegen, am Wahltag ihre Stimme abzugeben. Gleichzeitig hatte diese Werbung einen negativen Einfluss auf die Wahlbeteiligung der Liberalen, den Kernanhänger:innen der alternativen Kandidatin Hillary Clinton. Demobilisierung von Wähler:innen des anderen politischen Lagers kann selbstverständlich auch das Ziel einer Wahlkampagne sein. Zudem beeinflusste die Facebook-Werbung moderate und weniger informierte Wähler:innen dazu, für Trump zu stimmen. Diese Ergebnisse stehen im Einklang mit Berichten, welche zeigen, dass Trump im Vergleich zu Clinton mehr Mittel für die Nutzung sozialer Medien im Wahlkampf eingesetzt hat. Ein anderer Grund dafür könnte sein, dass seine Botschaften möglicherweise besser konzipiert waren.[147]

Eine gemeinsame Studie der Landesmedienanstalten Nordrhein-Westfalen, Bayern, Berlin-Brandenburg sowie Rheinland-Pfalz untersuchte den Effekt von Microtargeting in Deutschland auf die Ergebnisse der Europawahl 2019. Die zentrale Schlussfolgerung dieser Untersuchung ist, dass der Effekt von Microtargeting bei diesen Wahlen vergleichsweise gering war. Allerdings schließt die Studie nicht aus, dass deutlichere Effekte erzielt werden können, wenn die Microtargeting-Kampagnen adäquat konzipiert sind und sowohl in die Breite als auch in die Tiefe wirken.[148]

Einflussnahme von Social Bots

Die Ergebnisse der oben zuletzt genannten Studie deuten darauf hin, dass die Alternative für Deutschland (AfD) möglicherweise weniger auf Microtargeting im Sinne von Wahlwerbung setzt, sondern eher auf Einflussnahme durch Social Bots.

Bei der Analyse der Daten fällt auf, dass die AfD vergleichsweise wenige Anzeigen auf sozialen Plattformen schaltet. Gleichzeitig veröffentlicht die Partei zahlreiche Beiträge, die in ihrer Gestaltung stark an politische Werbung erinnern. Es gibt Anhaltspunkte dafür, dass bei diesen Aktivitäten auto-

145 Vgl. Tappin et al. 2023
146 Vgl. Redoano 2019
147 Vgl. Liberini et al. 2020
148 Vgl. Hegelich / Serrano 2019

matisierte Bots eingesetzt werden.[149] Es konnte beobachtet werden, dass hyperaktive Accounts eine systematische Interaktion mit jedem Post auf den AfD-Seiten zeigen, was zu einer Verzerrung des politischen Online-Diskurses führen kann.[150] Es könnte sein, dass Anhänger:innen der Partei diese Interaktionen aus politischer Überzeugung unentgeltlich generieren oder dass eine Investition unbekannten Ursprungs in eine Infrastruktur zur Förderung von Social-Media-Interaktionen erfolgt ist.

Auf der Plattform Facebook beispielsweise stellt dieses Vorgehen ein Problem dar, da die Strategie, die Interaktionsanzahl künstlich zu steigern, wirksam ist. Facebook hat seinen Newsfeed so konfiguriert, dass Meaningful Interactions die entscheidende Größe ist, die der Algorithmus zu optimieren versucht. Praktisch bedeutet dies, dass Posts, die mehr Reaktionen hervorrufen als wichtiger eingestuft werden, was dazu führt, dass sie vermehrt im Newsfeed erscheinen. Die genaue Auswirkung sowie der Umfang der Strategie der AfD ist jedoch noch nicht abschließend geklärt.[151] Die Aufdeckung der Aktivitäten von Social Bots ist ebenso komplex wie der Nachweis ihrer Wirkungen.[152]

In diesem Kontext ist ebenfalls relevant, dass Misinformation und Desinformation, die von Politiker:innen verbreitet wird, glaubhafter erscheinen und daher potenziell mehr Schaden anrichten können, als wenn diese Inhalte von gewöhnlichen Nutzer:innen stammen.[153]

3.4 Chancen und Risiken – Zusammenfassung

KI bietet zahlreiche Chancen im Medienbereich, es erwachsen aus dem Einsatz aber auch Risiken. Tabelle 3 gibt einen Überblick über die Chancen und Risiken, die sich aus den verschiedenen KI-Technologien ergeben.

ML-Algorithmen können dazu genutzt werden, gesuchte Inhalte schneller zu finden und zielgerichteter auf die Interessen der Medienkonsument:innen anzupassen. Dabei können insbesondere in sozialen Medien Inhalte, die Hass und Anfeindungen enthalten, herausgefiltert werden. Gleichzeitig kann eine derartige (Vor-)Auswahl von Inhalten die Entstehung von Filterblasen und Echokammern begünstigen und im schlimmsten Fall sogar zu einer Zensur führen.

Insbesondere im Bereich des investigativen Journalismus bieten KI-basierte Analysewerkzeuge die Möglichkeit, komplexe Sachverhalte mit großen Datenmengen systematisch zu analysieren. Aufgrund der Komplexität der Analysen ist der Wahrung der journalistischen Sorgfaltspflicht besondere Beachtung zu schenken. Diese Werkzeuge können darüber hinaus dazu beitragen, Media Bias zu verringern, indem sie in kurzer Zeit eine diverse Grundlage an Daten und Informationen liefern. Wichtig dabei ist, die Methoden und verwendeten Werkzeuge transparent zu machen (siehe dazu auch Kapitel 4.3.2).

Die wahrscheinlich größten Veränderungen haben und werden sich durch generative KI ergeben. Die Anwendungspotenziale reichen von der Übernahme von Routineaufgaben bis hin zur Unterstützung bei kreativen Arbeiten. Diese Werkzeuge können seriöse Medienhäuser darin unterstützen effizienter zu sein und damit auch dazu beitragen, im Wettlauf mit den sozialen Medien kompetitiver zu bleiben. Wie der Gesamteffekt auf die Qualität von Inhalten ausfällt, bleibt abzuwarten. Denn obschon seriöse Medienhäuser vom Einsatz dieser Technologien profitieren, steigt der Output unseriöser Quellen durch den Einsatz von KI um ein Vielfaches schneller an.[154] Umfassende Recherchen und gründliche Prüfungen durch menschliche Mitarbeiter:innen bleiben hierbei schlichtweg aus. Verbindliche rechtliche Rahmenbedingungen sollten möglichst zeitnah geschaffen werden, damit diese Technologien frei von rechtlichen Risiken angewandt werden können.

Seitens der Medienkonsument:innen ergeben sich durch generative KI ebenfalls zahlreiche Chancen. So kann allein schon die Nutzung dieser Werkzeuge die Medienkompetenz stärken, etwa indem persönliche Assistenten oder Lehrkräfte bei verschiedensten Aufgaben unterstützen. Allerdings könnte mit zunehmender Verbreitung dieser Anwendungen die Meinungsvielfalt reduziert werden. Denn während die Ergebnisse klassischer Suchmaschinen auf verschiedenste Seiten mit diversen Perspektiven und Darstellungen verweisen, könnte die Zusammenfassung durch persönliche KI-Assistent:innen deutlich knapper und inhaltlich eingeschränkter erfolgen.

Neben den genannten einzelnen Risiken ergibt sich durch die KI-Anwendungen auch ein systemisches Risiko für demokratische Staaten. Denn sowohl durch ML-Algorithmen als auch generative KI wird versucht, die öffentliche Debatte gezielt zu manipulieren. Der systematische Effekt auf die politische Meinungsbildung bleibt jedoch unklar. Es besteht die Befürchtung, dass das Internet zukünftig von Misinformation und Desinformation verschiedener Akteure mit vielfältigen Zielsetzungen überschwemmt wird. Angesichts der schieren Menge und möglichen Omnipräsenz hochwertiger Desinformation wird es für Rezipient:innen zunehmend schwierig, die Faktizität von Inhalten korrekt zu beurteilen. Die derzeit größte Schwäche von generativer KI – die Ermangelung einer Überprüfung der verfassten Inhalte auf Faktizität – wird bei der automatisierten Erstellung von Desinformation zu einer ungewollten Stärke.

Hyperaktive Accounts können den Diskurs auf sozialen Medien in eine Richtung drängen, die nicht der Meinung der Mehrheit entspricht. Mittels Microtargeting lassen sich Botschaften gezielt an Gruppierungen richten, die spezifische Eigenschaften bzw. Meinungen teilen. Deepfakes können dazu genutzt werden, um Falschinformationen zu verbreiten, gleichermaßen aber auch, um reale Geschehnisse als Fälschungen abzutun.

Die Entwicklung im KI-Bereich lässt sich nicht mehr umkehren. Damit die Potenziale gehoben werden können und gleichzeitig die Risiken kontrollierbar bleiben ist der proaktive Umgang mit diesen Technologien von entscheidender Bedeutung. In Kapitel 6 werden Handlungsempfehlungen dargelegt, wie dieses Ziel erreicht werden kann.

149 Vgl. Serrano et al. 2019
150 Vgl. Papakyriakopoulos et al. 2019
151 Vgl. Hegelich / Serrano 2019
152 Vgl. Martini et al. 2021
153 Vgl. Redoano 2019
154 Vgl. Hanley / Durumeric 2023

KI-Technologie	Chancen	Risiken
Vorschlagsalgorithmen (ML)	• Vereinfachtes Finden von Inhalten • Personalisierung von Inhalten • Schutz vor Hass und Anfeindungen in Sozialen Medien	• Filterblasen / Echokammern • Microtargeting • Zensur • Manipulation des öffentlichen Diskurses durch hyperaktive Accounts
Analyse- und Recherchewerkzeuge (ML)	• Reduzierung des Media Bias • Möglichkeit komplexe Datenanalysen durchzuführen	• Gefahr für journalistische Sorgfalt • Fehlende Transparenz der Algorithmen
Generative KI	• Arbeitsunterstützung sowohl im Routine- als auch im Kreativbereich • Steigerung der Medienkompetenz durch KI-Anwendung • Persönliche Assistent:innen und Lehrkräfte • Steigerung der Wettbewerbsfähigkeit von seriösen Medienhäusern	• (Teil-)automatisierte Erzeugung von Desinformation (Text, Bild, Sprache) • Deepfakes (z. B. Desinformation, Identitätsdiebstahl, Liars Dividend) • Konsequenzen für die Meinungsvielfalt • Rechtliche Herausforderungen
	Effekt auf die Qualität von Inhalten kann variieren	

Tabelle 3: Überblick über die Chancen und Risiken zentraler KI-Anwendungen. Quelle: Eigene Darstellung

4 Regulatorische und technologische Rahmenbedingungen

Medienkompetenz muss von den Bürger:innen angestrebt und erworben werden. Dies stellt angesichts der oben beschriebenen Entwicklungen eine große Herausforderung dar. Neben direkten Maßnahmen zur Steigerung der Medienkompetenz können medienpolitische und regulatorische Vorhaben sowie technische Angebote eine unterstützende Funktion erfüllen, um die Herausforderungen zu verringern und eine Überforderung der Bürger:innen zu verhindern. In diesem Kapitel werden zentrale politische Rahmenbedingungen (Kapitel 4.1) beschrieben sowie die Rolle von Selbstverpflichtungen und Standards (Kapitel 4.2) diskutiert. Anschließend werden technologische Maßnahmen zur Erkennung und Kennzeichnung von KI-Inhalten sowie zur Verifikation bzw. Authentifikation von menschlich erstellten Inhalten vorgestellt (Kapitel 4.3).

4.1 Politische Rahmenbedingungen

Zwei aktuelle Regulierungen der Europäischen Union (EU) sind zentral für das in diesem Gutachten diskutierte Thema: Der Digital Services Act (DSA) und der Artificial Intelligence Act (AI Act). Zusätzlich spielt auch die 2018 in Kraft getretene Datenschutzgrundverordnung (DSGVO) eine wichtige Rolle. Die Inhalte dieser Verordnungen werden im Folgendem kurz zusammengefasst.

Digital Services Act (DSA)

Der DSA wurde im Oktober 2022 veröffentlicht und hat seit Februar 2024 in der gesamten EU Geltung. Für die größten Betreiber im Bereich Online-Plattformen und Suchmaschinen (über 45 Millionen monatliche Nutzer:innen) gilt diese Regelung bereits seit August 2023 (hierunter fallen zum Beispiel X, vormals Twitter, und Facebook). Der DSA verfolgt als zentrales Ziel die Steigerung der Transparenz bei großen Plattformen, die digitale Dienste anbieten, und die Übernahme von Verantwortung sowie Haftbarkeit für gesellschaftliche Risiken, die aus ihrer Tätigkeit resultieren.

Konkret setzt der DSA klare Vorgaben in verschiedenen Bereichen: Plattformbetreiber sind dazu verpflichtet, die Funktionsweise ihrer algorithmischen Empfehlungssysteme offen zu legen und ihren Einfluss auf die Plattforminhalte, einschließlich Content Moderation, zu erklären. Zudem muss es Nutzer:innen möglich sein, Informationen darüber zu erhalten, aufgrund welcher Parameter ihnen personalisierte Werbung angezeigt wird, und wie sie diese Parameter beeinflussen können. Der DSA verbietet Werbung, die sich gezielt an Kinder richtet, ebenso wie die Nutzung sensibler Merkmale wie sexueller Orientierung oder religiöser Überzeugung zur Profilbildung. Anbieter digitaler Dienste müssen illegale Inhalte unverzüglich entfernen, sobald sie darauf aufmerksam gemacht werden.

Der DSA betrachtet große Plattformen als potenzielle gesellschaftliche Bedrohungen und identifiziert beispielsweise Risiken für den öffentlichen Diskurs und die faire Durchführung von Wahlen. Daher werden diese Plattformbetreiber dazu verpflichtet, regelmäßige formalisierte Prüfungen der von ihnen ausgehenden Risiken durchzuführen. Dabei ist nicht nur eine Selbstbewertung vorgesehen, sondern es wird auch der Zugang zu Daten für unabhängige Auditoren festgelegt.

In der Umsetzung des DSA gibt es noch viele ungeklärte Punkte, die von verschiedenen Akteuren adressiert werden. Schwierigkeiten ergeben sich unter anderem daraus, dass einige der Forderungen des DSA, wie beispielsweise die von den Plattformen geforderten Risikobewertungen, nicht klar definiert sind. Darüber hinaus muss jeder Mitgliedsstaat DSA-Bevollmächtigte ernennen, das

heißt Aufsichtsbehörden, die die Umsetzung des DSA in dem jeweiligen Staat für kleinere Anbieter digitaler Dienste regelt. Hierbei gibt es die Befürchtung, dass bereits bestehende nationale Aufsichtsstrukturen durch diese Prozesse geschwächt werden.[155]

Die Umsetzung des DSA in Deutschland wird durch das Digitale-Dienste-Gesetz durchgeführt. Hierbei bietet sich auch die Chance, teilweise veraltete Terminologie, die bisher im Medienrecht verwendet wird, zu ändern und auf die aktuellen Gegebenheiten anzupassen.

Artificial Intelligence Act (AI Act)
Der AI Act ist eine geplante EU-Verordnung zur Regulierung von Anwendungen, die auf KI basieren. Die EU legt damit eine der ersten Verordnungen zur Regulierung von KI vor. Der erste Entwurf für den AI Act wurde von der EU-Kommission im April 2021 veröffentlicht und hat seither unter anderem aufgrund der Entwicklungen im Bereich generativer KI bedeutende Anpassungen erfahren. Nach den erfolgreichen Trilogverhandlungen im Dezember 2023 steht aktuell die Verabschiedung im EU-Parlament aus. Zwei Jahre nach der Verabschiedung der endgültigen Fassung, also voraussichtlich 2026, kann der AI Act in Kraft treten.

Der AI Act setzt auf einen risikobasierten Ansatz, der sich auf vier Risikoklassen stützt. Jede dieser Klassen ist mit spezifischen Auflagen bezüglich Risikobewertung, Dokumentation und Monitoring verbunden. Die KI-Technologien werden dabei nicht abstrakt, sondern im Kontext ihres jeweiligen Einsatzgebiets bewertet. Die vier Risikoklassen sind wie folgt definiert:

KI-Systeme mit inakzeptablem Risiko: KI-Systeme in dieser Risikoklasse werden nach dem AI Act verboten. Darunter fallen beispielsweise Anwendungen, die eine biometrische Identifizierung in Echtzeit in öffentlich zugänglichen Räumen ermöglichen; Anwendungen, die Menschen auf intransparente Art und Weise manipulieren oder ihnen physisch schaden können; sogenannte Social-Scoring-Anwendungen, die Menschen auf Basis ihres Sozialverhaltens bewerten und daraus Handlungen ableiten.

KI-Systeme mit hohem Risiko: Diese Klasse umfasst Systeme, die ein Risiko für die Gesundheit, Sicherheit oder Grundrechte von Personen darstellen, wie KI-Systeme in kritischer Infrastruktur, Bewerbungsprozessen oder solche, die eine biometrische Identifikation durchführen. Für Hochrisiko-Systeme werden Pflichten wie die Einrichtung eines Risikomanagementsystems, Dokumentationspflichten, transparente Informationsbereitstellung für Nutzer:innen, Mindestanforderungen an Genauigkeit, Robustheit und Cybersicherheit sowie Überwachung durch menschliches Personal formuliert.

KI-Systeme mit begrenztem Risiko: Bei KI-Systemen, die in diese Klasse fallen, muss sichergestellt werden, dass für die Nutzer:innen aus der Interaktion mit dem System klar wird, dass sie mit einem KI-System interagieren. Beispiele für Anwendungen in dieser Risikoklasse sind Systeme zur Emotionserkennung oder biometrische Kategorisierungssysteme.

155 Vgl. zum Beispiel: Die Medienanstalten 2021b; Algorithmwatch 2022; Mast et al. 2023

KI-Systeme mit niedrigem Risiko: Anwendungen in dieser Klasse unterliegen im AI Act keinerlei rechtlichen Anforderungen. Hierbei handelt es sich vor allem um technische Anwendungen ohne direkte Interaktion mit Menschen, wie beispielsweise Spamfilter oder Systeme für vorausschauende Instandhaltung (Predictive Maintenance). Die Einstufung als Anwendung mit niedrigem Risiko erfordert eine technische Dokumentation und Risikobewertung gemäß den Bestimmungen des AI Act.

Expert:innen verbinden mit dem AI Act die Hoffnung, dass mit dieser Verordnung ein Standard geschaffen wird, der auch Ausstrahlungskraft auf andere Teile der Welt haben kann. Dies wäre insbesondere dann eine begründete Hoffnung, wenn der AI Act darin erfolgreich ist, bei Nutzer:innen Vertrauen in KI-Anwendungen zu schaffen und damit zu einer breiteren Akzeptanz der Technologie zu führen.

Um dieses Vertrauen zu stärken, sind weitere Schritte notwendig, darunter beispielsweise die klare und gut sichtbare Kennzeichnung von Anwendungen, die den Standards des AI Act entsprechen oder überprüft wurden (siehe Kapitel 4.3). Diese Maßnahmen sollen dazu beitragen, dass Nutzer:innen transparente Informationen über die Konformität von KI-Anwendungen erhalten.

Gleichzeitig gibt es allerdings auch Kritik an der Regulierung des AI Act. Insbesondere äußert sich diese Kritik in der Befürchtung, dass die Regelungen die Innovationsmöglichkeiten in der Europäischen Union erheblich einschränken könnten. Es wird betont, dass ein ausgewogenes Verhältnis zwischen Regulierung und Innovationsförderung gefunden werden muss, um sicherzustellen, dass die EU weiterhin eine treibende Kraft im Bereich der KI bleiben kann. Nichtregierungsorganisationen wie AlgorithmWatch kritisieren darüber hinaus, dass einige kritische Hochrisikosysteme, die in den meisten Fällen verboten sind, „durch die Hintertür“ ermöglicht werden, wenn es um die nationale Sicherheit oder die Strafverfolgung geht.[156]

Die praktische Umsetzung des AI Act steht vor zahlreichen Herausforderungen. Eine entscheidende Hürde besteht in der korrekten Zuordnung von Anwendungen oder Systemen zu den jeweiligen Risikoklassen. Eine Studie von appliedAI verdeutlicht, dass diese Zuordnung in der aktuellen Ausführung des AI Act oft nicht eindeutig abgeleitet werden kann.[157] Die Identifizierung geeigneter Institutionen, die die Prüfung und Kategorisierung dieser Zuordnungen übernehmen, wird unerlässlich sein. Einige solcher Institutionen befinden sich bereits in der Entwicklungsphase.

Die Herausforderungen für betroffene Unternehmen sind ebenfalls nicht zu unterschätzen. Es wird erforderlich sein, interne Strukturen für die Risikobewertung, Dokumentation und weitere erforderliche Maßnahmen aufzubauen, um den Anforderungen des AI Act gerecht zu werden.

Darüber hinaus beinhaltet der AI Act eine Vielzahl unbestimmter Rechtsbegriffe, wie beispielsweise den Begriff der Manipulation, deren genaue Auslegung erst im Zuge der Anwendung der Verordnung durch Behörden oder vor Gerichten präzisiert werden wird. Diese Unschärfen können zu rechtlichen Unsicherheiten führen und erfordern eine kontinuierliche Anpassung und Klärung im Verlauf der praktischen Umsetzung des Gesetzes.

Parallel zu den Fortschritten in der EU werden auch in anderen internationalen Organisationen, wie den Vereinten Nationen (UN), der G7[158] und der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD), sowie in den Staaten mit den meisten relevanten Technologieunternehmen,

156 Vgl. zum Beispiel: Müller/Spielkamp 2023
157 Vgl. appliedAI 2023
158 Vgl. G7 2023

insbesondere den USA und China[159], Leitlinien für den Umgang mit KI vorgeschlagen. Ein wiederkehrender Schwerpunkt in diesen Forderungen liegt dabei oft auf der Entwicklung von vertrauenswürdiger KI. Die Bemühungen um vertrauenswürdige KI erstrecken sich somit über globale Ebenen und reflektieren das gemeinsame Interesse daran, ethische und verantwortungsbewusste Standards für den Einsatz von KI zu etablieren.

Datenschutzgrundverordnung (DSGVO)
Seit 2018 regelt die DSGVO die Verarbeitung personenbezogener Daten in der Europäischen Union. Die DSGVO hat damit auch einen erheblichen Einfluss auf den Medienbereich und insbesondere Social-Media-Plattformen, da sich aus der DSGVO heraus eine Vielzahl an Verpflichtungen gegenüber ihren Nutzer:innen ableiten lässt.

Einwilligung und Transparenz
Die DSGVO verlangt von Medienunternehmen, klare und verständliche Informationen darüber bereitzustellen, wie personenbezogene Daten gesammelt, verarbeitet und genutzt werden. Die Nutzer:innen müssen aktiv und informiert ihre Einwilligung geben, bevor ihre Daten verwendet werden können. Dieser Grundsatz fördert nicht nur deren Bewusstsein für Datenschutz, sondern zwingt auch Medienunternehmen dazu, transparente Praktiken zu implementieren, was die Medienkompetenz stärken kann.

Recht auf Vergessenwerden
Die DSGVO gewährt Personen das Recht, die Löschung ihrer personenbezogenen Daten zu verlangen. Dies kann sich auf die Archivierung von Inhalten und Suchergebnissen auswirken. Medienunternehmen müssen daher sicherstellen, dass sie effektive Mechanismen zur Umsetzung dieses Rechts implementieren, was gleichzeitig die Sensibilität für den Schutz der Privatsphäre erhöht.

Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen
Ein weiterer wichtiger Aspekt der DSGVO ist die Forderung nach Datenschutz durch Technikgestaltung (Privacy by Design) und datenschutzfreundlichen Voreinstellungen (Privacy by Default). Medienunternehmen müssen Datenschutzprinzipien bereits bei der Entwicklung neuer Dienste und Anwendungen berücksichtigen. Dies fördert die Integration von Datenschutz in den Kern von technologischen Innovationen im Medienbereich.

Verarbeitung von besonderen Kategorien personenbezogener Daten
Besonders geschützte Datenkategorien wie politische Meinungen, religiöse Überzeugungen oder Gesundheitsdaten dürfen nur unter strengen Voraussetzungen verarbeitet werden. Dies soll verhindern, dass gezielte Werbung (Microtargeting, siehe Kapitel 3.4) auf Basis sensibler Daten durchgeführt wird.

Die Datenschutzbehörden können auf Basis der DSGVO deutlich höhere Strafen aussprechen, wenn wie im Fall von Cambridge Analytica die Einwilligung zur und die Information über die Datennutzung fehlen sowie besonders geschützte personenbezogene Daten wie die politische Meinung von Plattformnutzer:innen verwendet werden.[160] Aktuell wird diese Frage beispielsweise in einem

159 Vgl. Ye 2023
160 Vgl. Dachwitz / Rudl 2018

Fall verhandelt, in dem auf X (vormals Twitter) politische Werbung unterschiedlichen Plattformnutzer:innen gezeigt wurde, die aufgrund ihrer politischen oder religiösen Überzeugungen ausgewählt wurden.[161]

Einfluss auf die Medienkompetenz
Die regulatorischen Maßnahmen der Europäischen Union beeinflussen die Medienkompetenz der Nutzer:innen zwar nur indirekt, eröffnen jedoch durch die umfangreichen Forderungen nach Transparenz und Selbstbestimmung bedeutende Möglichkeiten zur Förderung der Medienkompetenz. Insbesondere die Auflage an Plattformbetreiber, die Funktionsweise ihrer Vorschlagsalgorithmen sowie die Parameter für personalisierte Werbung offenzulegen, bietet eine Gelegenheit für einen reflektierteren Umgang mit dieser Technologie.

Diese zur Verfügung gestellten Informationen ermöglichen nicht nur Nutzer:innen einen bewussteren Umgang, sondern können auch im Rahmen von Fortbildungs- oder Informationsveranstaltungen verwendet werden, um die Funktionsweise der Plattformen anschaulich zu vermitteln. Die Bereitstellung von Einblicken in die Mechanismen digitaler Dienste trägt somit dazu bei, dass Nutzer:innen besser informiert sind und ihre Medienkompetenz gezielt stärken können.

Gleichzeitig ist mit den regulatorischen Maßnahmen die Hoffnung verbunden, eine Überforderung der Nutzer:innen zu verhindern, indem klare und nachvollziehbare Regeln festgelegt werden, die das Vertrauen in KI-Anwendungen steigern und einige der Herausforderungen, die durch die neue Technologie entstehen, bereits adressieren, bevor die Nutzer:innen mit ihnen konfrontiert werden.

4.2 Selbstverpflichtungen / Standards

In Anbetracht der digitalen Transformation und der verstärkten Integration von KI in das gesellschaftliche Leben haben zahlreiche Organisationen und Unternehmen, insbesondere auch Medienhäuser, eigenständige Richtlinien für den Umgang mit KI etabliert. Diese gehen über die bestehenden regulatorischen Rahmenbedingungen hinaus und reflektieren eine proaktive Auseinandersetzung mit der gesellschaftlichen Verantwortung. Die Anwendung von Instrumenten der Selbstregulierung wird als unerlässlich erachtet, um eine erfolgreiche Gestaltung der digitalen Transformation zu gewährleisten.

Für Akteure in der Medienbranche ist es von zentraler Bedeutung, klare Positionen im Umgang mit KI einzunehmen, da diese Technologie die Arbeitsweise der Medien grundlegend verändert. Spezifische Richtlinien für den Umgang mit KI in der journalistischen Arbeit spielen dabei eine entscheidende Rolle. Sie dienen nicht nur als Orientierung für Anwender:innen, sondern sind vor allem essenziell, um das Vertrauen in die Medien zu erhalten und im Idealfall zu stärken. Eine erste Studie von Simon et al. analysiert Richtlinien, die sich Nachrichtenorganisationen in verschiedenen Ländern der Welt selbst gegeben haben und kommt zu dem Schluss, dass deren Inhalte bereits zu einigen zentralen Themen konvergieren, insbesondere der Notwendigkeit für Transparenz und menschliche Überwachung beziehungsweise Entscheidungskompetenz.[162] Diese Themen spiegeln sich beispielsweise auch in der Richtlinie wider, die sich die Deutsche Presse-Agentur (dpa) im Umgang mit KI gegeben hat.[163]

161 Vgl. Noyb 2023b
162 Vgl. Simon et al. 2023a
163 Vgl. Raabe 2023

Eine bemerkenswerte Form der Selbstverpflichtung ist die Einigung zwischen der Writers Guild und den Autor:innen in Hollywood.[164] Diese Vereinbarung erlaubt zwar die Nutzung von KI-Werkzeugen als unterstützendes Mittel für Autor:innen, stellt jedoch sicher, dass die KI deren Aufgaben nicht vollständig übernimmt. Dieses Modell könnte als wegweisendes Beispiel dienen, dem auch andere Medienschaffende und Kreativberufe folgen können. Es illustriert die Bedeutung einer ausbalancierten Integration von KI, bei der menschliche Expertise und Entscheidungsfähigkeiten weiterhin zentral bleiben.

Auf der europäischen Ebene hat das *Steering Committee on Media and Information Society* des Europarats im November 2023 Leitlinien zur verantwortungsbewussten Umsetzung von KI-Systemen im Journalismus verabschiedet.[165] Diese Leitlinien bieten praktische Anleitungen für die relevanten Akteure, insbesondere für Nachrichtenorganisationen, aber auch für Staaten, Technologieanbieter und digitale Plattformen, die Nachrichten verbreiten. Es wird detailliert erläutert, wie KI-Systeme eingesetzt werden sollten, um die Produktion von journalistischen Medieninhalten in verschiedenen Phasen zu unterstützen: angefangen bei der Entscheidung über den Einsatz von KI-Systemen über den Erwerb von KI-Werkzeugen, die Integration von KI-Systemen in die berufliche Praxis bis hin zur externen Dimension der Verwendung von KI in Redaktionen, insbesondere deren Auswirkungen auf das Publikum und die Gesellschaft. Die Leitlinien schlagen außerdem bestimmte Verantwortlichkeiten für Technologieanbieter und Plattformen sowie für Mitgliedsstaaten vor. Letztere können eine wichtige Rolle bei der Entwicklung von Standards für den verantwortungsbewussten Einsatz von KI spielen und Unterstützung bieten, einschließlich finanzieller Förderprogramme für die Entwicklung verantwortungsbewusster KI-Systeme im Journalismus.

Einen weiteren Vorstoß in Richtung anwendbarer und sektorenübergreifender Standards im Bereich der vertrauenswürdigen KI, verfolgt das Projekt Mission KI. Mission KI ist ein Projekt des Bundesministeriums für Digitales und Verkehr und hat das Ziel den Transfer von KI-Innovationen aus der Forschung in den Markt zu beschleunigen. Durch Mission KI werden unter anderem Prüfverfahren für vertrauenswürdige KI entwickelt und anhand von spezifischen Anwendungsfällen wie beispielsweise Empfehlungssystemen oder Chatbots erprobt. Das Projekt skizziert und testet somit ein Qualitätsversprechen für KI-Anwendungen im niedrigen Risikobereich, das marktfähig und anwendungsfallbasiert arbeitet sowie in verschiedenen Sektoren nutzbar ist. Für Unternehmen, die KI in einem niedrigen Risikobereich einsetzen – wie es zum Beispiel bei Nachrichtenorganisationen in der Regel der Fall ist –, bietet diese Selbstverpflichtung die Möglichkeit, sich klar zu positionieren sowie Wissen und Erfahrung im Bereich vertrauenswürdiger KI zu sammeln.

4.3 Kennzeichnung und Transparenz

Kennzeichnung stellt eine essenzielle Maßnahme dar, die neben zahlreichen weiteren positiven Effekten die Förderung der Medienkompetenz ermöglicht und die Entlastung der Bürger:innen im Umgang mit (möglicherweise) KI-erstellten Inhalten zum Ziel hat. Eine präzise und deutliche Kennzeichnung ermöglicht es den Bürger:innen, ohne erheblichen Aufwand sehr schnell und leicht zu erkennen, welche Anwendungen KI beinhalten und ob Inhalte mithilfe oder vollständig von KI erstellt wurden. Optional ist dabei die Information, wie die KI in Bezug auf diverse Werte wie Transparenz, Fairness oder Erklärbarkeit zu bewerten ist.

164 Vgl. Dampz 2023
165 Vgl. Council of Europe 2023

Eine wirkungsvolle Kennzeichnung trägt zudem dazu bei, dem Risiko der Desinformation entgegenzuwirken. Die Umsetzung von Kennzeichnungen wird dabei notwendigerweise je nach Kontext des jeweiligen Mediums oder der spezifischen Anwendung differieren, da beispielsweise auf digitalen Plattformen andere Informationen als relevant und erkennbar gelten als bei der Kennzeichnung einer käuflichen KI-Anwendung für journalistische Recherchezwecke.

Nachdem die Notwendigkeit von Kennzeichnungen (Kapitel 4.3.1) erläutert wurde, folgt eine Diskussion zur algorithmischen Neutralität (Kapitel 4.3.2) und im Anschluss die Vorstellung bereits bestehender Kennzeichnungen (Kapitel 4.3.3). Das Kapitel endet mit einer kurzen Zusammenfassung und Empfehlung (Kapitel 4.3.4).

4.3.1 Zur Notwendigkeit von Kennzeichnung

Durch die enorme quantitative wie auch qualitative Steigerung von KI-erzeugten Inhalten im Medienbereich wird es für Nutzer:innen zunehmend schwierig zwischen solchen synthetischen Medien einerseits und menschlicher Kommunikation und Medienproduktion andererseits zu unterscheiden. Schon heute ist eine solche Unterscheidung bei geschriebenen Texten für Laien kaum möglich. Die Kompetenz diese Unterscheidung treffen zu können, ist jedoch für den Aufbau kritischer Technikkompetenz essenziell. Zudem wird durch diese Unterscheidung die Integrität menschlicher Kommunikation gewahrt. Es wird daher argumentiert, dass eine Pflicht zur Kennzeichnung von KI-Inhalten geboten ist.[166]

In der Folge werden einige zentrale Vorteile einer Kennzeichnung von KI genauer ausgeführt.

Transparenz als Grundvoraussetzung

Eine erfolgreiche Kennzeichnung von KI-Inhalten ermöglicht Transparenz. Transparenz ist ein grundlegendes Element verschiedenster ethischer KI-Kodizes.[167] Sie bezieht sich nicht nur auf die Offenlegung von Algorithmen und Trainingsdaten, sondern auch auf die Kennzeichnung der Verwendung von KI und die gekennzeichneten Medieninhalte selbst. Transparenz ist eine notwendige Voraussetzung für die Identifikation und Korrektur von Rechtsverletzungen. Sie ist darüber hinaus auch entscheidend, um gesellschaftliche Debatten und den Aufbau von Vertrauen in KI-gestützte Anwendungen zu ermöglichen.

Digitale Inhalte und Medienkompetenz

Die Kennzeichnung von KI-Inhalten, insbesondere in Medien, erweist sich als sinnvoll, da sie es Medienkonsument:innen ermöglicht, bewusster mit digitalen Inhalten umzugehen. Durch die transparente Kennzeichnung erkennen Nutzer:innen, welche Inhalte durch KI generiert oder beeinflusst wurden. Dies fördert ein höheres Maß an Medienkompetenz, da Verbraucher:innen die Ursprünge und möglichen Einflüsse hinter den präsentierten Informationen besser verstehen können. Die Förderung von Medienkompetenz wird somit durch eine klare Kennzeichnung unterstützt und ermöglicht eine verantwortungsbewusste Nutzung digitaler Medien.

166 Vgl. zum Beispiel Heesen 2023
167 Vgl. zum Beispiel AI Ethics Impact Group 2020

Stärkung des Vertrauens in Medienquellen
Die Kennzeichnung von KI-Inhalten trägt zur Stärkung des Vertrauens in Medienquellen bei. Eine klare und eindeutige Markierung von mit KI erstellten Inhalten schafft Transparenz und ermöglicht es den Nutzer:innen, die Glaubwürdigkeit von Nachrichten und Informationen besser einzuschätzen. Automatisch generierte Desinformation kann so beispielsweise schneller und unkomplizierter erkannt werden. Dies ist insbesondere wichtig, um mögliche Bedenken hinsichtlich Manipulation und Authentizität zu adressieren, die mit dem Einsatz von KI in der Medienproduktion verbunden sein können. Vertrauen in Medienquellen ist ein Schlüsselelement für eine funktionierende Demokratie, da die Legitimation des Gemeinwohls aus dem öffentlichen Diskurs hervorgeht.[168]

Vorteil Datenqualität
Große KI-Sprachmodelle basieren in der Regel auf Daten, die aus öffentlich zugänglichen Internetquellen gewonnen werden. Gleichzeitig produzieren diese Modelle selbst Texte, Bilder sowie Audio- und Videobeiträge. Um die Qualität von Trainingsdaten zu sichern und die Reproduktion falscher Informationen zu vermeiden, ist eine eindeutige Kennzeichnung von durch KI erzeugten Inhalten hilfreich. Maschinenlesbare Kennzeichnungspflichten sind dabei entscheidend, um eine selbstreferenzielle Nutzung von KI-Quellen zu verhindern. Zusätzlich ermöglicht eine solche Kennzeichnung eine effektivere und effizientere Filterung von Inhalten auf Plattformen, was insbesondere im Hinblick auf die Qualität und Relevanz von Informationen von hoher Bedeutung ist.

Berücksichtigung journalistischer Standards und Normen
Die Kennzeichnung von KI-Inhalten steht im Einklang mit journalistischer Ethik. Medienproduzierende müssen die Herkunft und den Einfluss von Inhalten verstehen und offenlegen. Die Kennzeichnung unterstützt dies, indem sie den Nutzer:innen klare Informationen darüber gibt, ob es sich um KI-generierte Inhalte handelt. Dies trägt dazu bei, mögliche Verzerrungen zu verhindern und die Integrität des Journalismus zu wahren. Insbesondere in einer Zeit, in der Manipulation und Desinformation durch digitale Medien eine ernste Bedrohung darstellen, ist die transparente Kennzeichnung von KI-Inhalten ein essenzieller Bestandteil ethisch verantwortungsbewusster Medienpraktiken.

Mögliche Nachteile von Kennzeichnung
Neben den oben beschriebenen Vorteilen werden aber auch unterschiedliche Nachteile befürchtet, die sich aus einer Kennzeichnung von KI-Inhalten in Medien ergeben könnten.

Stigmatisierung und Vorurteile
Eine auffällige Kennzeichnung von KI-Inhalten könnte zu Stigmatisierung führen, wenn Nutzer:innen dazu neigen, automatisch anzunehmen, dass mit KI erstellte Inhalte von geringerer Qualität oder Glaubwürdigkeit sind. Dies könnte zu Vorurteilen gegenüber KI-generierten Inhalten führen, selbst wenn sie inhaltlich korrekt und relevant sind.

Manipulation und Umgehung
Wenn die Kennzeichnung von KI-Inhalten zu einfach und zu offensichtlich ist, könnten Produzenten von Inhalten versucht sein, diese zu umgehen oder zu manipulieren. Dies könnte zu einer Verringerung der Wirksamkeit der Kennzeichnung führen und das Vertrauen der Nutzer:innen in die Kennzeichnung insgesamt untergraben.

168 Vgl. Heesen et al. 2021

Komplexität der KI-Erkennung
Es könnte herausfordernd sein, eine eindeutige und standardisierte Kennzeichnung für KI-Inhalte zu entwickeln, insbesondere da KI-Algorithmen ständig weiterentwickelt werden und vielfältige Anwendungen haben. Die Dynamik und Komplexität von KI-Systemen könnten die Erstellung und Umsetzung klarer Kennzeichnungen erschweren.

Verwirrung bei Nutzer:innen
Eine übermäßige Kennzeichnung von KI-Inhalten könnte Nutzer:innen verwirren und dazu führen, dass sie die Bedeutung der Kennzeichnung nicht einordnen können. Dies könnte die beabsichtigte Transparenz beeinträchtigen und zu einer ineffektiven Nutzung der Kennzeichnung führen. Zusätzlich sind Nutzer:innen potenziell mit der Frage konfrontiert, ob die Kennzeichnung authentisch ist, was eine weitere Komplexitätsebene hinzufügt.[169]

Mangelnde Akzeptanz
Nutzer:innen könnten die Kennzeichnung von KI-Inhalten als überflüssig oder einschränkend empfinden, insbesondere wenn die Gründe für die Kennzeichnung nicht ausreichend kommuniziert sind. Dies könnte zu einer geringen Akzeptanz und möglicherweise zu einer Ablehnung der Kennzeichnung führen.

Bei allen potenziellen Nachteilen handelt es sich jedoch um Probleme, die bei einer guten Umsetzung zur Kennzeichnung von KI-Inhalten vermieden werden können, die also keine ausreichende Argumentation bieten, um die oben beschriebene Notwendigkeit einer Kennzeichnung in Abrede zu stellen. Sie sollten jedoch als wichtige Hinweise beim Erstellen sowie bei der Einführung von Kennzeichnungen verstanden werden, um zu einer erfolgreichen Kennzeichnung von KI-Inhalten zu gelangen.

4.3.2 Algorithmische Neutralität & Transparenz

KI-Anwendungen werden oft damit beworben, dass sie objektiver und neutraler als Menschen sind. Es ist jedoch inzwischen aus zahllosen Beispielen klar, dass insbesondere die neueren KI-Anwendungen, die aus dem Training auf Basis von großen Mengen an Daten entstehen, Bias enthalten. Diese Bias können zu Misrepräsentation und Diskriminierung führen. Bias in KI-Anwendungen entsteht vor allem dann, wenn die Daten, die zum Training des KI-Systems benötigt werden, nicht repräsentativ ausgewählt wurden. Wenn beispielsweise historische Daten systematische Ungleichheiten oder Vorurteile enthalten, wird der Algorithmus diese Muster lernen und möglicherweise verstärken.

Beispiele für algorithmische Voreingenommenheit finden sich in verschiedenen Bereichen, einschließlich Strafjustiz, Finanzwesen, Rekrutierung und Gesundheitswesen. Zum Beispiel könnten KI-gesteuerte Bewerbungsscreenings aufgrund historischer Vorurteile gegenüber bestimmten Gruppen unfaire Entscheidungen treffen. Auch für die in diesem Gutachten diskutierten Anwendungen gibt es hierfür Beispiele. Erste Studien zeigen zum Beispiel, dass Sprachmodelle geschlechtsbasierte Bias beinhalten.[170]

169 Vgl. Feng et al. 2023
170 Vgl. AlignedAI 2023

Um algorithmische Voreingenommenheit zu reduzieren, sind verschiedene Maßnahmen notwendig:[171]

Diversität der Datensätze: Es ist wichtig, diverse und repräsentative Datensätze zu verwenden, die verschiedene Gruppen angemessen abbilden.

Transparenz und Erklärbarkeit: Modelle sollten so konzipiert sein, dass ihre Entscheidungen nachvollziehbar sind, damit Entwickler:innen und Nutzer:innen verstehen können, wie und warum bestimmte Entscheidungen getroffen werden.

Überwachung und Evaluierung: Kontinuierliche Überwachung von Modellen in Bezug auf Bias und regelmäßige Überprüfung ihrer Leistung in verschiedenen Bevölkerungsgruppen sind notwendig, um potenzielle Probleme zu identifizieren und zu beheben.

Die vollständige Offenlegung von Daten und genutzten Algorithmen für die Öffentlichkeit birgt aufgrund von Datenschutzbedenken Herausforderungen und ist zugleich nicht zielführend für die Transparenz. Eine mögliche Lösung wäre es, eine externe unabhängige Instanz zu etablieren, welche die Umsetzung der Maßnahmen zur Reduzierung von algorithmischer Voreingenommenheit überprüft. Diese Instanz sollte nicht nur die Implementierung der Maßnahmen überwachen, sondern auch die Effektivität dieser Maßnahmen evaluieren. Eine solche unabhängige Überprüfung kann dazu beitragen, Vertrauen in die KI-Anwendungen zu stärken und sicherzustellen, dass ethische Standards eingehalten werden. Für die Kommunikation an die Nutzer:innen kann mithilfe einer Kennzeichnung oder Zertifizierung gearbeitet werden. Dies würde den Nutzer:innen klare Hinweise darauf geben, dass die KI-Anwendung auf algorithmische Voreingenommenheit überprüft wurde und den definierten Standards entspricht, wodurch Transparenz und Vertrauen gefördert werden können. Hierbei muss wiederum darauf geachtet werden, dass durch zusätzliche Komplexität in der Kennzeichnung, die Gefahr der Überforderung oder Ablehnung der Kennzeichnung steigt.

4.3.3 Bestehende Kennzeichnungen

Die Umsetzung einer Kennzeichnung kann auf verschiedenste Art und Weise geschehen. Im Falle der Energieverbrauchskennzeichnung wurde beispielsweise eine Kennzeichnungspflicht per EU-Verordnung eingeführt. Die Nutri-Score-Kennzeichnung zur Nährwertkennzeichnung von Lebensmitteln ist eine Kennzeichnung, die freiwillig genutzt werden kann.

Die Kennzeichnung von KI-Inhalten erfordert eine differenzierte Herangehensweise, die den Zweck der Kennzeichnung und die unterschiedlichen Adressatenkreise berücksichtigt. Es ist nicht zu erwarten, dass sich eine einzige Kennzeichnung über alle Anwendungsbereiche hinweg durchsetzen wird. Das Spektrum der zu kennzeichnenden KI reicht dabei von einfachen Filteranwendungen bis zu komplexen synthetischen Medien, die durch KI generiert wurden. Verursacher:innen von bösartigen oder manipulativ verbreiteten Deepfakes werden ihre Produkte natürlich nicht kennzeichnen und beginnen bereits damit, authentische Inhalte als Deepfakes zu deklarieren.[172] Dies unterstreicht die Notwendigkeit einer zuverlässigen Kennzeichnung, die nicht leicht löschbar oder manipulierbar ist. In diesem Kontext könnten Authentizitätszertifikate für kritische Medieninhalte oder Chatanwendungen eine Lösung sein, ähnlich den bereits für Identitätsprüfungen von Social-Media-Accounts vorhandenen Zertifikaten. Ergänzend können Detektionssysteme für generative KI-Inhalte zur

171 Vgl. dazu zum Beispiel Löser et al. 2023
172 Vgl. Pawelec 2022

maschinellen Identifikation solcher Inhalte eingesetzt werden, wobei jedoch die derzeit noch unzureichende Zuverlässigkeit dieser Systeme berücksichtigt werden muss. Dieser Ansatz zielt darauf ab, nicht auf eine freiwillige Kennzeichnung durch die Verantwortlichen zu setzen, sondern eine automatisierte Identifikation von KI-generierten Inhalten zu gewährleisten.

In der Folge werden einige bereits existierende Beispiele aus verschiedenen Bereichen vorgestellt.

Abbildung 7: AI Ethics Label.
Quelle: AI Ethics Impact Group 2020, S.13

AI Ethics Label

Das AI Ethics Label wurde von der AI Ethics Impact Group[173] konzipiert, um Entwickler:innen von KI-Systemen eine Leitlinie zu bieten und gleichzeitig Bürger:innen, Anwender:innen sowie Verbraucher:innen eine leicht zugängliche Darstellung der ethischen Charakteristika eines KI-Systems bereitzustellen. Innerhalb dieses Labels werden sechs grundlegende Werte, die für die ethische Beurteilung von KI-Systemen von zentraler Bedeutung sind, evaluiert: Transparenz, Nachvollziehbarkeit, Datenschutz, Fairness, Verlässlichkeit und ökologische Nachhaltigkeit.[174]

Die Bewertung dieser Werte erfolgt durch eine Analyse verschiedener Kriterien, die die Erfüllung oder Verletzung jedes einzelnen Wertes definieren. Dies geschieht anhand konkreter Messungen oder Informationen über das KI-System. Die Ergebnisse können in ein kompaktes Label überführt werden, das dem Energieeffizienzlabel nachempfunden ist und prägnante Informationen über die ethisch relevanten Charakteristika des KI-Systems liefert (siehe Abbildung 7).

In Verbindung mit beispielsweise auf Risikoklassen basierenden Anforderungen, wie sie im AI Act formuliert sind, können Entscheidungsträger:innen schnell entscheiden, ob ein KI-System für ihre Zwecke geeignet ist oder nicht. Die Bewertung der einzelnen Kriterien kann grundsätzlich von den Hersteller:innen selbst durchgeführt werden. Um jedoch ein hohes Maß an Vertrauen in die Bewertung zu gewährleisten, wäre eine unabhängige Überprüfung durch eine Prüfinstitution empfehlenswert. Die Etablierung einer solchen Institution befindet sich derzeit in der Umsetzung.

Im Medienbereich kann eine solche Kennzeichnung vor allem für Medienschaffende hilfreich sein, die entscheiden müssen, ob sie ein bestimmtes KI-basiertes Werkzeug verwenden möchten und ob es mit regulatorischen Rahmenbedingungen oder einer möglichen Selbstverpflichtung kompatibel ist. Abhängig vom Anwendungskontext können insbesondere die Werte Transparenz in Bezug auf die Trainingsdaten des Systems oder Nachvollziehbarkeit von zentraler Bedeutung sein. Bei Recherchearbeiten, die sensible Daten involvieren, würde hingegen der Wert Privacy selbstverständlich eine deutlich höhere Gewichtung erfahren.

173 https://www.ai-ethics-impact.org/de
174 Vgl. AI Ethics Impact Group 2020

Valid

Ein alternativer Ansatz zur Kennzeichnung, insbesondere im Kontext von Nachrichtenartikeln, wird von Valid angeboten (siehe Abbildung 8).[175] Anstatt Produkte zu kennzeichnen, die von KI erstellt wurden, fokussiert die Kennzeichnung von Valid auf die Validierung und Authentifizierung der Herkunft eines Artikels. Dies gewährleistet, dass der Inhalt tatsächlich von einer bestimmten Person oder einem bestimmten Medium (wie beispielsweise einer Zeitung) stammt. Technisch wird dies durch eine kryptografische Signatur realisiert. Die Inhalte werden durch einen unveränderbaren und transparenten Eintrag auf einer Blockchain gesichert, wodurch Nutzer:innen die Möglichkeit haben, diese einfach zu überprüfen.

Abbildung 8: Valid Label. Quelle: http://valid.tech/

Diese Form der Kennzeichnung als Authentifikation ermöglicht es den Leser:innen, sicherzustellen, dass die Inhalte von vertrauenswürdigen Quellen stammen und nicht manipuliert wurden. Autor:innen profitieren von dieser Kennzeichnung, da das Vertrauen in ihre Veröffentlichungen gestärkt wird. Plattformbetreiber können durch die Kennzeichnung und insbesondere durch die maschinenlesbare Kodierung automatisch zwischen validierten Versionen eines Artikels und manipulierten Fassungen unterscheiden. Dadurch können sie effektiv gegen Desinformation vorgehen und gleichzeitig den neuen Regelungen und Gesetzen entsprechen.

Coalition for Content Provenance and Authenticity (C2PA)

Ein vergleichbarer Ansatz zur Authentifizierung für das Medium Fotos wird von der Coalition for Content Provenance and Authenticity (C2PA) vorangetrieben. Kameras, die deren Technologie beinhalten, erzeugen direkt beim Fotografieren Metadaten mit Informationen über die Kamera, den oder die Fotograf:in sowie den Zeitpunkt der Aufnahme. Diese Daten werden mithilfe einer digitalen Signatur mit dem Bild verknüpft, sodass eine Verifizierung der Echtheit des Fotos mit einfachen Werkzeugen möglich ist.[176] Auch wenn verschiedene Möglichkeiten der Umgehung der Authentifizierung denkbar sind, wie beispielsweise die Entfernung oder Fälschung von Metadaten oder das Fotografieren eines hochwertigen Ausdrucks eines Fakes, ist diese Kennzeichnung dennoch ein wichtiges Werkzeug, um gegen Desinformation vorzugehen und das Vertrauen in digitale Bildinhalte zu stärken, da der Aufwand für gefälschte Inhalte deutlich steigt.

175 http://valid.tech/
176 Zum Beispiel auf der Seite der C2PA: https://contentcredentials.org/verify

4.3.4 Zusammenfassung und Empfehlung

Zusammenfassend lässt sich festhalten, dass die Einführung einer Kennzeichnung und/oder Zertifizierung für KI-Inhalte im Medienbereich sinnvoll und wünschenswert ist. Bereits vorhandene Studien identifizieren relevante Prüfkriterien, die als Grundlage für eine standardisierte Bewertung dienen können.[177] Für die konkrete Umsetzung kann man sich zusätzlich an verschiedenen bereits vorhandenen Kennzeichnungen orientieren bzw. diese für einige Anwendungsfälle auch bereits nutzen.

Um eine erfolgreiche Implementierung sicherzustellen, bedarf es gezielter Projekte zur Sensibilisierung und Kommunikation über die Kennzeichnungen sowie ihrer Vorteile für die jeweiligen Zielgruppen. Die Bekanntmachung und Verständlichkeit der Kennzeichnungen sind entscheidend, um eine breite Akzeptanz zu erreichen und potenzielle negative Konsequenzen zu vermeiden. Hierbei könnten Medienanstalten sowohl in der Entwicklung von Kennzeichnungen als Kooperationspartner agieren als auch bei der Durchführung von Projekten zur Bekanntmachung eine wichtige Rolle einnehmen. In den Medienanstalten sind sowohl die erforderlichen Kompetenzen bezüglich der spezifischen Anforderungen an eine Kennzeichnung im Medienbereich als auch das Wissen über die relevanten Zielgruppen für Kommunikationsprojekte vorhanden. Ihre Mitwirkung könnte die erfolgreiche Einführung und Anwendung von Kennzeichnungen im Medienbereich maßgeblich unterstützen.

177 Vgl. zum Beispiel Heesen et al. 2020b

5 KI und Medienkompetenz

In diesem Kapitel wird zuerst aus den obigen Inhalten des Gutachtens die These entwickelt, dass die Veränderungen der Medienlandschaft durch Künstliche Intelligenz (KI) zu einem Bedarf an neuen Kompetenzen im Umgang mit Medien geführt haben (Kapitel 5.1). Hierbei ist insbesondere eine Stärkung der Technologiekompetenz in Bezug auf KI essenziell, da sie eine Voraussetzung für Medienkompetenz geworden ist. Gerade im Kontext einer Zunahme an Desinformation ist die Stärkung der Medienkompetenz auch für die Demokratiekompetenz notwendig. In Kapitel 5.2 werden bereits existierende Maßnahmen zur Medienkompetenzvermittlung dargestellt sowie neue Herangehensweisen diskutiert, die auf die veränderten Bedarfe reagieren und KI zur Vermittlung von Medienkompetenz nutzbar machen. Abschließend wird darauf eingegangen, welche Rollen die Medienanstalten bei der Umsetzung der Maßnahmen einnehmen können (Kapitel 5.3).

5.1 Notwendigkeit neuer Medienkompetenzen für Demokratiekompetenz

Aufbauend auf den bisher beschriebenen Inhalten wird argumentiert, dass es einen Bedarf für eine Weiterentwicklung der Medienkompetenz und neue Potenziale für ihre Vermittlung gibt. Vor allem in der näheren Zukunft wird die Technologiekompetenz in Bezug auf KI eine wichtigere Rolle spielen, da sie als Grundlage für komplexere Medienkompetenzen fungiert (Kapitel 5.1.1). Darüber hinaus werden Forschungsbedarfe ermittelt, die für eine effektive und effiziente Vermittlung von Technologie- und Medienkompetenz notwendig sind (Kapitel 5.1.2). Schließlich wird dafür argumentiert, dass gerade durch die Veränderungen des Medienbereichs durch KI, Medienkompetenz eine noch wichtigere Voraussetzung für Demokratiekompetenz ist (Kapitel 5.1.3).

5.1.1 Technologiekompetenz und Medienkompetenz

Wie aus den vorangehenden Abschnitten hervorgeht, erlebt der Medienbereich eine zunehmende Durchdringung von KI-basierten Anwendungen. Diese reichen von der Personalisierung von Inhalten über die automatisierte Erstellung von Nachrichten bis hin zur Analyse von Nutzerverhalten. Um Medieninhalte effektiv zu verstehen, zu bewerten und kritisch zu hinterfragen, ist es für die Nutzer:innen von entscheidender Bedeutung, ein umfassendes Verständnis über die Funktionsweise und die Auswirkungen dieser KI-Anwendungen zu entwickeln. Technologiekompetenz, die bislang eine eher untergeordnete Rolle in der Medienkompetenz spielte, gewinnt im Angesicht dieser dynamischen Entwicklungen an Bedeutung. Durch die starke Interaktion des Medienbereichs mit KI wird die Ausbildung einer Technologiekompetenz zur unerlässlichen Voraussetzung für eine vollumfängliche Medienkompetenz. Erst auf diesem Fundament kann eine informierte Einordnung und kritische Bewertung von Medien erfolgen.

Technologiekompetenz umfasst dabei nicht nur das Verständnis technologischer Mechanismen, sondern beinhaltet auch die aktive Handhabung von Technologie. Im Zusammenhang mit KI bedeutet dies, Medieninhalte nicht nur passiv zu konsumieren, sondern die Algorithmen und Mechanismen hinter diesen Inhalten zu durchschauen und die kreativen Potenziale der Technologie zu entdecken. Eine wirksame Medienkompetenz erfordert daher nicht nur die Fähigkeit zur Analyse von Medieninhalten, sondern auch ein tiefgehendes Bewusstsein für die technologischen Grundlagen, die diese Inhalte formen. Darüber hinaus erfordern auch die in diesem Gutachten beschriebenen Kennzeichnungen und Authentifizierungen von Inhalten Technologiekompetenz. In Zukunft ist es daher noch wichtiger, bei der Stärkung von Medienkompetenz immer auch die Technologiekompetenz mitzudenken und diese als Grundlage für die Medienkompetenz zu verstehen.

5.1.2 Grundlagen- und Evaluationsforschung

Für die Entwicklung und Durchführung von Maßnahmen zur Stärkung von Technologie- und Medienkompetenz ist Wissen über den aktuellen Bildungsstand der jeweiligen Zielgruppen essenziell. Nur so können Konzepte, die auch die Bedarfe treffen, entwickelt werden und auf Stärken und Schwächen verschiedener demographischer Gruppen gezielt eingegangen werden. Aktuell ist die Datenlage in Bezug auf Digital Literacy, also die Fähigkeiten, die für eine erfolgreiche Teilhabe in einer digitalen Gesellschaft notwendig sind, vor allem bei Kindern und Jugendlichen sehr schlecht. Unterschiede in Bezug auf die Nutzung von und das Wissen über digitale Medien, die es beispielsweise bei verschiedenen Altersgruppen, Schulbildungen oder zwischen den Geschlechtern gibt, können aktuell nur vermutet werden.

Es bedarf daher mehr Forschung, wie beispielsweise die JIM-Studie des Medienpädagogischen Forschungsverbunds Südwest (MPFS), die den Medienumgang der 12- bis 19-jährigen in Deutschland untersucht.[178] Neben einer aktuellen Standortbestimmung dienen die Daten der Entwicklung von Strategien und Ansätzen für neue Konzepte in den Bereichen Bildung, Kultur und Arbeit. Dabei sollte vor allem auf die aktuellen Entwicklungen Bezug genommen werden und Daten über das Verständnis von KI und Algorithmen sowie der Umgang mit sozialen Netzwerken untersucht werden. Das Projekt „Algorithmen und Künstliche Intelligenz im Alltag von Jugendlichen“ der Ludwigs-Maximilian-Universität München, das von den Bayerischen Landesmedienanstalten gefördert wird, ist ein positives Beispiel in diese Richtung. Die Medienanstalten können hierbei sowohl als Förderer auftreten, um die Lücke in der Forschung zu schließen, und auch als Multiplikator, der die Verbreitung von Studienergebnissen an die relevanten Akteure unterstützt.

Weitere Beispiele für Forschungsprojekte im Kontext von Medienkompetenz sind die Projekte *DataSkop* und *Coding Public Value*. DataSkop ist ein gemeinsames Projekt von AlgorithmWatch, der Europa-Universität Viadrina, der Fachhochschule Potsdam, der Universität Paderborn und dem Verein Mediale Pfade, das die Idee von Datenspenden zur Erforschung intransparenter Algorithmen nutzt. Das dreijährige Vorhaben begann mit dem Pilotprojekt *Wahlempfehlung: Was zeigt dir der YouTube-Algorithmus zur Bundestagswahl?* im Herbst 2021. Hier wurden Datenspenden analysiert, um personalisierte Empfehlungen und Suchergebnisse des YouTube-Systems im Kontext der Bundestagswahl zu untersuchen. Das zweite Projekt von DataSkop, das Anfang 2023 begonnen hat, widmet sich dem Empfehlungsalgorithmus von TikTok. Neben der Forschung bietet DataSkop seit Herbst 2021 auch schulische und außerschulische Lernszenarien an. Das Hauptziel besteht darin, durch Datenspenden die undurchsichtigen Algorithmen sozialer Medien und automatischer Entscheidungssysteme zu erhellen, wobei ähnliche Initiativen wie der *Citizen Browser* von The Markup und *Mozilla Rally* als Beispiele dienen.[179]

In dem Projekt *Coding Public Value* wird die Frage thematisiert, wie Software so entwickelt werden kann, dass sie sowohl das Gemeinwohl als auch die Interessen der Nutzer:innen und die medienrechtliche Regulierung berücksichtig? Dabei stehen zwei zentrale Fragen im Mittelpunkt: Wie lassen sich rechtliche, politische und nutzerorientierte Anforderungen an öffentlich-rechtliche Medien in Software umsetzen? Welche institutionellen, politischen und organisatorischen Voraussetzungen müssen erfüllt werden, um Medienplattformen auf Grundlage eines gemeinwohlorientierten Software Engineering zu betreiben? In dem Projekt wurde ein Leitfaden für die Softwareentwicklung

178 Vgl. MPFS 2023
179 Vgl. DataSkop 2024

in öffentlich-rechtlichen Medienanstalten erstellt. Dieser Leitfaden bietet die Grundlage für die Erfassung und Dokumentation von rechtlichen, politischen und institutionellen Anforderungen unter Berücksichtigung öffentlicher Werte.[180]

Neben der Grundlagenforschung ist für die Entwicklung und Weiterentwicklung von Maßnahmen zur Stärkung der Technologie- und Medienkompetenz auch begleitende Evaluationsforschung ein wichtiger Baustein, um die Effektivität von Maßnahmen wissenschaftlich fundiert zu bewerten und entsprechende Anpassungen zu treffen.

5.1.3 Medienkompetenz für Demokratiekompetenz

Medienkompetenz ist eine notwendige Voraussetzung für Demokratiekompetenz. Demokratiekompetenz beinhaltet die Fähigkeit zur freien Meinungsbildung. Die Ausbildung einer Informations- und Meinungsbildungskompetenz ist eine notwendige Voraussetzung, um dies erfolgreich zu bewältigen. Hierfür sind ein Verständnis über die Medienlandschaft und die zugrundeliegenden Dynamiken und Entwicklungen notwendig. Dies wird besonders vor dem Hintergrund deutlich, dass zunehmend mit Desinformation versucht wird, Einfluss auf das Wahlverhalten zu nehmen. In einer Zeit, in der die Verbreitung von manipulativen Inhalten über verschiedene Medienkanäle rasant zunimmt, ist es entscheidend, dass Bürger:innen in der Lage sind, solche Versuche zu durchschauen. Dies erfordert ein hohes Maß an Medienkompetenz. Nur durch ein solides Verständnis der Funktionsweise von Medien, der Verbreitung von Informationen und der Mechanismen hinter Desinformationskampagnen können Bürger:innen resilient gegenüber solchen Kampagnen sein und sich weiterhin frei ihre Meinung bilden und Entscheidungen treffen. Medienkompetenz ist somit ein Schlüssel zur Stärkung der Demokratiekompetenz, indem sie die notwendige Widerstandsfähigkeit gegenüber manipulativen Einflüssen fördert.

5.2 Maßnahmen zur Medienkompetenzvermittlung

In diesem Kapitel werden verschiedene Methoden zur Medienkompetenzvermittlung betrachtet, um daraus abzuleiten, welche Maßnahmen auch in Zukunft von Relevanz sein werden. Dazu werden zuerst klassische Ansätze der Medienkompetenzvermittlung in der Praxis betrachtet (Kapitel 5.2.1). Anschließend werden Maßnahmen diskutiert, die KI für die Vermittlung von Medienkompetenz nutzen (Kapitel 5.2.2).

5.2.1 Medienkompetenzvermittlung in der Praxis

Im folgenden Abschnitt werden Beispiele aus unterschiedlichen Praxisbereichen präsentiert, die bereits in der Vergangenheit erfolgreich zur Förderung der Medienkompetenz beigetragen haben. Diese Beispiele erstrecken sich über diverse Formen, von lokal umgesetzten Projekten bis hin zu Online-Seminaren. Dabei handelt es sich um etablierte Muster, die bereits seit einigen Jahren praktiziert werden und auf bewährten Methoden basieren.

180 Vgl. bidt 2024

Medienkompetenz vor Ort im regionalen und lokalen Raum
Eine effektive Strategie zur Förderung von Medienkompetenz manifestiert sich in praxisorientierten Angeboten, die einen interaktiven Austausch mit Fachexperten ermöglichen. Diese Praxis wird bereits in verschiedenen Kontexten umgesetzt, sei es in schulischen Projekten, Konferenzen oder durch Medienunternehmen, die Einblicke in ihre Produktionsprozesse im Rahmen von Tagen der offenen Tür gewähren.

Ein beispielhaftes Projekt, das diesen Ansatz verfolgt, ist *YourStory*, initiiert von der Landesanstalt für Kommunikation (LFK) in Baden-Württemberg in Zusammenarbeit mit der Landesvereinigung Kulturelle Jugendbildung (LKJ).[181] In diesem Projekt erstellen Jugendliche in schulischen Einrichtungen Erklärvideos zu Themen, die sie persönlich begeistern. Das Projekt wird von zwei Medienreferenten der LKJ an Schulen oder Jugend(kultur)einrichtungen durchgeführt. Die Jugendlichen wählen in Kleingruppen Themen für ihre eigenen Filme und übernehmen verschiedene Rollen, wie die vor oder hinter der Kamera, die Erstellung von Grafiken oder den Videoschnitt. Dies ermöglicht nicht nur die Entdeckung neuer Stärken, sondern vermittelt auch wichtige Kenntnisse über die Funktionsweise von Video-Plattformen, den Schutz persönlicher Rechte und die Verdienstquellen von Influencer:innen. Jugendliche sollten erkennen, dass Produktplatzierungen nicht ausschließlich zweckgebundene Absichten repräsentieren, sondern dass Influencer:innen auch Einnahmen erzielen und ihre Reichweite für monetäre Zwecke nutzen. Deswegen werden Teilnehmer:innen in dem Projekt sensibilisiert, Werbung zu erkennen. Insgesamt fördert das Projekt die Entwicklung von Medienproduktionskompetenzen und reflektiert die Medienrezeption der Jugendlichen.

Ein weiteres Beispiel ist das Projekt *Echt Fake, ich schwör!* der LFK in Zusammenarbeit mit der LKJ.[182] Hier setzen sich LKJ-Medienpädagog:innen gemeinsam mit Jugendlichen an Schulen oder außerschulischen Bildungseinrichtungen in Baden-Württemberg mit den Mechanismen von Verschwörungsmythen auseinander und üben Quellenkritik. Um das Verständnis von Desinformation zu vertiefen, erfinden die Teilnehmer:innen ihre eigenen Verschwörungsgeschichten, führen Interviews mit vermeintlichen Expert:innen, schreiben Zeitungsartikel über angebliche Beweise oder erstellen Videos zur vermeintlich logischen Aufdeckung von Fakten. Die Medienpädagog:innen unterstützen sie dabei, überzeugende mediale „Beweise“ zu erstellen. Ein zentraler Bestandteil ist die gemeinsame Auflösung, bei der die Mechanismen der Überzeugungsversuche durch die erstellten Geschichten reflektiert werden. Durch diese praktische Erfahrung können die Teilnehmer:innen die Mechanismen und potenziellen Gefahren von Desinformation besser verstehen und werden dafür sensibilisiert. Dadurch werden kritisches Bewusstsein und Medienkompetenz im Umgang mit Nachrichten gefördert.

Insgesamt bieten Medienkompetenzaktivitäten vor Ort im regionalen und lokalen Raum eine facettenreiche und praxisorientierte Herangehensweise, um die Medienkompetenz von Jugendlichen zu stärken. Durch die Kombination von theoretischem Wissen und praktischer Erfahrung können Teilnehmer:innen auf die Anforderungen der digitalen Medienwelt vorbereitet werden.

181 Vgl. LKJ 2024a
182 Vgl. LKJ 2024b

Publikationen und Handbücher
Publikationen und Handbücher wirken in der Förderung von Medienkompetenz mit, da sie nicht nur Wissen vermitteln, sondern auch klare Leitfäden und praxisnahe Anleitungen bieten können. Ein Beispiel hierfür ist das Klicksafe Handbuch *Ethik macht Klick*.[183] Klicksafe ist eine Initiative, die sich insbesondere für die Sicherheit von Kindern und Jugendlichen im Internet einsetzt.

Das Klicksafe-Handbuch *Ethik macht Klick* bietet Einblicke in das Informationsverhalten von Jugendlichen und regt sie dazu an, wie man mit Desinformation umgehen kann. Es unterstützt beim Analysieren und Erkennen von Desinformationsstrategien und verdeutlicht die Auswirkungen von Falschinformationen auf die demokratische Gesellschaft. Im Fokus steht eine medienethische Roadmap – ein Kompass für verschiedene Kompetenzen, der Methodenkompetenz („Wie informiere ich mich?“), Sachkompetenz („Welches Wissen habe ich über Medien und digitale Öffentlichkeiten?“), Sozialkompetenz („Wie verhalte ich mich in Diskussionen?“) und ethische Kompetenz („Wie entwickle ich eine Haltung?“) umfasst.

Verschiedene Organisationen, wie etwa Deutschland sicher im Netz (DsiN), veröffentlichen regelmäßig Materialien, Informationsbroschüren und Leitfäden zu verschiedenen Aspekten der digitalen Sicherheit und Medienkompetenz. Auch auf der BLM-Webseite[184] sind kostenfreie Publikationen verfügbar, wie zum Beispiel die Publikation *Dein Algorithmus – Meine Meinung!*[185], die sich eingehend mit dem Thema Algorithmen und ihrer Bedeutung für die Meinungsbildung und Demokratie auseinandersetzt.

Online-Beratung und -Seminare
Online-Beratungen und -Seminare können Medienkompetenz fördern, indem sie einen barrierefreien Zugang zu Bildungsinhalten ermöglichen, unabhängig vom geografischen Standort. Diese Formate schaffen nicht nur flexiblere Teilnahmemöglichkeiten, sondern fördern auch eine aktive Beteiligung durch interaktive Elemente und Diskussionsforen. Die breit gefächerten Themen und die Option des selbstbestimmten Lernens tragen dazu bei, eine umfassende Medienkompetenz zu entwickeln.

In den letzten Jahren sind Faktenchecks immer mehr in den Fokus der Öffentlichkeit gerückt. Dieses Thema spielt insbesondere in der Medien- und Informationslandschaft eine wichtige Rolle, da es darum geht, falsche oder irreführende Aussagen zu identifizieren und zu korrigieren.[186] Viele verschiedene Organisationen, wie beispielweise Nachrichtensender, Suchmaschinenanbieter und die Deutsche Presse-Agentur (dpa) haben erkannt, welchen Einfluss sie als Multiplikatoren haben, und integrieren Faktenchecks als festen Bestandteil in ihren täglichen Betrieb. In einer kooperativen Initiative stellen die dpa und die Google News Initiative ihre Fachkenntnis im Bereich Faktenchecks unter dem Namen *Faktencheck23*[187] zur Verfügung. Die angebotenen Workshops vermitteln Journalist:innen grundlegende Fähigkeiten zum Durchführen von Faktenchecks und somit der Qualitätssicherung von medialen Inhalten.

JUUUport[188] bietet eine Plattform, auf der Jugendliche sich über wichtige Themen wie Datenschutz, Cyber-Mobbing, Online-Sicherheit und digitale Rechte informieren können. Jugendliche können Fragen stellen und erhalten Unterstützung von JUUUport-Scouts – junge Menschen, die speziell

183 Vgl. Medienanstalt Rheinland-Pfalz 2024
184 www.blm.de
185 Vgl. BLM 2017
186 Vgl. Rashkin et al. 2017
187 Vgl. dpa 2023
188 https://www.juuuport.de/

geschult wurden, um Altersgenossen in Bezug auf Online-Sicherheit zu beraten. Die Plattform bietet zudem ein Kontaktformular an, über das Jugendliche anonym Fragen zu verschiedenen Themen rund um die sichere Internetnutzung stellen können. Das JUUUport-Team setzt sich dann mit passenden Informationen oder Ratschlägen mit den Jugendlichen in Verbindung. JUUUport bietet auch Onlineseminare mit interaktiven Elementen wie Umfragen, Quiz-Spiele und einem Whiteboard-Tool an. So entstehen vielfältige Möglichkeiten für gemeinsames Brainstorming.

Weitere Online-Beratungsplattformen wie FLIMMO[189] geben Eltern Tipps dazu, wie sie ihre Kinder altersgerecht über seriöse Quellen informieren können, inklusive Strategien zur Erkennung von manipulierten Inhalten. ZEBRA[190], die Beratungsplattform der Landesanstalt für Medien NRW, bietet persönliche Beratung durch Medienexpert:innen an und erklärt, was jede/jeder Einzelne gegen Desinformation tun kann.[191]

Medienkompetenzkampagnen
Medienkompetenzkampagnen tragen zur Stärkung von Medienkompetenz bei, indem sie gezielte Aufklärung bieten, um ein bewusstes und kritisches Verständnis für Medieninhalte zu fördern. Durch die Sensibilisierung für Themen wie Faktenprüfung und sicheres Online-Verhalten ermöglichen diese Kampagnen eine sicherere Teilnahme am digitalen Raum. Sie tragen somit dazu bei, die Bürger:innen besser vor Desinformation zu schützen und ihre Fähigkeiten im Umgang mit Medien zu verbessern.

In Zusammenarbeit mit deutschen NGOs, Faktenchecker:innen, Wissenschaftler:innen und Desinformations-Expert:innen setzt die Google-Tochter Jigsaw auf eine innovative Prebunking-Kampagne. Das Konzept des Prebunking basiert auf wissenschaftlichen Erkenntnissen und hilft Nutzer:innen dabei, sich gegen zukünftige Versuche, sie mit falschen Informationen zu manipulieren, zu wappnen, indem kurze Videos, die gängige Manipulationstechniken der Desinformation thematisieren, gezeigt werden. In Polen, Tschechien und der Slowakei erreichten die Prebunking-Videos über 38 Millionen Aufrufe, was fast einem Drittel der Bevölkerung entspricht. Nach dem Ansehen der Videos stieg die Erkennungsrate von Desinformationstaktiken um bis zu acht Prozentpunkte. In enger Zusammenarbeit mit lokalen Expert:innen wird nun eine Videokampagne für Deutschland gestartet, um Menschen für Falschinformationen zu sensibilisieren.[192]

Auch die aktuelle Kampagne der Vodafone Stiftung Deutschland mit dem Hashtag *#WeissteBescheid*[193], setzt sich zum Ziel, Medienkompetenz zu stärken. Auf TikTok ermöglicht die neue Klickwinkel-Initiative eine kreative Begegnung zwischen den Generationen. Hier vermitteln erfahrene TikTok-Nutzer:innen und Senior:innen nicht nur Fähigkeiten zur Internetnutzung, sondern auch umfassende Kompetenzen für einen souveränen Umgang mit Social Media. Die Verbindung von Digital Natives mit älteren Generationen schafft eine entspannte Atmosphäre, in der Jung und Alt sich austauschen. Von Begriffen wie Cookies bis hin zu Deepfakes, von KI bis Online-Dating werden die verschiedenen Facetten des Internets gemeinsam erkundet. In einem respektvollen Dialog werden Fragen zur Internetnutzung gestellt, wodurch eine bereichernde Dynamik des gegenseitigen Lernens entsteht. Die Kampagne präsentiert dabei nicht nur Tipps und Tricks für Internet-Nutzer:innen jeden Alters, sondern setzt auch ein positives Zeichen für einen generationenübergreifenden Austausch im digitalen Raum.

189 https://www.flimmo.de/
190 https://www.fragzebra.de/
191 Vgl. Landesanstalt für Medien NRW 2023
192 Vgl. Goldberg 2023
193 Vgl. Vodafone Stiftung 2023

Im Auftrag des Ministeriums für Kultus, Jugend und Sport Baden-Württemberg führt das Landesmedienzentrum Baden-Württemberg die landesweite Kampagne *BITTE WAS?! Kontern gegen Fake und Hass*[194] durch. Die Initiative setzt sich aktiv gegen Hass, Fake News und Hetze in sozialen Netzwerken ein. Ziel ist es, Kinder und Jugendliche dazu zu ermutigen, sich für ein positives gesellschaftliches Miteinander, auch in der digitalen Welt, einzusetzen. Eine weitere Medienkompetenzkampagne der Landesmedienanstalten Saarland (LMS), die *Goldenen Medienregeln*[195], basiert auf dem Ansatz, Regeln zur Mediennutzung so ansprechend zu gestalten, dass sie gerne im häuslichen Umfeld aufgehängt werden. Durch die Gestaltung von attraktiven Informationsbroschüren, die leicht verständlich sind und im direkten Umfeld der Mediennutzung platziert werden, unterstützt die LMS Eltern und Kinder bei einem bewussten Umgang mit Medien.

Prüf- und Aufsichtspraxis
Die Ausübung von Prüf- und Aufsichtspraxen durch Medienanstalten ist von zentraler Bedeutung für die Sicherstellung verschiedener Aspekte der Medienlandschaft. Dies umfasst den Jugendschutz, indem sichergestellt wird, dass Medieninhalte altersgerecht sind und Kinder und Jugendliche vor schädlichen Inhalten geschützt werden. Medienanstalten können Qualitätsstandards für journalistische Ethik, Genauigkeit und Fairness setzen und überwachen deren Einhaltung. Rechtliche Bestimmungen, einschließlich Datenschutz und Wettbewerbsrecht, werden von Medienanstalten überprüft, um sicherzustellen, dass Medienunternehmen im Einklang mit dem Gesetz operieren. Die Förderung von Pluralität und Meinungsvielfalt sowie der Schutz vor Desinformation sind weitere wichtige Ziele dieser Praxen. Insgesamt tragen diese Überwachungsmechanismen dazu bei, die Integrität und Vertrauenswürdigkeit der Medien zu wahren.

Die Bayerischen Landesmedienanstalt (BLM) setzt sich im Rahmen ihrer Tätigkeiten im Jugendschutz und in der Medienpädagogik in verschiedener Hinsicht aktiv gegen Extremismus, Antisemitismus und damit verbundene Problemfelder ein. In einem ersten Schritt hat die BLM sich für ein Indizierungsverfahren bei der Bundesprüfstelle für jugendgefährdende Medien (BPjM) eingesetzt. Als Resultat wäre die betroffene Webseite künftig beispielsweise nicht mehr über gängige Suchmaschinen auffindbar. Weitere Handlungsmöglichkeiten werden derzeit noch geprüft.[196]

Die Medienanstalt Hamburg/Schleswig-Holstein (MA HSH) reagierte auf ein YouTube-Video des Österreichers Martin Sellner, Mitglied der Identitären Bewegung, das Bilder eines Tränengaseinsatzes gegen Geflüchtete an der türkisch-griechischen Grenze zeigte. Das Video präsentierte Menschen, die erheblichen physischen und psychischen Leiden ausgesetzt waren, ohne dass ein berechtigtes Interesse an dieser Darstellung vorlag. Die Darstellung des Leidens wurde vielmehr zu Zwecken der Misinformation genutzt und Sellner dekontextualisierte die Inhalte. Die MA HSH meldete das Video wegen Desinformation sowie möglichen Verstoßes gegen die Menschenwürde (§ 4 Abs. 1 Satz 1 Nr. 8 JMStV) bei YouTube. Parallel dazu wurden über siebzig strafrechtlich relevante Hasskommentare von der MA HSH an die Plattform gemeldet. YouTube löschte daraufhin die Kommentare und sperrte das Video für Nutzer:innen in Deutschland. Die Reaktion der MA HSH auf das YouTube-Video von Sellner zeigt die Bedeutung effektiver Aufsichtsmaßnahmen im digitalen Raum.[197]

194 https://bitte-was.de/
195 Vgl. Landesmedienanstalt Saarland 2018
196 Vgl. BLM 2024
197 Vgl. MA HSH 2021

Unterstützung von Bildungseinrichtungen
Medienpädagog:innen und Journalist:innen engagieren sich in Bildungseinrichtungen, um praxisnahe Einblicke in Medien und Journalismus zu vermitteln. Dies trägt dazu bei, dass Schüler:innen die Funktionsweise von Medien besser verstehen und ein bewusstes Medienverhalten entwickeln. Projekte zu Themen wie Cyber-Mobbing, Desinformation und Sozialen Netzwerke fördern kritische Medienreflexion und ermöglichen einen sensiblen Umgang mit digitalen Inhalten.

Das medienpädagogische Referentennetzwerk Bayern spielt eine bedeutende Rolle in der Unterstützung bayerischer Bildungseinrichtungen, darunter Schulen, Kindergärten und Familienzentren. Das Netzwerk bietet kostenfrei Referent:innen für medienpädagogische Informationsveranstaltungen in verschiedenen Formaten an, sowohl in Präsenz als auch online. Dabei werden relevante Themen wie Cyber-Mobbing, digitale Spiele, Internet und Smartphone sowie Soziale Netzwerke und Messenger behandelt. Das Projekt wird von der Bayerischen Staatskanzlei gefördert, um eine umfassende medienpädagogische Unterstützung zu gewährleisten.

Ein weiteres Projekt ist *Journalismus macht Schule* der Medienanstalt Berlin und Brandenburg (MABB).[198] Hier engagieren sich Journalist:innen ehrenamtlich, um Schulklassen ab der 9. Jahrgangsstufe über ihre Arbeit zu informieren. Diese Initiative, aufgelegt von der Süddeutschen Zeitung und unterstützt durch zahlreiche Medienanbieter, dient als niedrigschwellige Zugangsmöglichkeit zum Thema Medienkompetenz. Ziel ist es, an Schulen das Bewusstsein für Nachrichten- und Informationskompetenz zu schärfen und jungen Menschen einen kritischen Umgang mit Desinformation zu vermitteln.

5.2.2 Rolle von KI bei der Medienkompetenzvermittlung

KI kann den Zugang zu Informationen erleichtern, personalisierte Lernangebote schaffen und innovative Lehrmethoden unterstützen. KI-basierte Analysewerkzeuge könnten dazu beitragen, Fehlinformationen und Desinformationen schneller zu identifizieren. Zusätzlich ermöglichen algorithmische Empfehlungssysteme die Bereitstellung personalisierter Inhalte, die den individuellen Lernbedürfnissen entsprechen. Besonders im Kontext der Verbreitung von falschen oder manipulierten Informationen im Internet ist es wichtig, dass Individuen die Glaubwürdigkeit von Quellen überprüfen können und sich kritisch mit den Inhalten auseinandersetzen können. KI kann einerseits dazu verwendet werden, Medienkompetenz zu vermitteln und somit als unterstützendes Instrument dienen, um Menschen dabei zu helfen, ihre Fähigkeiten im Umgang mit Medien zu entwickeln. Andererseits können unter Zuhilfenahme von KI herausfordernde Inhalte oder Situationen auf eine zugängliche Art und Weise präsentiert werden. Nachfolgende Beispiele erläutern die verschiedenen Rollen, die KI in diesem Kontext einnehmen kann.

Automatisierte Lernplattformen – KI-basiertes Learning Management System (LMS)
Im Gegensatz zum traditionellen Lernmanagementsystem (LMS), das sich auf die Organisation und Bereitstellung von Kursen konzentriert, hat das KI-basierte LMS die moderne Personal- und Organisationsentwicklung durch intelligente Funktionen und Algorithmen transformiert. Lernplattformen sind in der Lage, maßgeschneiderte Lernpfade zu erstellen, die individuell auf die Bedürfnisse und Fähigkeiten der Lernenden zugeschnitten sind. Dies kann dazu beitragen, Medienkompetenz auf verschiedene Aspekte, wie den kritischen Umgang mit Informationen im Internet oder die Erkennung von Fehlinformationen, auszudehnen.

198 Vgl. MABB 2024

So können KI-basierte Lernplattformen Medienkompetenzkurse anbieten und Lernenden interaktive Module zur Verfügung stellen. Diese Module konzentrieren sich auf Themen wie Quellenkritik, Datenschutz im Internet und sicheres Online-Verhalten. Ein Beispiel hierfür ist die *Khan Academy*[199], die maschinelles Lernen einsetzt, um maßgeschneiderte Lernressourcen bereitzustellen und den individuellen Lernfortschritt zu verfolgen.

Chatbots und Virtuelle Assistenten
KI-gestützte Chatbots oder virtuelle Assistenten bieten die Möglichkeit, Faktenchecks durchzuführen, Ratschläge zur Verbesserung der Medienkompetenz und zielgerichtete Informationen bereitzustellen. Diese interaktiven Anwendungen fördern selbstgesteuertes Lernen und bieten einen unkomplizierten Zugang zu relevanten Inhalten. Darüber hinaus können Chatbots auch bei der Entwicklung von Medienkompetenz unterstützen.

Ein exemplarisches Szenario in diesem Kontext umfasst die Integration oder die Referenz eines Chatbots auf Webseiten, der Nutzer:innen zur Verfügung steht, um sachkundige Ratschläge zu erteilen, Anfragen zu beantworten und auf relevante Ressourcen zu verweisen. Ein konkretes Beispiel für eine solche Implementierung ist auf der Plattform Correctiv[200] zu finden. Hierbei ermöglicht ein Whatsapp-Chatbot nicht nur das Melden von Falschmeldungen, sondern versorgt die Nutzer:innen zudem mit instruktiven Hinweisen zur Identifizierung von Desinformation und informiert über neuerlich aufgedeckte Täuschungsversuche.

Analysewerkzeuge für Medieninhalte
KI kann dazu verwendet werden, Medieninhalte automatisch zu analysieren und zu bewerten. Dies könnte beispielsweise die Identifizierung von Fehlinformationen, Hassrede oder unangemessenen Inhalten erleichtern. Solche Werkzeuge können Mediennutzenden helfen, fundiertere Entscheidungen über die Inhalte zu treffen, denen sie im Internet begegnen.

Verschiedene Plattformen tragen zur Förderung von Medienkompetenz bei, indem sie automatisiert Nachrichteninhalte analysieren und die Vertrauenswürdigkeit dieser Informationen anzeigen. Darüber hinaus haben Medienkonsument:innen die Möglichkeit, eigenständig Faktenchecks durchzuführen und erhalten Hintergrundinformationen zu den geprüften Inhalten. Beispiele hierfür sind das *KIVI-KI-Tool* der Medienanstalten NRW[201], welches durch die automatisierte Überwachung von Social-Media-Plattformen und Webseiten mögliche Gesetzesverstöße erkennt, und der Google *Fact Check Explorer*[202], der als Suchmaschine für Faktenchecks hilft, zwischen Fakten und manipulierten Inhalten zu unterscheiden. Zudem ermöglicht die Plattform *Deep Fake Total*[203] des Fraunhofer AISEC, KI-generierte Audio-Deepfakes zu erkennen. Diese Plattformen unterstützen die Ausbildung einer umfassenden Medienkompetenz, indem sie Nutzer:innen Werkzeuge zur Verfügung stellen, um Nachrichten kritisch zu hinterfragen und auf ihre Faktizität zu prüfen.

Personalisierte Nachrichtenaggregatoren
KI kann für personalisierte Nachrichtenaggregatoren eingesetzt werden, die auf individuellen Interessen und Präferenzen basieren. Dies fördert eine ausgewogene Informationsaufnahme und unterstützt die Nutzer:innen dabei, verschiedene Perspektiven und Quellen einzubeziehen. Der Begriff

199 https://de.khanacademy.org/
200 Vgl. Correctiv 2024
201 Vgl. Landesanstalt für Medien NRW 2021
202 https://toolbox.google.com/factcheck/explorer
203 https://deepfake-total.com/

ausgewogene Informationsaufnahme im Kontext von personalisierten Nachrichtenaggregatoren meint, dass die personalisierte Zusammenstellung von Inhalten darauf abzielt, der Nutzer:in eine Vielfalt von Informationen zu präsentieren. Das Ziel ist es, sicherzustellen, dass die präsentierten Inhalte nicht einseitig oder ausschließlich auf die bestehenden Vorlieben und Ansichten des Nutzers ausgerichtet sind. Hinsichtlich der Gefahr, dass Nutzer:innen durch diese personalisierten Aggregatoren nur noch Inhalte erhalten, die ihren (politischen) Ansichten entsprechen, besteht die Herausforderung darin, diese Aggregation so zu gestalten, dass sie nicht in einer Filterblase resultiert. Filterblasen können entstehen, wenn Nutzer:innen ausschließlich mit Inhalten konfrontiert werden, die ihre bestehenden Ansichten bestätigen. Um dem entgegenzuwirken, ist es entscheidend, dass personalisierte Nachrichtenfeeds auch Diversität und Ausgewogenheit fördern. Die KI kann so konfiguriert werden, dass sie der Nutzer:in bewusst eine Vielfalt von Meinungen und Standpunkten präsentiert, um eine umfassendere Perspektive zu gewährleisten. Ein weiterer wichtiger Aspekt ist die Transparenz solcher KI-gesteuerten Systeme. Transparenz ermöglicht es den Nutzer:innen, besser zu verstehen, wie ihre personalisierten Nachrichtenfeeds zusammengestellt werden, und fördert ein Verständnis für die Vielfalt der präsentierten Informationen. Idealerweise sollte die Transparenz schon bei der Entwicklung als grundlegende Anforderung gelten (Transparency by Design).

Ein Beispiel für einen KI-gesteuerten Nachrichtenaggregator ist *Google News*. Diese Plattform analysiert Benutzerpräferenzen und erstellt personalisierte Nachrichtenfeeds, wodurch Nutzer:innen eine breite Palette von Perspektiven erhalten und ihre Informationsquellen diversifizieren können. Durch den Einsatz von maschinellem Lernen kann Google News individuelle Interessen und Lesegewohnheiten der Nutzer:innen verstehen und darauf basierend maßgeschneiderte Nachrichtenfeeds generieren. Diese personalisierte Herangehensweise ermöglicht es den Nutzer:innen nicht nur, relevante und aktuelle Informationen zu erhalten, sondern eröffnet auch die Möglichkeit, verschiedene Standpunkte und Quellen zu erkunden. Dies trägt dazu bei, eine umfassende und vielfältige Informationsgrundlage zu schaffen, die den individuellen Präferenzen und Interessen der Nutzer:innen gerecht wird.

Spiele und Simulationen
KI kann in Serious Games und Simulationen integriert werden, um realitätsnahe Situationen zu schaffen, in denen die Nutzer:innen ihre Medienkompetenz testen und verbessern können. Ein Serious Game verfolgt ernsthafte Ziele jenseits der reinen Unterhaltung, wie etwa Bildung, Training, Gesundheitsförderung oder Bewusstseinsbildung. Die Integration von spielerischen Elementen erleichtert das Lernen und motiviert die Nutzer:innen.

Ein konkretes Beispiel hierfür ist *Quandary*, das vom Learning Games Network entwickelt wurde. Dieses Spiel regt Jugendliche dazu an, ethische Entscheidungen im digitalen Raum zu treffen. Durch das Eintauchen in eine virtuelle Welt und das Bewältigen herausfordernder Situationen sollen die Spieler:innen ihre Fähigkeiten im Umgang mit digitalen Medien verbessern. Diese Art von spielerischen Ansätzen bietet eine effektive Möglichkeit, Medienkompetenz zu fördern, indem sie das Lernen durch Interaktion und Spaß unterstützen.

5.3 Verschiedene Rollen der Medienanstalten

In vorherigen Abschnitten wurden beispielhaft Projekte zur Förderung der Medienkompetenz vorgestellt, bei denen die Medienanstalten bereits häufig vertreten waren und vielfältige Funktionen übernahmen. Die Medienanstalten spielen eine entscheidende Rolle bei der Verbreitung von

Informationen, der Festlegung von Standards und der Vermittlung bewährter Methoden in der Medienbildung. In ihrer Rolle als Multiplikatoren sorgen sie für die Verbreitung relevanter Publikationen und Informationen. Als Förderer setzen sie gezielte Unterstützung und finanzielle Mittel für Projekte ein, die die Medienkompetenz in der Bevölkerung stärken sollen. Zudem agieren sie als Kooperationspartner, indem sie mit verschiedenen Institutionen zusammenarbeiten, um gemeinsame Ziele im Bereich der Medienregulierung und -förderung zu erreichen.

In ihrer Funktion als Multiplikatoren verbreiten sie Informationen, Richtlinien und bewährte Praktiken im Bereich der Medienregulierung, tragen zur Schaffung eines einheitlichen Verständnisses für regulatorische Anforderungen bei und sorgen für Transparenz in Bezug auf rechtliche Rahmenbedingungen. Vor Ort fördern sie durch praxisorientierte Angebote, Publikationen, Handbücher und Online-Beratungen eine Vielzahl von Medienkompetenzaktivitäten.

In der Rolle als Förderer stellen die Medienanstalten gezielte Unterstützung und finanzielle Mittel für verschiedene Medienbildungsprojekte bereit. Diese finanziellen Zuwendungen tragen dazu bei, eine vielfältige Medienlandschaft zu fördern und innovative Ansätze zur Stärkung der Medienkompetenz zu unterstützen. Dies schließt die Förderung von Bildungsprogrammen, Forschungsprojekten, Medieninitiativen sowie Kampagnen und die Übernahme einer Aufsichtsrolle im Netz ein.

Als Kooperationspartner arbeiten die Medienanstalten mit verschiedenen Institutionen, Organisationen und Akteuren zusammen, um gemeinsame Ziele im Bereich der Medienregulierung und -förderung zu erreichen. Durch solche Partnerschaften können sie ihre Ressourcen effizienter nutzen, Fachkenntnisse bündeln und innovative Lösungen entwickeln. Dies schließt die Förderung von Forschungsprojekten in den Themenkomplexen Medienkompetenz und Integration von KI in die Medienkompetenzvermittlung ein.

Die Nutzung von KI-gesteuerten Technologien etwa in automatisierten Lernplattformen, Chatbots, Analysewerkzeugen, personalisierten Nachrichtenaggregatoren und Serious Games eröffnet neue Möglichkeiten, die Medienkompetenz auf innovative und effektive Weise auszubilden. Durch die Zusammenarbeit mit Technologieunternehmen und Bildungseinrichtungen können die Medienanstalten als Kooperationspartner dazu beitragen, den Einsatz von KI in der Medienkompetenzvermittlung voranzutreiben. Diese Verbindung von etablierten Praktiken und modernen Technologien stellt einen zukunftsweisenden Ansatz dar, um die Medienkompetenz in unserer digitalen Gesellschaft nachhaltig zu fördern.

6 Handlungsfelder

KI eröffnet ein breites Spektrum an Möglichkeiten und kann diverse Mehrwerte generieren. Allerdings bergen falsche oder kriminelle Anwendungen dieser Technologie erhebliche Risiken sowohl für individuelle Akteure als auch für ganze Staaten. Um dieser ambivalenten Natur von KI zu begegnen, sind angemessene Strategien erforderlich. Ein singulärer Ansatz erweist sich hierbei als unzureichend; vielmehr bedarf es einer kombinierten Herangehensweise, die die Beteiligung verschiedener Akteure einschließt.

Die nachstehenden Handlungsfelder wurden mit dem Ziel konzipiert, Optionen aufzuzeigen, wie die Potenziale von KI erschlossen und gleichzeitig die damit verbundenen Risiken minimiert werden können. Neben technologischen Lösungen und regulatorischen Maßnahmen werden Wege aufgezeigt, wie die Medienkompetenz gesteigert werden kann. Eine gut ausgebildete Medienkompetenz ist der Schlüssel, um die vermeintlich konkurrierenden Ziele „Hebung der Potenziale“ und „Reduzierung der Risiken“ parallel zu erreichen.

Jedem Handlungsfeld sind relevante Akteure zugeordnet, um eine ganzheitliche Herangehensweise zu gewährleisten.

Handlungsfeld Technologie
Technologische Ansätze können dabei unterstützen, Vertrauen zu schaffen. Die Sicherstellung der Identität von Personen, die einfache Erkennung synthetischer Inhalte, die Vorbeugung von Fälschungen und die Nachvollziehbarkeit der Herkunft von Inhalten tragen zu diesem Ziel bei.

Es ist die Aufgabe der institutionellen und privatwirtschaftlichen Forschung die genannten Ansätze konsequent weiterzuentwickeln, sodass diese fälschungssicher und unlöschbar sind und auch bleiben. Außerdem ist eine möglichst nutzerfreundliche Anwendbarkeit anzustreben. Seitens der Plattformen und der Medienschaffenden ist eine weitreichende Anwendung und Implementierung der genannten Ansätze zu forcieren, insbesondere im Bereich der Kennzeichnungen und Herkunftsnachweise, denn nur so können diese ihre volle Effektivität entfalten. Die Medienkonsument:innen sind gleichermaßen angehalten, diese Methoden anzuwenden. Dafür ist es von zentraler Bedeutung, dass die dafür notwendige Medienkompetenz vermittelt wird. Dies umfasst sowohl das Verständnis, warum derartige Maßnahmen notwendig sind, als auch die Kompetenzen diese anwenden zu können.

Authentifizierung
Um Vertrauen zu schaffen, ist es entscheidend, in der digitalen Welt sowohl Inhalte als auch, in gewissen Kontexten, Personen zu authentifizieren.

Kryptographische Herkunftsnachweise können dazu genutzt werden, um nachzuweisen, welchen Ursprung Inhalte haben und ob Modifikationen am Original vorgenommen wurden. Diesem Ansatz widmen sich bereits verschiedene Initiativen wie etwa Valid, Truepic und C2PA. Anwenden sollten diese Werkzeuge sowohl die Urheber:innen von Inhalten, als auch Plattformbetreiber. Übergreifende, vereinheitlichte Kennzeichnungen können mit externen Faktencheck-Initiativen geteilt werden, um sicherzustellen, dass verifizierte reale, falsche und bewusst irreführende Inhalte auf allen Plattformen bei einem erneuten Hochladen ohne Zeitverzögerung identisch identifiziert werden.

Die Implementierung von Authentifizierungsprotokollen in relevanter Software ist notwendig, um zu verhindern, dass Individuen sich in virtuellen Besprechungen oder Veranstaltungen mithilfe von Deepfake-Technologien oder interaktiven synthetischen Avataren als andere Personen ausgeben können.

Seitens der Medienschaffenden und Anwendenden ist eine weitreichende Implementierung dieser Ansätze essenziell, denn dies ist die Voraussetzung für eine effektive Wirksamkeit. Weitere Forschung wird benötigt, um diese Methoden fälschungssicher und unlöschbar zu gestalten. Gleichzeitig müssen Umsetzungen gefunden werden, die gleichermaßen einfach zu implementieren und zu erfassen sind.

Detektion
Um eine Umgehung der Authentifizierungsmaßnahmen zu erkennen, werden fortschrittliche Methoden und Werkzeuge benötigt, die imstande sind, synthetische Inhalte und Avatare zu detektieren. Weitere Forschungsbedarfe bestehen, um eine zuverlässige Erkennung sicherzustellen. Darüber hinaus müssen Prozesse etabliert werden, um falschpositive Einordnungen anfechten zu können.

Ein weitreichender öffentlicher Zugang zu derartigen Werkzeugen bietet, verbunden mit der entsprechenden technischen Kompetenz, Bürger:innen die Möglichkeit selbstständig zu prüfen, ob Inhalte vertrauenswürdig sind.

Kennzeichnung
Die Kennzeichnung der Charakteristika von KI-Anwendungen bietet für Entwickler:innen und für Anwender:innen von KI-Systemen eine Leitlinie und schafft somit Sicherheit und Vertrauen. Entscheidungsträger:innen können schnell erkennen, ob ein KI-System für ihre Zwecke geeignet ist oder nicht. Die Vergabe dieser Kennzeichnungen sollte idealerweise durch ein unabhängiges Prüfinstitut geschehen, um eine einheitliche und konsistente Kennzeichnung zu gewährleisten.

Handlungsfeld Regulierung
In gewissen Bereichen müssen gesetzliche Regularien dabei unterstützen, Missbrauch von technologischen Potenzialen zu verhindern. Es ist die Aufgabe der Politik, Technologien, die inakzeptable Risiken bergen, zu verbieten. Gleichzeitig dürfen Innovationen dadurch nicht gehemmt werden. Es können jedoch nicht alle Bereiche durch den Gesetzgeber reglementiert werden. An dieser Stelle können freiwillige Standards und Selbstregulierungen Transparenz schaffen und damit Vertrauen in die mit Unterstützung von KI erstellten Inhalte erhöhen.

Umsetzung von EU-Regulierungen
Europa hat mit dem Digital Services Act (DSA) und dem Artificial Intelligence Act (AI Act) weltweit eine Vorreiterrolle bei der Regulierung von KI eingenommen, die das Potenzial hat, global Akzente zu setzen. Die zentrale Aufgabe ist es nun, diese Regulierungen auf nationaler und europäischer Ebene entsprechend umzusetzen und die dafür nötigen Strukturen und Organisationen zu schaffen.

Standards und Selbstregulierung
Zahlreiche deutsche und internationale Medienhäuser haben sich bereits zur Einhaltung von Standards und Selbstregulierungen verpflichtet. Um das Vertrauen in die Medien zu stärken, ist eine breite Anwendung dieser Standards und Selbstregulierungen wünschenswert. Die Schaffung von Selbstregulierungsorganisationen kann das Vertrauen weiter erhöhen, indem die Einhaltung unabhängig überprüft wird.

Handlungsfeld Medienkompetenz
Technologische Ansätze etwa zur Kennzeichnung oder Erkennung synthetischer Inhalte können genauso wie regulatorische Maßnahmen immer nur unterstützend wirken. Im Zentrum steht eine umfassende Medienkompetenz. Diese muss Technologiekompetenz – also das Verständnis der Nutzer:innen über die Anwendungen, Implikationen und Möglichkeiten der KI-Technologie – einschließen, denn dies bildet die Grundlage für weitere Kompetenzen, wie etwa Medienanalyse oder Medienkritik. Der Vermittlung letztgenannter Kompetenzen kommt dabei weiterhin eine hohe Bedeutung zu.

Im Folgenden werden Ansätze dargestellt, wie Medienkompetenz vermittelt werden kann. Neben den klassischen Bildungseinrichtungen wie Schulen und Universitäten spielen hierbei private und öffentliche Initiativen, aber insbesondere auch die Medienanstalten eine zentrale Rolle, indem sie in diesem Kontext als Förderer, Kooperationspartner und Multiplikatoren tätig werden.

Damit die Maßnahmen zur Vermittlung von Medienkompetenz den gewünschten systemischen Effekt erzielen, ist es entscheidend, dass große Teile der Bevölkerung erreicht werden. Daher sollten die verschiedenen Ansätze möglichst vielen Menschen, mit verschiedensten Hintergründen und Lebenssituationen möglichst leicht zugänglich gemacht werden.

Forschungsförderung
Für den nachhaltigen Erfolg von Maßnahmen zur Medienkompetenzsteigerung sind begleitende Forschungsprojekte notwendig. Es ist einerseits entscheidend zu wissen, welche Kompetenzen in verschiedenen Bevölkerungsschichten bereits vorhanden sind und wie die jeweiligen Bevölkerungsschichten erreicht werden können. Andererseits ist eine konsequente Evaluierung der Effekte von Maßnahmen zentral, um diese stetig weiterentwickeln zu können. Es sollten daher mehr Forschungsprojekte zum Status der Technologie- und Medienkompetenz, insbesondere bei Kindern und Jugendlichen, angestoßen werden und bestehende Maßnahmen systematischer evaluiert werden.

Klassische Methoden zur Medienkompetenzvermittlung
Klassische pädagogische Methoden bleiben bei der Förderung von Medienkompetenz weiterhin relevant. Dazu zählen Handbücher und Leitfäden, die es ermöglichen, Inhalte jederzeit nachzulesen. Gleichermaßen wichtig sind praxisorientierte Angebote, die persönliche Erfahrungen sowie einen interaktiven Austausch mit Expert:innen ermöglichen. Derartige Formate können online oder vor Ort stattfinden. Das entsprechende Lehrmaterial sollte regelmäßig an die aktuellen Anforderungen angepasst werden. Klassische pädagogische Methoden sollten daher weiterhin mit der Zielsetzung möglichst viele Menschen zu erreichen, eingesetzt werden.

Unterstützung von Bildungseinrichtungen
Das Engagement von Medienpädagog:innen und Journalist:innen in Bildungseinrichtungen vermittelt praxisnahe Einblicke in Medien und Journalismus. Diese Zusammenarbeit fördert das Verständnis der Schüler:innen für Medien und trägt zu einem bewussten Medienverhalten bei. Derartige Angebote sollten idealerweise über alle Alters- und Bildungsniveaus hinweg angeboten werden.

Medienkompetenzvermittlung auf digitalen Plattformen
Gerade um Jugendliche gezielt zu erreichen, sollten neben klassischen Methoden aber auch Kampagnen auf Plattformen, die Jugendliche verstärkt nutzen, initiiert werden. Solche Kampagnen können ein bewusstes und kritisches Verständnis für Medieninhalte fördern, indem sie gezielt auf Inhalte eingehen, die auf der jeweiligen Plattform relevant sind.

Prebunking-Kampagnen
Durch das Zeigen entsprechender Hinweise (Prebunking) werden Menschen nachweislich weniger empfänglich für manipulative Botschaften. Derartige Hinweise sollten Anleitungen geben, wie kritische Nutzer:innen von Nachrichten wahrheitsgemäße Informationen von absichtlich irreführenden Quellen unterscheiden können, ohne dabei blind auf bestimmte Quellen zu vertrauen. Um möglichst viele Bürger:innen mit diesen Kampagnen zu erreichen, sollten Kooperationen mit großen Plattformbetreibern sowie dem öffentlich-rechtlichen und dem privaten Rundfunk angestrebt werden.

KI-basierte Lernplattformen
Der Einsatz von KI-basierten Lernplattformen ermöglicht personalisierte Medienkompetenzkurse. Individuelle Lernpfade fördern eine differenzierte Medienkompetenzentwicklung, während die kritische Auseinandersetzung mit KI ein reflektiertes Verständnis der Technologie fördert. Um die Medienkompetenz gezielt zu fördern, wird empfohlen, eine kritische Auseinandersetzung mit KI in Medienkompetenzkurse zu integrieren. Dies trägt dazu bei, ein reflektiertes Verständnis der Technologie zu entwickeln, und stärkt die Fähigkeit, KI-generierte Inhalte kritisch zu hinterfragen. Die Entwicklung solcher Lernplattformen sollte mit dem Wissen aus den traditionellen Maßnahmen unterstützt und die resultierenden Plattformen anschließend möglichst vielen Bürger:innen zugänglich gemacht werden.

Serious Games
Serious Games können durch ihren spielerischen Charakter dabei unterstützen, jüngeren Menschen ernste Themen zu vermitteln. So können beispielsweise Fragen aus dem Bereich der Technologie- und Medienkompetenz zielgruppengerecht aufbereitet werden. KI kann in Serious Games und Simulationen integriert werden, um realitätsnahe Situationen zu schaffen und beispielsweise menschenähnliche Dialoge zu generieren und damit zu einer ansprechenden Umsetzung beitragen. Die Chancen von Serious Games für die Medienkompetenzvermittlung sollten genutzt werden, um gerade Kinder und Jugendliche für das Thema zu sensibilisieren.

Medienkompetenzvermittlung für Medienschaffende
KI-Anwendungen bieten für Medienschaffende zahlreiche Potenziale. Damit diese Chancen genutzt werden können, ist jedoch ein umfangreiches Verständnis im Umgang mit den neuen Technologien Voraussetzung. Der Umgang mit KI-Anwendungen sollte daher fester Bestandteil der journalistischen Ausbildung sein und etwa in Journalistenschulen und Universitäten vermittelt werden. Der Kompetenzaufbau in Redaktionen sollte unterstützt werden, beispielsweise durch aktive Vermarktung entsprechender Angebote oder die Unterstützung neuer Maßnahmen, die sich gezielt an Journalist:innen richten.

7 Ausblick

Künstliche Intelligenz (KI) hat die Medienlandschaft bereits signifikant verändert und gerade durch das Aufkommen generativer KI deutlich geprägt. Wenngleich Prognosen in einem derart dynamischen Feld kaum möglich bzw. sinnvoll erscheinen, kann wohl davon ausgegangen werden, dass KI-Technologien in den kommenden Jahren sowohl seitens der Medienschaffenden als auch der Medienkonsument:innen immer breiter genutzt werden und daher der Einfluss von KI in den kommenden Jahren weiter zunehmen wird.

Im folgenden Kapitel werden einzelne Entwicklungen, die sich bereits abzeichnen, dargestellt. Diese Entwicklungen haben das Potenzial, für große Umwälzungen in der näheren Zukunft zu sorgen, daher ist es entscheidend, sich die damit verbundenen Chancen und Risiken bewusst zu machen und entsprechend zu handeln. Die bereits bestehenden Herausforderungen sollten dabei jedoch nicht in den Hintergrund treten, denn zukünftige Entwicklungen können sich ganz anders manifestieren als hier dargestellt.

Automatische Inhaltsgenerierung
Eine Entwicklung, die bereits losgetreten wurde, ist die Nutzung von KI zur automatisierten Generierung von Inhalten. Bisher waren derartige Einsätze bei seriösen Medienhäusern eher auf Formate beschränkt, die festen Regeln folgten, wie beispielsweise Fußball- oder Börsenberichterstattungen. Es kann aber davon ausgegangen werden, dass sich dieser Trend mit zunehmenden sprachlichen und analytischen Fähigkeiten von KI in Zukunft nicht nur fortsetzen, sondern sogar intensivieren wird. In der künftigen Entwicklung der Textgeneration wird ein bedeutender Fortschritt darin bestehen, dass Systeme in der Lage sein werden, eigenständig Fakten zu recherchieren und diese nahtlos in generierte Texte zu integrieren. Diese Forschungsrichtung wird durch verschiedene Ansätze vorangetrieben, darunter die direkte Nutzung des in den Trainingsdaten enthaltenen Wissens, wie von Nayeon Lee et al.[204] beschrieben, oder die Implementierung von Plug-Ins, wie von Miaoran Li et al.[205] vorgeschlagen. Dabei stellen sich mehrere Herausforderungen, einschließlich des sorgfältigen Kuratierens der Trainingsdaten, des effektiven Abgleichs des Modells mit aktuellen Informationen und der Bewältigung der Frage, wie Falschinformationen zuverlässig herausgefiltert werden können.

Darüber hinaus wird voraussichtlich die Bedeutung von Audio- und Videoformaten für die Nachrichtenvermittlung zunehmen.[206] Durch KI-Werkzeuge, wie etwa Text-to-Speech oder synthetische Avatare, wird die Erzeugung von derartigen Formaten deutlich günstiger und einfacher.

Dies bedeutet für Redaktionen Umstellungen, gerade in den Anforderungen, die an Mitarbeiter:innen gestellt werden. Aber auch rechtliche Herausforderungen, etwa im Kontext vom Urheberrecht, werden damit verbunden sein. Die Schaffung von Rechtssicherheit seitens des Gesetzgebers und der Gerichte ist eine wichtige Voraussetzung, um die damit verbundenen Potenziale zu heben.

Regulatorik und Kennzeichnung
Mit der Implementierung des Digital Services Act (DSA) ist zu erwarten, dass die Regulierungsbehörden in Europa verstärkt darauf drängen werden, die Vorschriften für digitale Plattformen durchzusetzen und gegebenenfalls von ihrem Sanktionsrecht Gebrauch zu machen. Der Schwerpunkt liegt hierbei insbesondere auf der Bekämpfung von Hassrede und Desinformation, insbesondere

204 Vgl. Lee et al. 2020
205 Vgl. Li et al. 2023
206 Vgl. Reuters 2023

im Zusammenhang mit Wahlprozessen. Ebenso beabsichtigt der voraussichtlich 2026 in Kraft tretende Artificial Intelligence Act (AI Act), Maßnahmen gegen Desinformation zu ergreifen. Zu den Zielen gehört unter anderem die Verpflichtung großer Plattformen, KI-Inhalte zu identifizieren und entsprechend zu kennzeichnen. Die Konzerne Meta und Alphabet haben bereits Richtlinien erlassen, die politische Werbung dazu verpflichten, jegliche Verwendung von KI bei der Erstellung von Werbeinhalten offenzulegen. Bisher fehlen jedoch transparente Mechanismen zur Gewährleistung der Einhaltung dieser Richtlinien. Es wird somit von Interesse sein, wie sich die Dynamik zwischen Nutzer:innen, Plattformen und Regulierungsbehörden in diesem Bereich entwickeln werden.[207]

Persönliche Assistenten
Einer der übergeordneten Trends, die von KI und insbesondere von generativer KI getrieben werden, ist die immer stärkere Personalisierung von Medieninhalten (siehe Kapitel 3.2.2). Der Trend hat positive wie negative Effekte: Einerseits ermöglicht eine Personalisierung, beispielsweise der Sprache oder der Komplexität der Darstellung, einem größeren Teil der Bevölkerung Zugang zu Inhalten. Andererseits besteht die Gefahr, dass die Personalisierung von Nachrichteninhalten die Polarisierung der Gesellschaft weiter vorantreibt. Denn wenn neben der Struktur auch der Inhalt von Nachrichten verändert, beziehungsweise gefiltert wird, kann dies die Entstehung von Echokammern bzw. Filterblasen begünstigen.

Dieser Trend könnte noch verstärkt werden, wenn generative KI, wie von vielen Expert:innen erwartet, dazu genutzt wird, persönliche Assistenten zu schaffen, die das (digitale) Leben der Bürger:innen unterstützen und verwalten. Es wäre beispielsweise denkbar, dass ein solcher Assistent dafür verwendet wird, Nachrichten für Nutzer:innen individuell zu personalisieren (siehe Kapitel 5.2.2). Bisher eignen sich Chatbots kaum als Nachrichtenquelle und Recherchemedium, was sich jedoch zeitnah ändern könnte. Entsprechende Ansätze existieren bereits.[208] Über entsprechende Prompts können relevante Nachrichten zu verschiedenen Themen gesammelt werden und entsprechend den Vorgaben in Sprache, Länge und Komplexität angepasst werden. Fehler, die beispielsweise bei der Übersetzung oder Zusammenfassung von Inhalten geschehen, könnten dazu beitragen, dass Inhalte sowie die darin enthaltenen Botschaften und Nuancen verfälscht werden. Hier ist es wichtig, Regeln für die Auswahl der Inhalte zu implementieren und transparente KI-Systeme einzusetzen, bei denen die Wahrscheinlichkeit derartiger Fehler möglichst gering ausfällt. Ebenfalls denkbar wäre eine Zusammenarbeit mit Nachrichtenorganisationen, die ihre Inhalte passend an die beliebtesten Formate anpassen, prüfen und dann persönlichen Assistenten zur Verfügung stellen.

Persönliche Assistenten könnten auch zu Gesprächspartnern werden. Daraus ergeben sich beispielsweise Möglichkeiten für einsame Menschen oder im Bereich der Bildung. Ähnlich wie bei Online-Videospielen könnte diese Anwendung sogar dazu beitragen, soziale Kompetenzen zu erweitern, wobei jedoch die potenzielle Gefahr der sozialen Isolation in der physischen Welt nicht unterschätzt werden sollte.[209]

207 Vgl. Reuters 2023
208 Vgl. zum Beispiel Li et al. 2023
209 Vgl. Tushya / Abraham 2023

Im Kontext von Medienkompetenzbildung würden sich hier Chancen ergeben, den Nutzer:innen Einordnungen der präsentierten Nachrichten oder anderer Inhalte mitzugeben und die Möglichkeit zur Diskussion über Medienkompetenz anzubieten. Um ethischen Fragestellungen zu begegnen, ist es entscheidend, dass diese Funktionalitäten in der Design- und Entwicklungsphase persönlicher Assistenten bereits mitgedacht werden.

Auch diese Entwicklungen werfen rechtliche Fragen, etwa zum geistigen Eigentum, auf. Für Medienhäuser bietet dies jedoch die Möglichkeit neue Geschäftsmodellinnovationen voranzutreiben. So könnten beispielsweise fortschrittliche Personalisierungsmöglichkeiten nur über Abo-Modelle zugänglich gemacht werden.[210]

210 Vgl. Reuters 2023

Projekt

Auftragnehmer
acatech – Deutsche Akademie der Technikwissenschaften

Projektleitung
Dr. Anna Frey, acatech Geschäftsstelle

Autor:innen
Dr. Paul Grünke, acatech Geschäftsstelle
Simon Litsche, acatech Geschäftsstelle
Sandra Starchenko, acatech Geschäftsstelle

Befragte Expert:innen
Dr. Alessandro Barberi, Universität Wien
Prof. Dr. Christoph Bieber, Center for Advanced Internet Studies
Prof. Dr. Thomas Goll, Technische Universität Dortmund
Uli Köppen, Bayerischer Rundfunk
Prof. Dr. Christoph Neuberger, Freie Universität Berlin
Prof. Dr. Jan-Hendrik Passoth, European New School of Digital Studies
Prof. Dr. Dr. Frauke Rostalski, Universität zu Köln
Prof. Dr. Thorsten Thiel, Universität Erfurt
Prof. Dr. Antje von Ungern-Sternberg, Universität Trier
Prof. Dr. Ruth Wendt, Ludwig-Maximilians-Universität München

Projektlaufzeit
09/2023 – 02/2024

Literaturverzeichnis

AI Ethics Impact Group, (2020). From Principles to Practice: An interdisciplinary framework to operationalise AI ethics. Verfügbar unter https://www.ai-ethics-impact.org/resource/blob/1961130/c6db9894ee73aefa489d6249f5ee2b9f/aieig---report---download-hb-data.pdf

Fraunhofer AISEC, (2023). *»Deepfakes«: Mit KI-Systemen Audio- und Videomanipulationen verlässlich entlarven*. Verfügbar unter https://www.aisec.fraunhofer.de/de/das-institut/wissenschaftliche-exzellenz/Deepfakes.html

Aldhous, P. (2017). *We Trained A Computer To Search For Hidden Spy Planes. This Is What It Found*. Verfügbar unter https://www.buzzfeednews.com/article/peteraldhous/hidden-spy-planes

AlgorithmWatch, (2022). *Ein Leitfaden zum Digital Services Act: Das neue EU-Gesetz soll den großen Tech-Konzernen Zügel anlegen*. Verfügbar unter https://algorithmwatch.org/de/dsa-erklaert/

AlignedAI, (2023). *Using faAIr to measure gender bias in LLMs*. Verfügbar unter https://buildaligned.ai/blog/using-faair-to-measure-gender-bias-in-llms

Angrick, A. (2023). „„Neue Geschichten vom Pumuckl': Kult-Kobold läuft im Dezember im Stream und Free-TV." Verfügbar unter https://www.augsburger-allgemeine.de/panorama/tv/pumuckl-neuauflage-auf-rtl-start-free-tv-besetzung-handlung-folgen-von-neue-geschichten-vom-pumuckl-infos-am-13-12-2023-id62138636.html

Antenne Deutschland, (2023). *ABSOLUT RADIO AI: Der erste AI-moderierte Radio-Stream ist da!* Verfügbar unter https://www.antenne-deutschland.de/absolut-radio-ai-der-erste-ai-moderierte-radio-stream-ist-da/

appliedAI, (2023). Risikoklassifizierung_appliedAI_Final_März-27-2023. Retrieved from https://aai.frb.io/assets/files/Studie-Risiskoklassifizierung_appliedAI_Final_M%C3%A4rz-27-2023.pdf

ARD/ZDF-Forschungskommission, (2024). „ARD/ZDF-Onlinestudie 2023: Normalisierung der Internetnutzung nach den Corona-Jahren." Verfügbar unter https://www.ard-zdf-onlinestudie.de/

Argyle, L. P., Busby, E., Gubler, J., Bail, C., Howe, T., Rytting, C., & Wingate, D. (2023). AI Chat Assistants can Improve Conversations about Divisive Topics. arXiv preprint arXiv:2302.07268.

Arnold, M., Bellamy, R. K., Hind, M., Houde, S., Mehta, S., Mojsilović, A., ... & Varshney, K. R. (2019). FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development*, 63 (4/5), 6–1

Baacke, D. (1996). „Medienkompetenz: Begrifflichkeit und sozialer Wandel." In: Rein, A. v. (Hrsg.), *Medienkompetenz als Schlüsselbegriff*. Bad Heilbrunn, S. 112–124.

Babakar, M. & Sud, A. (2023). New features coming to Fact Check Explorer. Verfügbar unter https://blog.google/products/news/new-features-coming-to-fact-check-explorer/

Baker, E. (2023). *Voice Cloning 101: The Technology Behind Authentic-Sounding Voice Simulations*. Verfügbar unter https://www.veritonevoice.com/blog/voice-cloning-101/

Baker, S., Kocieniewski, D. & Smith, M. (2018). Cambridge Analytica's History of Dubious Election Tricks. Verfügbar unter https://www.bloomberg.com/politics/articles/2018-03-20/cambridge-analytica-has-long-history-of-dubious-election-tricks

Balamurali, B. T., Lin, K. E., Lui, S., Chen, J. M., & Herremans, D. (2019). Toward robust audio spoofing detection: A detailed comparison of traditional and learned features. *IEEE Access,* 7, 84229–84241.

Beam, M. A. (2014). "Automating the news: How personalized news recommender system design choices impact news reception." *Communication Research,* 41(8), 1019–1041.

Beck, D. (2023). *So könnte das Radio der Zukunft aussehen*. Verfügbar unter https://www.tagesschau.de/wissen/technologie/zukunft-radio-100.html

Behringer, W. & Jerouschek, G.. (2000). Heinrich Kramer (Institoris). *Der Hexenhammer. Malleus maleficarum. München.*

Bewarder, K., et al. (2023). „Russland provoziert die Türkei – Die perfide Methode des Kremls." Verfügbar unter https://www.tagesschau.de/investigativ/ndr-wdr/russland-provokationen-tuerkei-100.html

bidt, (2024). „Coding Public Value: Gemeinwohlorientierte Software für öffentlich-rechtliche Medienplattformen." Verfügbar unter https://www.bidt.digital/forschungsprojekt/coding-public-value-gemeinwohlorientierte-software-fu%CC%88r-oeffentlich-rechtliche-medienplattformen/

Bayerische Landeszentrale für neue Medien (BLM), (2017). „Dein Algorithmus – meine Meinung! Algorithmen und ihre Bedeutung für Meinungsbildung und Demokratie." München.

Bayerische Landeszentrale für neue Medien (BLM), (2024). „Extremismus im Netz." Verfügbar unter https://www.blm.de/de/wir-regulieren/jugend_und_nutzerschutz/themen-im-fokus/extremismus.cfm

Bond, R., Fariss, C., Jones, J. et al. (2012). „A 61-million-person experiment in social influence and political mobilization." *Nature*, 489, 295–298. https://doi.org/10.1038/nature11421

Brown, A. (2019). "Scammer Successfully Deepfaked CEO's Voice To Fool Underling Into Transferring $243,000" Verfügbar unter https://gizmodo.com/scammer-successfully-deepfaked-ceos-voice-to-fool-under-1837835066

Cercone, J. (2022). "There are no U.S.-run biolabs in Ukraine: Contrary to social media posts." Verfügbar unter https://www.politifact.com/factchecks/2022/feb/25/tweets/there-are-no-us-run-biolabs-ukraine-contrary-socia/

Clark, M. (2023). *LinkedIn's flood of 'collaborative' articles start out with AI prompts*. Verfügbar unter https://www.theverge.com/2023/3/4/23624241/linkedin-collaborative-articles-ai-prompts-content

Cole, M. (2023). "A DIY coder created a virtual AI waifu." Verfügbar unter https://www.vice.com/en/article/jgpzp8/a-diy-coder-created-a-virtual-ai-waifu-chatgpt

Correctiv, (2024). „Falschmeldungen einreichen: CORRECTIV.Faktencheck ist auf Whatsapp erreichbar." Verfügbar unter https://correctiv.org/faktencheck/ueber-uns/2020/05/12/correctiv-faktencheck-ist-auf-whatsapp-erreichbar/

Council of Europe, (2023). “Guidelines on the responsible implementation of artificial intelligence.” Verfügbar unter https://rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6

Dachwitz, I. & Rudl, T. (2018). *Was wir über den Skandal um Facebook und Cambridge Analytica wissen.* Verfügbar unter https://netzpolitik.org/2018/cambridge-analytica-was-wir-ueber-das-groesste-datenleck-in-der-geschichte-von-facebook-wissen/#netzpolitik-pw

Dampz, N. (2023). *Hollwood-Autoren beenden Streik.* Verfügbar unter https://www.tagesschau.de/wirtschaft/unternehmen/hollywood-streik-ende-102.html

DataSkop, (2024). *Was ist DataSkop?* Verfügbar unter https://dataskop.net/ueber/

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media.* Cambridge, MA: Harvard University Press.

Diaz, J. (2023). “AI feels like a magic act. By 2033, it will be a horror movie.” Verfügbar unter https://www.fastcompany.com/90873422/ai-feels-like-a-magic-act-by-2033-it-will-be-a-horror-movie

Die Medienanstalten, (2021a). „Fakt oder Fake? Jugendschutz, Medienkompetenz und Desinformation.“ Verfügbar unter https://www.die-medienanstalten.de/publikationen/jugendschutz-medienkompetenzbericht/fakt-oder-fake-jugendschutz-medienkompetenz-und-desinformation

Die Medienanstalten, (2021b). „Stellungnahme der Direktorenkonferenz der Landesmedienanstalten zur Digital Services Act (DSA) und zum Digital Markets Act (DMA).“ Verfügbar unter https://www.die-medienanstalten.de/fileadmin/user_upload/die_medienanstalten/Ueber_uns/Positionen/20210330_DSA_DMA_Stellungnahme_DLM_final.pdf

DiResta, R. (2020). “AI-generated text is the scariest deepfake of all.” Verfügbar unter https://www.wired.com/story/ai-generated-text-is-the-scariest-deepfake-of-all/

dpa, (2023). *Faktencheck23 – Ein Projekt der dpa und der Google News Initiative.* Verfügbar unter https://www.dpa.com/de/faktencheck23

Dreyer, S., Stanciu, E., Potthast, K. C., & Schulz, W. (2021). *Gutachten Desinformation.* Verfügbar unter https://www.medienanstalt-nrw.de/fileadmin/user_upload/NeueWebsite_0120/Themen/Desinformation/Leibnitz-Institut_LFMNRW_GutachtenDesinformation.pdf

Eady, G., Paskhalis, T., Zilinsky, J., et al. (2023). “Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior.” *Nature Communications, 14,* 62. https://doi.org/10.1038/s41467-022-35576-9

Egelhofer, J. L., Aaldering, L., Eberl, J. M., Galyga, S., & Lecheler, S. (2020). From novelty to normalization? How journalists use the term “fake news” in their reporting. *Journalism Studies,* 21(10), 1323–343.

European Commission, (2022a). “Code of Practice on Disinformation.” Verfügbar unter https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

European Commission, (2022b). “Signatories 2022 – Strengthened Code of Practice on Disinformation.” Verfügbar unter https://digital-strategy.ec.europa.eu/en/library/signatories-2022-strengthened-code-practice-disinformation

European Commission, (2023a). “Press release: European Commission proposes measures to tackle disinformation online.” Verfügbar unter https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6453

European Commission, (2023b). *Share of respondents who used the following media every day or almost every day in the European Union from 2011 to 2023.* Statista. Verfügbar unter https://www.statista.com/statistics/422572/europe-daily-media-usage/

European Parliament, (2023). „Eurobarometer: Fake News and Disinformation Online.” Verfügbar unter https://europa.eu/eurobarometer/surveys/detail/3153

Fallis, D. (2015). What is disinformation?. *Library trends,* 63(3), 401–426.

Feng, K. K., Ritchie, N., Blumenthal, P., Parsons, A., & Zhang, A. X. (2023). “Examining the Impact of Provenance-Enabled Media on Trust and Accuracy Perceptions.” *Proceedings of the ACM on Human-Computer Interaction,* 7(CSCW2), 1–42.

Flawless AI, (2023). *Flawless Demo.* Verfügbar unter https://vimeo.com/781894404

Fletcher, R., Schifferes, S., & Thurman, N. (2020). “Building the ‘Truthmeter’: Training algorithms to help journalists assess the credibility of social media sources.” *Convergence,* 26(1), 19–34. https://doi.org/10.1177/1354856517714955

Foote, K. D. (2022). The history of machine learning and its convergent trajectory towards AI. *Machine Learning and the City: Applications in Architecture and Urban Design,* 1299–142.

G7, (2023). “G7 Leaders’ Statement.” Verfügbar unter https://www.mofa.go.jp/files/100573473.pdf

Galileo, (2023). „Künstliche Intelligenz: Die Risiken und Chancen der KI.“ Verfügbar unter https://www.prosieben.de/serien/galileo/news/kuenstliche-intelligenz-die-risiken-und-chancen-der-ki-329392

Giardina, C. (2015). “How ‘Furious 7’ Brought Back Late Paul Walker.” Verfügbar unter https://www.hollywoodreporter.com/movies/movie-news/how-furious-7-brought-late-845763/

Giardina, C. (2016). “Rogue One: How Grand Moff Tarkin and Princess Leia Were Resurrected.” Verfügbar unter https://www.hollywoodreporter.com/movies/movie-news/rogue-one-how-grand-moff-tarkin-peter-cushing-returned-957258/

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT press.

Goldberg, Y. (2023). „Prebunking-Kampagne gegen Falschinformationen.“ Verfügbar unter https://blog.google/intl/de-de/unternehmen/technologie/prebunking-kampagne-gegen-falschinformationen/

Goldhammer, K., et al. (2019). *Künstliche Intelligenz, Medien und Öffentlichkeit. Wissenschaftlicher Bericht im Auftrag des Bundesamts für Kommunikation – BAKOM, Bern.* Verfügbar unter https://www.bakom.admin.ch/dam/bakom/de/dokumente/bakom/elektronische_medien/Zahlen%20und%20Fakten/Studien/studien-kuenstliche-intelligenz-medien-oeffentlichkeit.pdf.download.pdf/K%C3%BCnstliche%20Intelligenz,%20Medien%20und%20%C3%96ffentlichkeit.pdfent

Growcoot, M. (2023). “Photographer Creates Lifelike Social Media Influencer Using Only AI.” Verfügbar unter https://petapixel.com/2023/05/23/photographer-creates-lifelike-social-media-influencer-using-only-ai/

Haase, M. (2023). *KI in der Musik: „In manchen Aspekten wesentlich kreativer als wir".* Verfügbar unter https://www.xplr-media.com/de/xplr-magazin/ki-in-der-musik-in-manchen-aspekten-wesentlich-kreativer-als-wir.html

Hanley, H. W. & Durumeric, Z. (2023). Machine-Made Media: Monitoring the Mobilization of Machine-Generated Articles on Misinformation and Mainstream News Websites. *arXiv preprint arXiv:2305.09820.*

Hamborg, F., et al. (2019). "Automated identification of media bias in news articles: an interdisciplinary literature review." *International Journal on Digital Libraries,* pp. 391-415. https://doi.org/10.1007/s00799-018-0261-y

Heesen, J., et al. (2020a). *Ethik-Briefing: Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen.* Whitepaper aus der Plattform Lernende Systeme, München. Verfügbar unter: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG3_Whitepaper_EB_200831.pdf

Heesen, J., et al. (2020b). *Zertifizierung von KI-Systemen: Kompass für die Entwicklung und Anwendung vertrauenswürdiger KI-Systeme.* Whitepaper aus der Plattform Lernende Systeme, München. Verfügbar unter: https://www.acatech.de/publikation/zertifizierung-von-ki-systemen-kompass-fuer-die-entwicklung-und-anwendung-vertrauenswuerdiger-ki-systeme/

Heesen, J., et al. (2021). *KI-Systeme und die individuelle Wahlentscheidung – Chancen und Herausforderungen für die Demokratie.* Whitepaper aus der Plattform Lernende Systeme, München. https://doi.org/10.48669/pls_2021-1

Heesen, J. (2022). „KI." In: *Journalistikon – Das Wörterbuch der Journalistik.* Hg. v. Katja Artsiomenka/Horst Pöttker, Herbert von Halem Verlag, 2022. Online unter: https://journalistikon.de/ki/

Heesen, J. (2023). Kennzeichnungspflichten für KI aus Perspektive der Ethik. *BvD-News* 2/23, 10–13

Heesen et al, (2023): *Künstliche Intelligenz im Journalismus. Potenziale und Herausforderungen für Medienschaffende.* Whitepaper aus der Plattform Lernende Systeme, München. https://doi.org/10.48669/pls_2023-1

Hegelich, S., & Serrano, R. A. (2019). „Microtargeting in Deutschland bei der Europawahl 2019." Verfügbar unter: https://www.medienanstalt-nrw.de/fileadmin/user_upload/lfm-nrw/Foerderung/Forschung/Dateien_Forschung/Studie_Microtargeting_DeutschlandEuropawahl2019_Hegelich_web2.pdf

Holzer, S., & Sengl, M. (2020, September). Quelle gut, alles gut? Glaubwürdigkeitsbeurteilung im digitalen Raum. In *Fake News und Desinformation* (pp. 155–178). Nomos Verlagsgesellschaft mbH & Co. KG.

Horvitz, E. (2022, November). On the horizon: Interactive and compositional deepfakes. In *Proceedings of the 2022 International Conference on Multimodal Interaction* (pp. 653–661).

Hunger, M. (2021). "Exploring the Pandora Papers with Neo4j." Verfügbar unter https://neo4j.com/developer-blog/exploring-the-pandora-papers-with-neo4j/

Kimmel, B. (2021). „Mit klicksafe Fake News und Verschwörungstheorien erkennen." Seiten 84 und 85 in: „Fakt oder Fake? Jugendschutz, Medienkompetenz und Desinformation." Verfügbar unter https://www.die-medienanstalten.de/publikationen/jugendschutz-medienkompetenzbericht/fakt-oder-fake-jugendschutz-medienkompetenz-und-desinformation

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat?. *Business Horizons,* 63(2), 135–146.

Kultusministerkonferenz (KMK), (2017). „Digitalstrategie für das Bildungswesen." Verfügbar unter https://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2018/Digitalstrategie_2017_mit_Weiterbildung.pdf

Konstan, J. A., & Riedl, J. (2012). "Recommender Systems: From Algorithms to User Experience." *User Modeling and User-Adapted Interaction*, 22, 101–123.

Kreye, T. (2021). „Die rote Linie." Verfügbar unter https://www.sueddeutsche.de/medien/kuenstliche-intelligenz-fake-news-recherche-1.5204699

Landesanstalt für Medien NRW, (2021). M*it künstlicher Intelligenz zu einer modernen Medienaufsicht*. Verfügbar unter https://www.medienanstalt-nrw.de/zum-nachlesen/recht-und-aufsicht/mit-kuenstlicher-intelligenz-zu-einer-modernen-medienaufsicht.html

Landesanstalt für Medien NRW, (2023). *ZEBRA gibt Tipps zum Erkennen und Eindämmen von Desinformation.* Verfügbar unter https://www.medienanstalt-nrw.de/presse/pressemitteilungen/pressemitteilungen-2023/default-19e9b2687c/zebra-gibt-tipps-zum-erkennen-und-eindaemmen-von-desinformation.html

Landesmedienanstalt Saarland, (2018). D*ie goldenen Medienregeln für Kinder und Eltern*. Verfügbar unter https://www.lmsaar.de/medienkompetenz/projekte/die-goldenen-medienregeln-fuer-kinder-und-eltern/

Langer, M. (2019). *Google beschränkt Microtargeting für politische Werbung*. Verfügbar unter https://www.nzz.ch/international/google-beschraenkt-microtargeting-fuer-politische-werbung-ld.1523660

Lardinois, F. (2023). *Microsoft brings the new AI-powered Bing to mobile and Skype, gives it a voice.* Verfügbar unter https://techcrunch.com/2023/02/22/microsoft-brings-the-new-ai-powered-bing-to-mobile-and-skype/

Lee, N., Li, B. Z., Wang, S., Yih, W. T., Ma, H., & Khabsa, M. (2020). "Language models as fact checkers?" *arXiv preprint arXiv:2006.04102.*

Leica, (2023). *Partnership for greater trust in digital photography: Leica and Content Authenticity Initiative.* Verfügbar unter https://leica-camera.com/en-int/news/partnership-greater-trust-digital-photography-leica-and-content-authenticity-initiative#

Li, M., Peng, B., & Zhang, Z. (2023). "Self-Checker: Plug-and-Play Modules for Fact-Checking with Large Language Models." *arXiv preprint arXiv:2305.14623.*

Liberini, F., Redoano, M., Russo, A., Cuevas, Á., & Cuevas, R. (2020). "Politics in the Facebook Era: Evidence from the 2016 US Presidential Elections." *CESifo Working Paper No. 8235*. Verfügbar unter http://dx.doi.org/10.2139/ssrn.3584086

Lima-Strong, R. (2021). "Facebook knew ads microtargeting could be exploited by politicians. It accepted the risk." Verfügbar unter https://www.washingtonpost.com/politics/2021/10/26/facebook-knew-ads-microtargeting-could-be-exploited-by-politicians-it-accepted-risk/

Landesvereinigung Kulturelle Jugendbildung (LKJ), Baden-Württemberg e.V. (2024a). YourStory. Verfügbar unter https://www.lkjbw.de/schule-kultur-medien/yourstory/

Landesvereinigung Kulturelle Jugendbildung (LKJ), Baden-Württemberg e.V. (2024b). Echt Fake, ich schwör! Verfügbar unter https://www.lkjbw.de/schule-kultur-medien/echt-fake-ich-schwoer/

Löser, A., Tresp, V. et al. (2023): Große Sprachmodelle – Grundlagen, Potenziale und Herausforderungen für die Forschung. Whitepaper aus der Plattform Lernende Systeme, München. https://doi.org/10.48669/pls_2023-3

Long, D., & Magerko, B. (2020). „What is AI literacy? Competencies and design considerations." *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.*MA HSH 2021

Medienanstalt Hamburg/Schleswig-Holstein (MA HSH), (2021). „Bericht 2020." Verfügbar unter https://www.ma-hsh.de/files/infothek/publikationen/Jahresberichte/MA-HSH-Bericht-2020.pdf

Medienanstalt Berlin-Brandenburg (mabb), (2024). „Journalismus macht Schule." Verfügbar unter https://www.mabb.de/journalismusmachtschule.html

Marcus, G. (2020). The next decade in AI: four steps towards robust artificial intelligence. arXiv preprint arXiv:2002.06177.

Maroni, D., Köhler, J., Fisseler, J., & Becker, S. (2020). Die ARD Mining-Plattform.

Martini, F., Samula, P., Keller, T. R., & Klinger, U. (2021). "Bot, or not? Comparing three methods for detecting social bots in five political discourses." *Big Data & Society,* 8(2). https://doi.org/10.1177/20539517211033566

Mast, T., Kettemann, M. C., Schulz, W., & Dreyer, S. (2023). Zum DDG-Referentenentwurf des BMDV vom 1. August 2023: Kurzstellungnahme. Verfügbar unter https://doi.org/10.21241/ssoar.89372

Medienanstalt Rheinland-Pfalz, (2024). „Ethik macht Klick – Meinungsbildung in der digitalen Welt." Verfügbar unter https://www.klicksafe.de/materialien/ethik-macht-klick-meinungsbildung-in-der-digitalen-welt/

Mehta, I. (2023a). "Meta wants to use generative AI to create ads." Verfügbar unter https://techcrunch.com/2023/04/05/meta-wants-to-use-generative-ai-to-create-ads/

Mehta, I. (2023b). "Meta announces generative AI features for advertisers.""Verfügbar unter https://techcrunch.com/2023/05/11/meta-announces-generative-ai-features-for-advertisers/

Meta, (2024). „Verwendung von Informationen für personalisierte Werbung." Verfügbar unter https://de-de.facebook.com/business/help/167836590566506?id=288762101909005

Metzger, N. (2024). Wie mit Fake-Vorwürfen Stimmung gemacht wird. ZDF heute. Verfügbar unter: https://www.zdf.de/nachrichten/politik/deutschland/demonstration-manipulation-desinformation-bilder-rechtsextremismus-100.html

Meßmer, A.-K., Sängerlaub, A., & Schulz, L. (2021). „Quelle: Internet"? Digitale Nachrichten- und Informationskompetenzen der deutschen Bevölkerung im Test. *Stiftung Neue Verantwortung e. V., 2021.*

Microsoft, (2022). "Microsoft Research: Understanding the news ecosystem to improve the health of democracy." Verfügbar unter https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE50KOK

Möller, J., Hameleers, M., & Ferreau, F. (2020). „Gutachten zur Entwicklung der gesamtgesellschaftlichen Verantwortung der Medien." Verfügbar unter https://www.die-medienanstalten.de/fileadmin/user_upload/die_medienanstalten/Publikationen/Weitere_Veroeffentlichungen/GVK_Gutachten_final_WEB_bf.pdf

Moh, C. (2023). *AI Voice Cloning: What It Is and How It Works.* Verfügbar unter https://lovo.ai/post/ai-voice-cloning-what-it-is-and-how-it-works

Medienpädagogischer Forschungsverbund Südwest (MPFS), (2023). „JIM 2023: Jugend, Information, Medien – Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger in Deutschland." *Stuttgart.*

Müller-Brehm, J. (2021). „Künstliche Intelligenz in Redaktionen – Ist Journalismus berechenbar?". In: Landesanstalt für Medien NRW (Hrsg.): *tbd – Der Datenmonitor der Landesanstalt für Medien NRW, Ausgabe 3 August 2021.*

Müller, J., & Spielkamp, M. (2023). "AI Act: The deal, key safeguards, and dangerous loopholes." Verfügbar unter https://algorithmwatch.org/en/ai-act-deal-key-safeguards-and-dangerous-loopholes/

Murphy, H. & Criddle, C. (2023). "Google to deploy generative AI to create sophisticated ad campaigns." In: *Financial Times.* Verfügbar unter https://www.ft.com/content/36d09d32-8735-466a-97a6-868dfa34bdd5.

Muscionico, D. (2022). *So funktioniert der Krieg der Bilder – heute und schon viel früher.* Verfügbar unter https://www.tagblatt.ch/kultur/krieg-der-bilder-im-krieg-ist-die-wahrheit-das-erste-opfer-wieso-kriegsfotografie-ein-revival-erlebt-ld.2267674

Neubauer, J. (2014). „Copyrightverletzungen: YouTube setzt weiterhin auf Content-ID-System." Verfügbar unter https://de.ign.com/news/87816/copyrightverletzungen-youtube-setzt-weiterhin-auf-content-id-system

Noyb,(2023a). "NOYB files complaint against EU Commission over targeted Chat Control ads." Verfügbar unter https://noyb.eu/en/noyb-files-complaint-against-eu-commission-over-targeted-chat-control-ads

Noyb, (2023b). "GDPR complaint against Twitter over illegal micro-targeting Chat Control ads." Verfügbar unter https://noyb.eu/en/gdpr-complaint-against-x-twitter-over-illegal-micro-targeting-chat-control-ads

Noyb, (2023c). "Political Microtargeting: Facebook Election Promise Just You." Verfügbar unter https://noyb.eu/en/political-microtargeting-facebook-election-promise-just-you

Oeftering, T. (2024). *Medienkompetenz politisch denken! Chancen und Herausforderungen der Digitalisierung für die politische Bildung.* Verfügbar unter https://www.politische-medienkompetenz.de/medienkompetenz-politisch-denken/

Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: statistical mechanics and its applications*, 540, 123174.

Paolo, F., Kroodsma, D., Raynor, J., et al. (2024). Satellite mapping reveals extensive industrial activity at sea. *Nature*, 625, 85–91. https://doi.org/10.1038/s41586-023-06825-8

Papakyriakopoulos, O., Shahrezaye, M., Serrano, J. C. M., & Hegelich, S. (2019, April). Distorting political communication: The effect of hyperactive users in online social networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops* (INFOCOM WKSHPS) (pp. 157–164). IEEE.

Pawelec, M. (2022). Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *DISO*, 1, 19. https://doi.org/10.1007/s44206-022-00010-6

Pew Research Center, (2012). Internet Gains Most as Campaign News Source but Cable TV Still Leads. Verfügbar unter https://www.pewresearch.org/journalism/2012/10/25/social-media-doubles-remains-limited/

Pew Research Center, (2014, October). Political Polarization and Media Habits.

Pew Research Center, (2016, February). The 2016 Presidential Campaign – a News Event That's Hard to Miss.

Pew Research Center, (2018, September). News Use Across Social Media Platforms 2018.

Potthast, M., Köpsel, S., Stein, B., & Hagen, M. (2016). Clickbait detection. In *Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20–23, 2016. Proceedings 38* (pp. 810–817). Springer International Publishing.

Raabe, S. (2023). *Offen, verantwortungsvoll und transparent – Die Guidelines der dpa für Künstliche Intelligenz.* Verfügbar unter https://innovation.dpa.com/2023/04/03/kuenstliche-intelligenz-fuenf-guidelines-der-dpa/

Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE Access, 10,* 25494–25513.

Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017, September). Truth of varying shades: Analyzing language in fake news and political fact-checking. *Proceedings of the 2017 conference on empirical methods in natural language processing, 2931–2937.*

Redoano, M. (2019). *Politics in the Facebook era: Examining the effects of voter "micro-targeting" in the 2016 US presidential election.* Warwick University. Verfügbar unter: https://warwick.ac.uk/fac/soc/economics/staff/mredoanocoppede/advantage.pdf

Reul, A. (2022). *Indiana Jones 5 Trailer Unveils Harrison Ford Deaged and the Dial of Destiny.* Variety. Verfügbar unter: https://variety.com/2022/film/news/indiana-jones-5-trailer-harrison-ford-deaged-dial-of-destiny-1235410524/

Reuter, C., Hughes, A. L., & Kaufhold, M. A. (2018). Social media in crisis management: An evaluation and analysis of crisis informatics research. *International Journal of Human – Computer Interaction,* 34(4), 280–294.

Reuters Institute, (2023). Digital News Report 2023. Verfügbar unter https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023

Ropek, A. (2021). *Bank Robbers in the Middle East Reportedly Cloned Someone's Hand to Steal Millions.* Gizmodo. Verfügbar unter https://gizmodo.com/bank-robbers-in-the-middle-east-reportedly-cloned-someo-1847863805

Rowlatt, M. (2023). *AI could predict hurricane landfall sooner – report.* BBC News. Verfügbar unter https://www.bbc.com/news/science-environment-67383755

Ryan-Mosley, T. (2023). *How generative AI is boosting the spread of disinformation and propaganda.* MIT Technology Review. Verfügbar unter: https://www.technologyreview.com/2023/10/04/1080801/generative-ai-boosting-disinformation-and-propaganda-freedom-house/

Salesforce, (2023). *Introducing the ChatGPT App for Slack.* Verfügbar unter https://www.salesforce.com/news/stories/chatgpt-app-for-slack/

Satariano, A., & Mozur, P. (2023). *The People Onscreen Are Fake. The Disinformation Is Real.* The New York Times. Verfügbar unter https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html

Sato, M. (2023). LinkedIn is adding AI tools for generating profile copy and job descriptions. Verfügbar unter https://www.theverge.com/2023/3/15/23640947/linkedin-ai-profile-job-description-tools

Schoenert, U. (2012). *Verständnisvolle Geräte.* In: Zeit Online. 14. Februar 2012. Verfügbar unter https://www.zeit.de/zeit-wissen/2012/02/Spracherkennung

Serrano, J. C. M., Shahrezaye, M., Papakyriakopoulos, O., & Hegelich, S. (2019, July). The rise of Germany's AfD: A social media analysis. In *Proceedings of the 10th international conference on social media and society* (pp. 214–223).

Silverman, C. (2016). *This analysis shows how viral fake election news stories outperformed real news on Facebook.* Verfügbar unter https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

Silverman, C., Strapagiel, L., Shaban, H., Hall, E. & Singer-Vine, J. (2016). *Hyperpartisan Facebook pages are publishing false and misleading information at an alarming rate.* Verfügbar unter https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis

Simmons, L., Thaler, C. & Nguyen, C. (2022). *Betrüger gibt sich im Telefonat mit Giffey als Vitali Klitschko aus.* In: Zeit Online. 25. Juni 2022. Verfügbar unter https://www.zeit.de/politik/2022-06/franziska-giffey-telefonat-betrug-deep-fake

Simon, F. M., Becker, K. B., & Crum, C. (2023a). Policies in parallel? A comparative study of journalistic AI policies in 52 global news organisations.

Simon, F.M., Altay, S. & Mercier, H. (2023b). *Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown.* Verfügbar unter https://misinforeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/

Spotify, (2023). *AI Voice Translation Pilot.* Verfügbar unter https://newsroom.spotify.com/2023-09-25/ai-voice-translation-pilot-lex-fridman-dax-shepard-steven-bartlett/

Stapf, I. (2021, February). „Fake News" als eine (mögliche) Frage der Wahrheit? Medienethische Perspektiven auf Wahrheit im Kontext der Digitalisierung. In *Medien und Wahrheit* (pp. 95–120). Nomos Verlagsgesellschaft mbH & Co. KG.

Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK),(2023). *Large Language Models und ihre Potenziale im Bildungssystem*. Impulspapier der Ständigen Wissenschaftlichen Kommission (SWK) der Kultusministerkonferenz. Verfügbar unter http://dx.doi.org/10.25656/01:28303

Tagesschau,(2023). *X erlaubt wieder politische Werbung*. Verfügbar unter https://www.tagesschau.de/ausland/amerika/x-twitter-musk-100.html

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. *Digital journalism,* 6(2), 137–153.

Tappin, B. M., Wittenberg, C., Hewitt, L. B., Berinsky, A. J., & Rand, D. G. (2023). Quantifying the potential persuasive returns to political microtargeting. *Proceedings of the National Academy of Sciences,* 120(25), e2216261120.

TargetLeaks, (2024). Verfügbar unter https://targetleaks.de/

Techvangart, (2023). *The Rise of AI News Anchors*. Verfügbar unter https://techvangart.com/2023/08/17/the-rise-of-ai-news-anchors/

Thiel, T., & Rostalski, F. (2021). Künstliche Intelligenz als Herausforderung für demokratische Partizipation. In V*erantwortungsvoller Einsatz von KI? Mit menschlicher Kompetenz!* (pp. 56–63). Berlin: Berlin-Brandenburgische Akademie der Wissenschaften.

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, 33(3), 460–480.

Thwaites, K. (2023) *Carvana Thanks Customers with One-of-a-Kind Videos Detailing the Day They Met Their Car.* Verfügbar unter https://www.businesswire.com/news/home/20230509005451/en/CarvanaThanks-Customers-with-One-of-a-Kind-Videos-Detailing-the-Day-They-Met-Their-Car

Tushya, Chhabra, D., & Abraham, B. (2023). Social Networking or Social Isolation? A Systematic Review on Socio-Relational Outcomes for Members of Online Gaming Communities. Games and Culture, 15554120231201760.

UK Parliament, (2019). Disinformation and "fake news": Final Report. Verfügbar unter: https://committees.parliament.uk/committee/378/digital-culture-media-and-sport-committee/news/103668/fake-news-report-published-17-19/

Vodafone Stiftung Deutschland gGmbH, (2023). *#WeissteBescheid-Kampagne bringt Jung und Alt das Thema Medienkompetenz näher*. Verfügbar unter https://www.vodafone-stiftung.de/
Wardle, C. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making (2017).

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Vol. 27, pp. 1–107). Strasbourg: Council of Europe.

World Economy Forum (WEF), (2024). *Global Risks Report 2024*. Verfügbar unter https://www.weforum.org/publications/global-risks-report-2024/

Weikmann, T., & Lecheler, S. (2023). Cutting through the Hype: Understanding the Implications of Deepfakes for the Fact-Checking Actor-Network. *Digital Journalism,* 1–18.

Wilczek, B. & Haim, M. (2022). Wie kann Künstliche Intelligenz die Effizienz von Medienorganisationen steigern? *Medienwirtschaft,* 19(4), 44–50. Verfügbar unter https://dx.doi.org/10.15358/1613-0669-2022-4-44

Woodward, A. (2020). *"Fake news": A guide to Trump's favourite phrase – and the dangers it obscures*. Verfügbar unter https://www.independent.co.uk/news/world/americas/us-election/trump-fake-news-counter-history-b732873.html

Ye, J. (2023). *China says generative AI rules to apply only to products for the public.* Verfügbar unter https://www.reuters.com/technology/china-issues-temporary-rules-generative-ai-services-2023-07-13/

Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3836–3847).

Zimmermann, F., & Kohring, M. (2018). „Fake News"als aktuelle Desinformation. Systematische Bestimmung eines heterogenen Begriffs. *M&K Medien & Kommunikationswissenschaft,* 66(4), 526–541.

Zweig, K. A., Deussen, O., & Krafft, T. D. (2017). Algorithmen und Meinungsbildung: eine grundlegende Einführung. *Informatik-Spektrum,* 40, 318–326.