

Zeitschriftenartikel*Begutachtet***Begutachtet:**Prof. Dr. Ulrike Verch HAW Hamburg
Deutschland**Erhalten:** 06. November 2024**Akzeptiert:** 17. November 2024**Publiziert:** 04. Februar 2025**Copyright:**

© Felina Wellner.

*Dieses Werk steht unter der Lizenz
Creative Commons Namens-
nennung 4.0 International (CC BY 4.0).***Empfohlene Zitierung:**

WELLNER, Felina, 2025: Die Rolle der Tonalität von generativer KI bei der Verbreitung und Wahrnehmung von Informationen. In: *API Magazin* 6(1) [Online] Verfügbar unter: [DOI 10.15460/apimagazin.2025.6.1.219](https://doi.org/10.15460/apimagazin.2025.6.1.219)

Die Rolle der Tonalität von generativer KI bei der Verbreitung und Wahrnehmung von Informationen

Chancen und Herausforderungen automatischer Textgenerierung in der journalistischen Praxis

Felina Wellner¹ ¹ Hochschule für Angewandte Wissenschaften Hamburg, Deutschland

Studierende im 3. Semester des Masterstudiengangs Digitale Transformation der Informations- und Medienwirtschaft

* Korrespondenz: redaktion-api@haw-hamburg.de

Zusammenfassung

Der Beitrag untersucht die Bedeutung der Tonalität generativer Künstlicher Intelligenz (KI) für die Verbreitung und Wahrnehmung von Informationen sowie mögliche Strategien zur Minimierung von Halluzinationen und Desinformation. Der Schwerpunkt liegt darauf, wie emotionale und stilistische Aspekte die Rezeption beeinflussen und welche Ansätze zur Sicherstellung der Zuverlässigkeit generierter Inhalte beitragen können. Dabei wird die Funktionsweise von KI-Chatbots und die Wirkung von Techniken wie Reinforcement Learning, Fine-Tuning und Retrieval Augmented Generation (RAG) analysiert, um ihren potenziellen Einfluss auf Qualität und Konsistenz der Tonalität zu verstehen.

Schlagwörter: Generative KI, Halluzination, Desinformation, Tonalität, Chatbot, Automatisierter Journalismus

The Impact of Generative AI Tonality on the Spread and Perception of Information

Opportunities and challenges of automatic text generation in journalistic practice

Abstract

The article examines the significance of the tonality of generative artificial intelligence (AI) for the dissemination and perception of information as well as possible strategies for minimising hallucinations and disinformation. The focus is on how emotional and stylistic aspects influence reception and which approaches can contribute to ensuring the reliability of generated content. It will analyse the functioning of AI chatbots and the impact of techniques such as reinforcement learning, fine-tuning and retrieval augmented generation (RAG) to understand their potential influence on the quality and consistency of tonality.

Keywords: Generative AI, Hallucination, Disinformation, Tonality, Chatbot, Automated Journalism

1 Einleitung

Welche Rolle spielt die Tonalität bei der Verbreitung und Wahrnehmung von Informationen? Der Begriff „generative KI“ bezeichnet „generative, vortrainierte Sprachmodelle“, die in der Lage sind, plausible Antworten für eine Vielzahl von Aufgaben zu generieren ([Lernende Systeme 2023](#), S. 12). Die Technologie umfasst Methoden zur Erzeugung neuer Daten in Form von Text, Bild, Audio oder Video. Im Rahmen dieser Arbeit wird die Tonalität generativer KI als die Art und Weise definiert, wie KI-Systeme Informationen präsentieren und kommunizieren, einschließlich des emotionalen und stilistischen Charakters der Textgenerierung ([Zumbühl 2024](#)).

Nach Zumbühl ([2024](#)) stehen folgende Eigenschaften im Fokus:

- Emotionale Färbung: Die Fähigkeit, Emotionen zu erkennen und zu nutzen, um die Intention der sprechenden Person zu vermitteln. Als Beispiel sei hier die Affektivität, die Stimmung sowie die Empathie genannt, welche in Rückkopplung auf die Eingaben der Nutzenden stehen.
- Kontextabhängigkeit: Anpassung der Sprache an den spezifischen Kontext und die Zielgruppe, unter Berücksichtigung von Wortwahl, Struktur und Komplexität.
- Stilistische Konsistenz: Beibehaltung eines einheitlichen Stils über verschiedene Kommunikationsformen hinweg, beispielsweise im Tonfall.

Die Relevanz der Tonalität von generativer KI liegt in ihrer Fähigkeit, die Wahrnehmung und das Vertrauen der Nutzenden in die bereitgestellten Informationen zu beeinflussen. Die vorliegende Arbeit widmet sich der zentralen Fragestellung, in welcher Weise die Tonalität von generativer KI die Verbreitung und Wahrnehmung von Informationen beeinflusst und welche Maßnahmen Unternehmen ergreifen können, um Halluzinationen und Desinformationen zu minimieren.

Die folgenden Kapitel analysieren die Integration von Automatisierungstechnologien in die Medienbranche und die Funktionsweise und Tonalität generativer KI anhand von GPT-basierten Systemen. Zudem werden die Gefahren von Halluzinationen und Desinformation sowie Präventionsstrategien thematisiert. Ein Experteninterview gibt Einblicke in die praktische Anwendung und mögliche Optimierungsmethoden. Abschließend werden die Chancen und Grenzen von generativer KI als neuer Bestandteil der Informationslandschaft diskutiert, um ein umfassendes Bild der aktuellen und zukünftigen Entwicklungen zu zeichnen.

2 Automatisierung und generative KI in der modernen Informationslandschaft

Zunächst soll der aktuelle Stand der Integration von Automatisierungstechnologien und generativer KI in die Informationslandschaft analysiert werden. Es wird aufgezeigt, wie weitreichend diese bereits zur Informationsvermittlung eingesetzt werden, um in den folgenden Kapiteln zu diskutieren, welche Rolle die Tonalität bei der Gestaltung und Präsentation von Inhalten spielt. Ein weiterer Fokus liegt auf der Abgrenzung von Automatisierung und dem Einsatz von (generativer) KI, insbesondere am Beispiel des Journalismus.

2.1 Medienwandel durch Automatisierung

Die Medienlandschaft befindet sich in einem anhaltenden Veränderungsprozess. Ein zentraler Aspekt dieser Transformation ist die Auflösung des traditionellen Gatekeeper-Monopols journalistischer Akteur*innen ([Neuberger 2022](#), S. 163). Kontrollierten in der Vergangenheit wenige große Medienhäuser den Informationsfluss, haben alternative Angebote wie Soziale Plattformen als „vermittelnde Dritte“ diesen Einfluss bereits reduziert. Auf Sozialen Plattformen entsteht ein sogenannter „Netzwerkjournalismus“ als „Knoten in einem weit verzweigten, dezentralen und dynamischen öffentlichen Netzwerk, in dem er im interaktiven Austausch mit Quellen und Publikum steht“ ([Neuberger 2022](#), S. 172). Es ist davon auszugehen, dass sich dieser Trend mit der Einführung von Automatisierungstechnologien, insbesondere der generativen KI, weiter verstärken wird. Denn neue Technologien wie ChatGPT (siehe Kapitel 3) erweitern die Möglichkeiten der Informationsabfrage und -produktion erheblich, was zu einer Vielzahl neuer Informationsangebote führt und gleichzeitig zu einer zunehmenden Unübersichtlichkeit und Fragmentierung der Gesellschaft beiträgt. Diese Entwicklungen markieren einen bedeutenden Medienwandel, der durch nachhaltige Veränderungen der technischen Kommunikationskanäle, der Organisationen und der gesellschaftlichen Institutionen, die die Kommunikation in der Gesellschaft beeinflussen, gekennzeichnet ist ([Körner 2023](#), S. 20).

Der von Körner ([2023](#), S. 20) beschriebene „Medienwandel der Automatisierung“ bezeichnet die Gesamtheit der Automatisierungstechnologien, die in die moderne Medienlandschaft eingeflossen sind und zu tiefgreifenden Veränderungen geführt haben. Diese reichen von verbesserter Barrierefreiheit über neue Möglichkeiten der Moderation von Kommentaren in Sozialen Medien bis hin zu neueren Entwicklungen wie der automatisierten Erstellung von Texten und Bildern ([Lernende Systeme 2023](#), S. 10). Die Automatisierung durch KI unterscheidet sich von herkömmlichen Automatisierungstechnologien durch ihre Fähigkeit, kreative Aufgaben zu übernehmen. So können KI-Systeme qualitativ hochwertige Texte generieren, die eine hohe Ähnlichkeit zu menschlich verfassten Texten aufweisen: In einer Umfrage auf dem Nachrichtenportal „kress“ ist nur einer Minderheit der Lesenden (39 Prozent) aufgefallen, dass der Text von einer KI und nicht von einem Menschen erstellt wurde ([Bartl 2019](#)). KI-Systeme können die menschliche Kreativität und Produktionsmöglichkeiten er-

weitern und die Erkennung komplexer Datenmuster und Ableitung neuer Handlungsoptionen ermöglichen ([Deutscher Ethikrat 2023](#), S. 177ff.). Allerdings entstehen auch neue Abhängigkeiten und Anpassungszwänge: Softwaresysteme können das Verhalten der Nutzenden steuern, indem sie die politische Kommunikation strukturieren und Suchergebnisse filtern, was zu einer schleichenden Vereinheitlichung führen kann ([Deutscher Ethikrat 2023](#), S. 179). Die Anpassung an technische Vorgaben kann unbewusst erfolgen und langfristige Auswirkungen auf die Meinungsbildung haben. Zudem besteht das Risiko, dass KI die Ungleichheit zwischen Nachrichtenorganisationen verschärfen könnte ([Beckett 2019](#), S. 55f.). Größere Organisationen verfügen über mehr Ressourcen, um in KI-Technologien zu investieren und diese effektiv einzusetzen, wodurch kleinere Akteure einen Wettbewerbsnachteil haben. Dies könnte sich negativ auf die Medienvielfalt und die Zukunftsfähigkeit des Journalismus auswirken.

2.2 Einbindung in die journalistische Praxis

Die Integration moderner Technologien in alle Bereiche der journalistischen Praxis lässt sich unter dem Begriff „Computational Journalism“ zusammenfassen ([Haim und Graefe 2018](#), S. 140; [Körner 2023](#), S. 16). Diese Prozesse, die in den 1960er Jahren in der investigativen Recherche begannen, umfassen heute zahlreiche Anwendungen von der Themenfindung über die Produktion bis zur Distribution ([Beckett 2019](#), S. 20; [Körner 2023](#), S. 21). Der Teilbereich der Textgenerierung, bekannt als „automatisierter Journalismus“, hat in den vergangenen Jahren erhebliche Fortschritte gemacht ([Haim und Graefe 2018](#), S. 141; [Körner 2023](#), S. 22). „In einigen deutschen Nachrichtenredaktionen gehört der Einsatz von automatisierten Prozessen zur Textgenerierung längst zum Standard“ ([Lernende Systeme 2023](#), S. 11). „Automatisierter Journalismus“ soll im Mittelpunkt dieses Kapitels stehen. Häufige Anwendungsbereiche sind Sport-, Finanz-, Wetter- und Verkehrsnachrichten, da sich diese durch eine hohe Datenverfügbarkeit und standardisierte Berichterstattungsmuster auszeichnen ([Beckett 2019](#), S. 25; [Körner 2023](#), S. 44). Ein weiterer Anwendungsbereich ist die automatisierte Erstellung und Veröffentlichung von Beiträgen in Sozialen Medien durch Nachrichtenmedien, wie aus einer globalen Umfrage unter Journalist*innen hervorgeht ([Beckett 2019](#), S. 24).

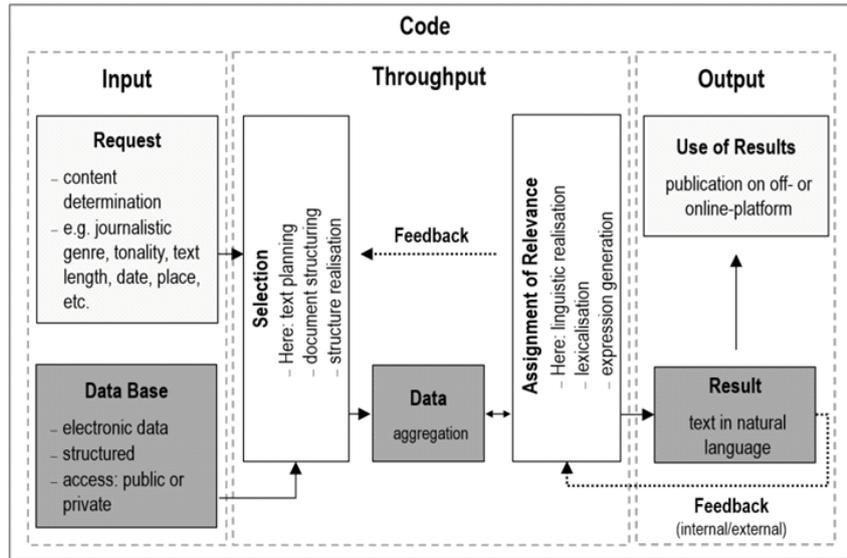


Abb. 1: Eingabe-Ausgabe-Verarbeitungsmodell des „automatisierten Journalismus“ (Quelle: [Dörr 2015](#))

Der „automatisierte Journalismus“ folgt dem „Eingabe-Verarbeitungs-Ausgabemodell“ (Abb. 1), das die Erhebung von Daten, die Verarbeitung durch Algorithmen und die Textausgabe umfasst ([Körner 2023](#), S. 26). Nach Haim und Graefe ([2018](#), S. 141) erfordert dieser Prozess hinreichend strukturierte Daten als Input, wobei die anschließende Verarbeitung ohne menschliches Zutun erfolgt, was oft als „Black Box“ bezeichnet wird. Die Grenzen des Modells liegen in der notwendigen journalistischen Mitwirkung bei der Datenverarbeitung und dem Training der Algorithmen ([Körner 2023](#), S. 26). Darüber hinaus lassen sich hybride Texte von rein automatisiert erstellten Texten unterscheiden. Die hybride Form kombiniert menschliche Fähigkeiten zur Kontextualisierung und Qualitätskontrolle mit algorithmischem Output ([Lernende Systeme 2023](#), S. 13; [Körner 2023](#), S. 19). Ein Beispiel hierfür ist „Reporters and data and robots“ (RADAR) der Press Association. Das Angebot kombiniert die Fähigkeiten von Menschen und Maschinen, um große Datenmengen zu analysieren und daraus lokalisierte Nachrichten für bestimmte Zielgruppen zu generieren ([Beckett 2019](#), S. 25). Körner ([2023](#), S. 20) beschreibt die aktuellen Entwicklungen im „automatisierten Journalismus“ als Medieninnovation im übergeordneten Medienwandel der Automatisierung ([Körner 2023](#), S. 45) (siehe Kapitel 3.1). Kriterien für eine Medieninnovation sind: Neuheit, Verwertbarkeit im wirtschaftlichen Kontext und kommunikative Folgen ([Körner 2023](#), S. 20).

Das Kriterium der Neuheit kann durch eine veränderte Form der Nachrichtenproduktion belegt werden, die die traditionelle journalistische Praxis ergänzt ([Körner 2023](#), S. 20). „Automatisierter Journalismus“ nutzt zwei Haupttechnologien: Lückentextmodelle und natürlichsprachliche Textgenerierung (Natural Language Generation, NLG) ([Körner 2023](#), S. 26). Lückentextmodelle verwenden vorgefertigte Textvorlagen (Templates), in die spezifische Informationen aus strukturierten Datenquellen

eingefügt werden ([Körner 2023](#), S. 27). NLG-Systeme hingegen generieren Texte vollständig neu, basierend auf maschinellem Lernen und neuronalen Netzen, um kontextbezogene Texte in natürlicher Sprache zu erstellen ([Körner 2023](#), S. 28). Auf die Funktionsweise der generativen KI wird an späterer Stelle (siehe Kapitel 3.1) näher eingegangen. An dieser Stelle lässt sich bereits der Unterschied zwischen automatisiertem und KI-Journalismus festhalten: Im Vergleich zu „einfachen“ Automatisierungstechnologien können mit Hilfe von NLG-Systemen und generativer KI nicht nur bestehende Informationen analysiert, sondern auf Basis dieser Analysen auch neue Inhalte erstellt werden. Dies führt zu einer fundierteren und umfassenderen Berichterstattung. Damit geht KI-Journalismus über die reine Automatisierung hinaus und ermöglicht die Erfüllung journalistischer Aufgaben, die traditionell menschliches Urteilsvermögen und kreative Fähigkeiten erfordern.

Die wirtschaftliche Verwertung des automatisierten Journalismus zeigt sich in der Entwicklung des ökonomischen Marktes, der sich in vier Phasen unterteilen lässt: Beginn der Anwendungen in den USA (2010-2012), frühe Marktphase, in der auch in Europa und später weltweit mit der Textgenerierung experimentiert wurde (2012-2016), Institutionalisierung mit einer stetigen Ausweitung der Themen und Anwendungen (ab 2016) und die flächendeckende Verfügbarkeit moderner „Generativer vortrainierter Transformer“ (GPT, Generative Pre-trained Transformer) (ab 2022) ([Körner 2023](#), S. 30ff.). Das Potenzial für wirtschaftlichen Erfolg beweist beispielsweise die US-amerikanische Tageszeitung „Washington Post“, die mit ihrer KI-Technologie „Heliograf“ durchschnittlich 850 zusätzliche Geschichten veröffentlicht, was zu einer Steigerung des Website-Traffics um mehr als 500.000 Klicks führte ([Lovrinovic und Volland 2019](#), S. 14).

Mit Blick auf den dritten Aspekt zur Beschreibung einer Medieninnovation lässt sich festhalten, dass die generierten Nachrichten auch kommunikative Folgen haben ([Körner 2023](#), S. 21) (siehe Kapitel 3.1). „Automatisierter Journalismus“ ermöglicht unter anderem die Berichterstattung über Nischenthemen, die bisher ökonomisch nicht rentabel waren. Darüber hinaus könnte die zunehmende Personalisierung von Nachrichten mittel- und langfristige Auswirkungen auf gesellschaftliche Kommunikationsprozesse haben. Algorithmische Personalisierungstendenzen verstärken die menschliche Neigung, Inhalte zu bevorzugen, die den eigenen Überzeugungen und Werten entsprechen, den sogenannten „Confirmation Bias“ ([Beckett 2019](#), S. 57).

3 Tonalität und Mechanismen generativer KI am Beispiel von GPT-basierten Systemen

Zur Untersuchung von Halluzinationen und der Rolle der Tonalität in generativer KI ist es essenziell, zunächst die Funktionsweise von KI-Chatbots zu analysieren. Diese Analyse bildet die Grundlage, um die Mechanismen zu verstehen, die zu Halluzinationen und deren Verbreitung führen sowie die Bedeutung der Tonalität in der Kommunikation mit KI-Systemen zu erörtern.

3.1 Funktionsweise von Chatbots auf Basis generativer KI

Die Entwicklung klassischer Chatbots verlief in mehreren Phasen. Frühere Modelle wie „ELIZA“ (1966) und „A.L.I.C.E.“ (1995) basierten auf festen Regeln und Entscheidungsbäumen, was zu begrenzten Fähigkeiten in der natürlichen Sprachverarbeitung führte ([Schöbel et al. 2024](#), S. 738). Die nächste Generation, die so genannten „Conversational Agents“, kombinierte maschinelles Lernen mit regelbasierten Systemen und ermöglichte fortschrittlichere Interaktionen ([Schöbel et al. 2024](#), S. 737f.). Diese konnten gesprochene Sprache verstehen und einfache Befehle ausführen, waren aber nur begrenzt lernfähig. Mit der Weiterentwicklung der Technologie und der zunehmenden Leistungsfähigkeit von KI-Systemen hat sich die Funktionsweise von Chatbots grundlegend verändert ([Schöbel et al. 2024](#), S. 739; [Körner 2024](#), S. 258). Die jüngste Entwicklung basiert auf Large Language Models (LLMs) wie GPT-3 und GPT-4 von OpenAI. Diese nutzen NLG zur Erzeugung menschenähnlicher und kontextbezogener Texte, was die Kommunikation in natürlicher Sprache ermöglicht (siehe Kapitel 3.2).

LLMs revolutionieren die Verarbeitung natürlicher Sprache (Natural Language Processing, NLP), indem sie Maschinen befähigen, menschliche Sprache besser zu verstehen und zu erzeugen ([Paaß und Giesselbach 2023](#), S. 1). „Aktuelle Systeme sind imstande, unterschiedliche Informationen aus den Daten zu extrahieren, diese zu bewerten, zu priorisieren und daraus schließlich ein adäquates Narrativ zu konstruieren“ ([Haim und Graefe 2018](#), S. 150). Ein wichtiger Schritt in diesem Prozess ist die Tokenisierung, die Texte in kleinere Einheiten (Wörter, Wortteile oder einzelne Zeichen) zerlegt ([Paaß und Giesselbach 2023](#), S. 4f.). Diese Tokens werden in hochdimensionale Vektorräume eingebettet, wobei ihre jeweiligen Positionen die Bedeutungen und Beziehungen zueinander repräsentieren. Wörter, die häufig in ähnlichen Kontexten vorkommen, haben ähnliche Positionen, wodurch die Bedeutungen im Kontext leichter erkannt werden können ([Paaß und Giesselbach 2023](#), S. 39).

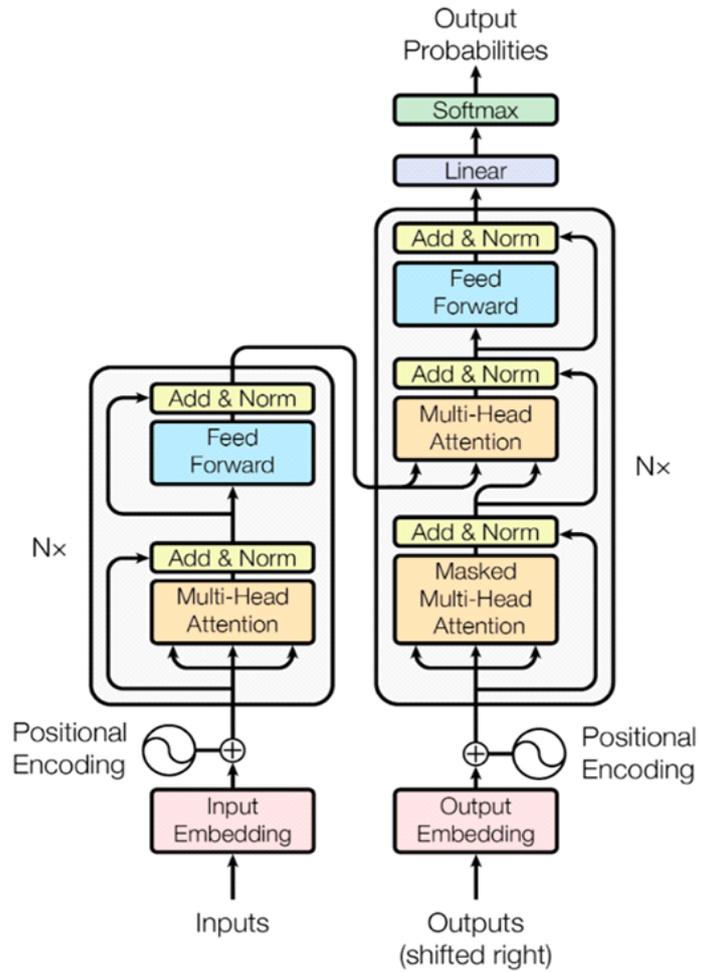


Abb. 2: Transformer-Architektur, links: Encoder und rechts: Decoder (Quelle: [Vaswani et al. 2017](#), S. 3)

Die Tokenisierung ist in eine komplexe Transformer-Architektur eingebettet, die eine „Encoder-Decoder-Struktur“ (Abb. 2) verwendet ([Vaswani et al. 2017](#), S. 1). Der Encoder verarbeitet die Eingabesequenz und fügt sogenannte „Positional Encodings“ hinzu, die Informationen über die Position der Tokens in der Sequenz enthalten ([Vaswani et al. 2017](#), S. 5f.). Auf diese Weise kann die Reihenfolge von Tokens und Wörtern erfasst werden. „Self-Attention“ ermöglicht es dem Modell, jedes Wort oder Token mit allen anderen in der Sequenz in Beziehung zu setzen, wodurch Abhängigkeiten und Beziehungen zwischen den Wörtern erkannt werden ([Vaswani et al. 2017](#), S. 2). Von Bedeutung ist dabei der sogenannte „Multi-Head Attention“-Mechanismus, der beschreibt, dass mehrere solcher „Self Attention“-Mechanismen parallel ausgeführt werden, um Informationen aus verschiedenen Aspekten der Eingabe gleichzeitig zu verarbeiten und eine reichhaltigere Informationsverarbeitung zu ermöglichen ([Vaswani et al. 2017](#), S. 4f.). Der Decoder verwendet die vom Encoder erzeugten Repräsentationen, um das nächste Token auf Basis des gelernten Kontextes vorherzusagen ([Vaswani et al. 2017](#), S. 3). Dieser Prozess wird als „Feed Forward“ bezeichnet. Bei ausreichenden Kontextinformationen können so Synony-

me oder identische Wörter, die im Satzkontext jedoch eine andere Bedeutung haben, erkannt werden ([Paaß und Giesselbach 2023](#), S. 44).

Die beschriebene parallele Verarbeitung aller Tokens verbessert das Sprachverstehen, verkürzt die Trainingszeit und erhöht die Leistungsfähigkeit der Modelle, insbesondere im Vergleich zu älteren Modellen, die Tokens nacheinander verarbeiten ([Vaswani et al. 2017](#), S. 2). Allerdings gilt auch hier: Die Genauigkeit der Ergebnisse hängt stark von den Trainingsdaten ab, die daher immer überprüft werden sollten. Beispielsweise weiß das Modell nur dann, dass ein Apfel an einem Baum wächst, wenn diese Information häufig in seinen Trainingsdaten vorkommt und als richtig erkannt wurde.

Der Trainingsprozess einer Transformer-Architektur erfolgt in zwei Schritten ([Brown et al. 2020](#), S. 3). Der erste Schritt ist das Pre-Training mit großen Textmengen. GPT-3 basiert auf Datenquellen wie Common Crawl (60 %), WebText (22 %), Büchern (16 %) und Wikipedia (3 %) ([Brown et al. 2020](#), S. 9). In einem zweiten Schritt können weitere Optimierungsstrategien angewendet werden, um das LLM zu verbessern und auf einen Chatbot zu überführen. Beim „Supervised Fine-Tuning“ wird das Modell auf Grundlage eines überwachten Datensatzes trainiert und an spezifische Anwendungsfälle angepasst ([Brown et al. 2020](#), S. 6). Dieser traditionelle und immer noch weit verbreitete Ansatz aktualisiert die Parameter (Gewichte) des Modells für die gewünschte Aufgabe ([Brown et al. 2020](#), S. 6). Dieses Verfahren erfordert große Mengen an qualitativ hochwertigen Daten, typischerweise werden tausende bis hunderttausende von annotierten Beispielen verwendet. Beim „Reinforcement Learning from Human Feedback“ (RLHF) wird das Modell durch menschliches Feedback optimiert sowie daran gekoppelt Belohnungen und Bestrafungen ([Paaß und Giesselbach 2023](#), S. 143). OpenAI verwendet ein trainiertes Belohnungsmodell, das auf menschlichen Präferenzen aus realen Gesprächsdaten basiert ([Gao et al. 2022](#), S. 2). Der Trainingsablauf umfasst eine Aufgabenstellung, einen Antwortversuch durch den Chatbot, eine Bewertung der Antwortqualität, Anpassungen basierend auf der Bewertung und iterative Wiederholungen mit dem verbesserten Modell. „RLHF“ wird häufig nach dem „Supervised Fine-Tuning“ eingesetzt, um die Modelle weiter zu verbessern und menschliche Präferenzen besser zu berücksichtigen.

Eine alternative Methode zum Fine-Tuning ist das Prompt-Engineering, wodurch Modelle mit minimalem Training zur Ausführung von Aufgaben befähigt werden sollen ([Brown et al. 2020](#), S. 6f.). Dazu zählen Techniken wie „Zero-Shot-, One-Shot- und Few-Shot-Learning“. „Zero-Shot-Learning“ bezeichnet die Fähigkeit eines Modells, Aufgaben zu lösen, ohne zuvor explizit für diese spezifischen Aufgaben trainiert worden zu sein. Das Modell greift auf allgemeines Wissen zurück, um neue Anfragen auf Grundlage eines umfassenden Pre-Trainings zu bearbeiten ([Brown et al. 2020](#), S. 7). „One-Shot-Learning“ erweitert diesen Ansatz, indem das Modell mit einem Beispiel trainiert wird, um eine neue Aufgabe zu erlernen. Dies erfordert eine

hohe Generalisierungsfähigkeit des Modells, um von einem minimalen Trainingsatz auf die Lösung der Aufgabe zu schließen ([Brown et al. 2020](#), S. 6f.). „Few-Shot-Learning“ geht noch einen Schritt weiter und bezieht sich auf das Training eines Modells mit wenigen Beispielen, typischerweise zwei bis fünf ([Brown et al. 2020](#), S. 6).

LLMs wie GPT-3 und GPT-4, die als allgemeine Sprachmodelle entwickelt wurden, werden zu Chatbots, wenn sie in Anwendungen wie ChatGPT integriert werden, das 2022 von OpenAI eingeführt wurde ([OpenAI 2022](#)). ChatGPT nutzt die Fähigkeiten von GPT-Modellen und kann durch zusätzliche Funktionen wie Websuchen oder hochgeladene Dokumente ein breites Spektrum von Anfragen präzise beantworten und dabei eine natürliche und menschliche Tonalität beibehalten. Seit November 2023 sind zur individuellen Anpassung sogenannte CustomGPTs verfügbar, mit denen Nutzende eigene Versionen von ChatGPT erstellen können, die auf spezifische Bedürfnisse zugeschnitten sind ([OpenAI 2023](#)).

3.2 Generierung von Tonalität

Die Tonalität generativer KI wird in dieser Arbeit als die Art und Weise definiert, wie KI-Systeme Informationen präsentieren und kommunizieren, wobei der emotionale und stilistische Charakter der Textgenerierung eine zentrale Rolle spielt (siehe Kapitel 1). Am Beispiel von ChatGPT wird untersucht, wie Chatbots diese Tonalität entwickeln und anwenden. Schließlich ist das Ziel des langwierigen Trainingsprozesses, dass die produzierten Texte sowohl sinnvoll zusammenhängend als auch im gewünschten Stil und Sprachgebrauch ausgegeben werden ([Lernende Systeme 2023](#), S. 11).

Die Tonalität von ChatGPT ist entscheidend für die User Experience und beeinflusst die Wahrnehmung und Bewertung der Informationen. Der Chatbot kann verschiedene Tonalitäten annehmen, um empathischer, sachlicher oder informeller zu wirken ([Zumbühl 2024](#)). Diese Flexibilität verbessert die Interaktion und erhöht die Akzeptanz der Informationen ([OpenAI 2024](#)). Die Tonalität ergibt sich aus dem komplexen Trainingsprozess und der zugrundeliegenden Datenbasis (siehe Kapitel 3.1). ChatGPT lernt durch die Analyse der bereitgestellten Daten, wie verschiedene Tonalitäten in unterschiedlichen Kontexten verwendet werden, indem es Muster in der Sprache erkennt und reproduziert ([Zumbühl 2024](#)).

Im Pre-Training erlernt der Chatbot allgemeine Sprachstrukturen, Grammatik und grundlegende kontextuelle Zusammenhänge (siehe Kapitel 3.1). Die Datenbasis umfasst ein breites Spektrum an Stilen (stilistische Konsistenz; siehe Kapitel 1) und Emotionen (emotionale Einfärbung; siehe Kapitel 1), was es ChatGPT ermöglicht, eine allgemein verstandene menschliche Tonalität anzunehmen. Im „Fine Tuning“ (siehe 3.1) werden die Datensätze auf spezifische Stile, Fachgebiete oder Anwendungsfälle abgestimmt, wodurch die Sprache und Tonalität von ChatGPT an ausge-

wählte Kontexte angepasst werden können (Kontextabhängigkeit; siehe Kapitel 1). Darüber hinaus kann das Modell auch feinere Nuancen der Sprache erlernen. Zumbühl (2024) beschreibt den Prozess des „Reinforcement Learning with Human Feedback“ im Kontext der Tonalität als „adaptives Lernen“, bei dem Nutzende Feedback zu den generierten Antworten geben, wodurch das Modell seine Parameter anpasst und kontinuierlich verbessert. Zusätzlich können spezifische Prompts verwendet werden, um ChatGPT spezifische Anweisungen zu geben (Zumbühl 2024). So kann ChatGPT anhand spezifischer Daten oder Beispiele trainiert werden (Zero-Shot-, One-Shot- und Few-Shot-Learning; siehe 3.1), um den „richtigen Ton“ zu treffen und beispielsweise eine formale oder freundliche Tonalität zu verwenden (OpenAI 2024; Zumbühl 2024).

Die Steuerung der Tonalität durch generative KI birgt das Potenzial, den Stil, den Kontext und die emotionale Einfärbung des Outputs an die menschliche Sprache und die spezifischen Aufgabenbereiche anzupassen. Herausforderungen bestehen jedoch in der Erfassung subtiler menschlicher Tonalitäten und kontextueller Besonderheiten, die Chatbots wie ChatGPT oft nicht vollständig verstehen (Zumbühl 2024). Ironie und Sarkasmus können beispielsweise besonders leicht missverstanden werden und zu Kommunikationsfehlern führen. Darüber hinaus variiert die Tonalität in verschiedenen Kulturen stark, was ebenfalls zu Missverständnissen führen kann, wenn kulturelle Nuancen nicht erkannt und angepasst werden (Zumbühl 2024). Außerdem können ChatGPT und Co Schwierigkeiten haben, Emotionalitäten hinter menschlichen Ausdrücken richtig zu interpretieren (Zumbühl 2024). Eine ungenaue Emotionserkennung kann zu unangemessenen Antworten und einer unpassenden Tonalität führen. Eine weitere Gefahr besteht in der Übernahme von Vorurteilen (Bias) aus den Trainingsdaten, wenn diese kulturelle, soziale oder geschlechtsspezifische Stereotype enthalten (Beckett 2019, S. 57f.; Brown et al. 2020, S. 36ff.). Diese Verzerrungen können die Art und Weise beeinflussen, wie das Modell antwortet und bestimmte Tonalitäten bevorzugt oder unangemessen anwendet. Brown et al. (2020, S. 36ff.) haben festgestellt, dass GPT-3-Modelle Geschlechter- und Herkunftsbias aufweisen, indem sie Berufe häufiger mit männlichen Identifikatoren und bestimmte Ethnien oder Religionen negativ konnotiert darstellen. Es ist daher wichtig, Vorurteile zu erkennen und zu minimieren, um eine faire und konsistente Tonalität zu gewährleisten (Beckett 2019, S. 58).

4 Gefahr von Halluzinationen und Desinformationen durch generative KI

Neben der Tonalität ist vor allem die Qualität der Informationen ein entscheidender Faktor für das Nutzungserlebnis und die gesellschaftliche Informiertheit. Generative KI kann jedoch die Verbreitung von Desinformationen begünstigen (siehe Kapitel 2). In diesem Kapitel wird zwischen unbeabsichtigten Fehlinformationen durch Halluzinationen und absichtlichen Falschinformationen (Fake News) unterschieden. Auch

Deepfakes, bei denen maschinelles Lernen eingesetzt wird, um realistisch aussehende, aber gefälschte Videos oder Audiodateien zu erstellen, werden betrachtet.

Die Verbreitung von Fehlinformationen stellt eine große Gefahr dar, da Chatbots nur zufällig richtige Antworten produzieren ([Bai et al., 2024](#), S. 26). Sie generieren Texte auf Basis von Wahrscheinlichkeiten und können statistisch plausible, aber faktisch falsche Informationen erzeugen. Eine mögliche Ursache ist unzureichendes Training trotz erfolgreicher Tonalitätskontrolle (siehe Kapitel 3.2). Eine robuste Transformer-Architektur (siehe Kapitel 3.1), qualitativ hochwertige Daten und effiziente Trainingsmethoden sind essenziell für die Qualität der generierten Informationen ([Bai et al. 2024](#), S. 7ff.). Unvollständige, veraltete oder verzerrte Trainingsdaten können dazu führen, dass das Modell fehlerhafte Muster lernt und Halluzinationen erzeugt ([Bai et al. 2024](#), S. 7f.). Visuelle Informationen sind ebenfalls von zentraler Bedeutung, um die Kontextualität zu erweitern. Bild-Text-Paare von schlechter Qualität oder unzureichender Menge können falsche Verknüpfungen und fehleranfällige Vorhersagen verursachen, was die Gefahr von Halluzinationen ebenfalls erhöht. Außerdem können eine schwache visuelle Modellkomponente und die Dominanz sprachlicher Elemente dazu führen, dass visuelle Informationen falsch interpretiert oder ignoriert werden ([Bai et al. 2024](#), S. 8f.). Auch die Qualität der Trainingsmethoden zur Analyse komplexer räumlicher Strukturen von Bildern spielt eine wichtige Rolle ([Bai et al. 2024](#), S. 9).

Halluzinationen treten insbesondere bei komplexen oder spezifischen Anfragen auf, die präzise Detailinformationen erfordern ([Bai et al. 2024](#), S. 8). Techniken wie „Supervised Fine-Tuning“ und „RLHF“ können die Genauigkeit und Relevanz der Antworten verbessern (siehe Kapitel 3.1), jedoch auch unbeabsichtigte Voreingenommenheit einführen ([Bai et al. 2024](#), S. 2). Halluzinationen können schwerwiegende Folgen haben, da sie häufig als wahr angenommen werden.

Generative KI kann auch bewusst zur Erzeugung von Fake News genutzt werden. Anders als Menschen können diese Technologien in großen Mengen und in kurzer Zeit falsche Inhalte produzieren ([Deutscher Bundestag 2020](#), S. 435). Manipulierte Daten und Tonalitätssteuerung (siehe Kapitel 3) ermöglichen die Erstellung überzeugend wirkender, aber falscher Inhalte. Dies ist besonders im Kontext demokratischer Prozesse bedenklich: Die Enquete-Kommission Künstliche Intelligenz ([Deutscher Bundestag 2020](#), S. 435) nennt beispielhaft die Präsidentschaftswahlen in den USA im Jahr 2016, bei denen Facebook politische Anzeigen mit falschen Inhalten aus Kreml-nahen Quellen identifizieren musste.

Darüber hinaus schaffen generative KI-Modelle auch neue Möglichkeiten für die Erstellung und Verbreitung von Deep Fakes. Der Begriff setzt sich aus „Deep Learning“ und „Fake“ zusammen und beschreibt Audio- und Videoinhalte, die Handlungen oder Aussagen realer Personen darstellen, die diese nie gemacht haben ([Deut-](#)

[scher Bundestag 2020](#), S. 462). Der mehrstufige Trainingsprozess wird, ebenso wie bei der Generierung von Fake News, gezielt manipuliert. Dafür ist eine kritische Masse an Rohdaten wie Fotos und Videos erforderlich. „Maschinelles Lernen ist das perfekte Werkzeug für die Generierung von Deepfakes und gefälschten Inhalten. Dies wird ein großes Problem für Nachrichten und glaubwürdige Medien darstellen. Auch die Überprüfung von Inhalten wird viel schwieriger und/oder werkzeuggeständig werden“, so eine anonyme Antwort im Rahmen einer globalen Umfrage von Journalist*innen ([Beckett 2019](#), S. 57). Die zunehmende Verfügbarkeit und Qualität dieser Inhalte erschwert es, zwischen echten und gefälschten Inhalten zu unterscheiden, was zu Unsicherheit bei der Informationsbeschaffung führen kann. Darüber hinaus werfen Deepfakes auch rechtliche Fragen auf, da sie das Recht am eigenen Bild verletzen und gegen die Datenschutzgrundverordnung (DSGVO)¹ verstoßen können, indem sie personenbezogene Daten ohne Zustimmung manipulieren und verbreiten ([von Oelffen 2020](#), S. 179). Die DSGVO definiert personenbezogene Daten als „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“ (Art. 4 Nr. 1 DSGVO). Sie fordert den Schutz der Privatsphäre und Integrität personenbezogener Daten, die durch Deepfakes gefährdet sind. Aus Platzgründen werden die rechtlichen Aspekte hier jedoch nicht weiter vertieft.

5 Experteninterview: Generative KI in der Anwendung

Nachdem im vorangegangenen Kapitel die Funktionsweise von generativer KI und Chatbots am Beispiel von OpenAI aus der Perspektive des Anwenders betrachtet wurde, liegt der Fokus nun auf dem Einsatz dieser Systeme in Unternehmen. Im Zuge der digitalen Transformation spielen Chatbots und generative KI-Systeme im Unternehmenskontext eine immer wichtigere Rolle (siehe Kapitel 2.2). Ein schriftliches Experteninterview mit Jan Fischer liefert vertiefende Einblicke in die spezifischen Anforderungen, Herausforderungen und Optimierungsmöglichkeiten dieser Technologien in der praktischen Anwendung. Jan Fischer ist wissenschaftlicher Mitarbeiter im Forschungs- und Transferzentrum Business Innovation Lab an der Hochschule für Angewandte Wissenschaften Hamburg (HAW Hamburg) am Department Wirtschaft, und an verschiedenen Projekten beteiligt. Dazu gehören „EDIH4urbanSAVE“ und „Mittelstand-Digital Zentrum Hamburg“, die sich der Unterstützung von kleinen und mittleren Unternehmen (KMU) bei digitalen Transformationsprozessen widmen.

5.1 Optimierung durch Datenzufuhr und manuelles Training

Unternehmen können generative KI-Systeme durch gezielte Datenzuführung und manuelles Training an ihre spezifischen Bedürfnisse anpassen. Die Aussagen aus dem Basisinterview ([Fischer 2024](#)) mit Jan Fischer bieten wertvolle Einblicke und praktische Hinweise hierzu, auf die im Folgenden näher eingegangen wird. Laut

¹ Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/64/EG (Datenschutz-Grundverordnung) vom 27.04.2016, ABl. L 119, zuletzt geändert am 04.03.2021 [ELI: <http://data.europe.eu/eli/reg/2016/679/>].

Fischer ist die Fähigkeit, korrekte und nützliche Informationen bereitzustellen, ein zentrales Merkmal moderner Chatbots. Um diese Qualität zu gewährleisten, empfiehlt er den Einsatz von „Retrieval Augmented Generation“ (RAG). Diese Methode nutzt unternehmenseigene Wissensdatenbanken anstelle des allgemein trainierten Wissens eines Sprachmodells und ermöglicht es, maßgeschneiderte Informationen aus vordefinierten Datenquellen abzurufen ([Reifschneider 2024](#)). Für Unternehmen sei „RAG“ geeigneter als traditionelles „Fine Tuning“, da sich auf diese Weise die kontinuierlich verändernden Daten leichter integrieren und erneuern lassen. Er hebt außerdem die Bedeutung von „read-only APIs“ (Application Programming Interfaces) hervor, die es ermöglichen, unternehmensspezifische „hard facts“ wie Terminkalender nur abzurufen, ohne die Daten zu verändern. Diese Praxis minimiert das Risiko von Datenmanipulation und -verlust und sichert die Datenintegrität ([Marić 2022](#)).

Eine große Herausforderung bei der Entwicklung von Chatbots ist die Vermeidung von Halluzinationen und Fehlinformationen (siehe Kapitel 4). Obwohl Halluzinationen nie vollständig ausgeschlossen werden können, kann das Risiko laut Fischer durch die Verwendung von „RAG“ und die Anpassung der „Temperatur“ des Modells minimiert werden. Die „Temperatur“ ist ein Parameter, der in Modellen zur Verarbeitung natürlicher Sprache verwendet wird, um das Vertrauen des Modells in seine wahrscheinlichste Antwort zu erhöhen oder zu verringern ([Paaß und Giesselbach 2023](#), S. 133f.). „Man erhält präzisere, aber weniger kreative Antworten bzw. weniger Varianz in den Antworten“, erklärt Fischer ([2024](#)). Dies sei vor allem im Unternehmenskontext von Vorteil, da hier häufig spezifische und verlässliche Informationen benötigt werden. Fischer bestätigt auch, dass KI-Systeme die spezifische Tonalität und den Stil eines Unternehmens vermitteln können. Dies könne durch die Analyse vorhandener Texte geschehen, um daraus die gewünschte Tonalität abzuleiten und diese durch gezieltes Prompting anzuwenden. „Vermutlich besser“ sei jedoch eine Optimierung durch „Fine Tuning“ ([Fischer 2024](#)) (siehe Kapitel 3.2).

Langfristig würden generative KI-Systeme, insbesondere LLM-basierte Chatbots, eine integrale Rolle bei der digitalen Transformation von Unternehmen spielen, da sie zunehmend in bestehende Produkte integriert würden. Beispiele wie der Microsoft Office 365 Copilot oder Apple Intelligence verdeutlichten diese Entwicklung. Die Nutzung solcher Systeme werde immer alltäglicher und so selbstverständlich „wie heute die Rechtschreibprüfung“, so Fischer ([2024](#)). Darüber hinaus könnten Chatbots Routineaufgaben im Kundensupport oder in hochautomatisierten Workflows übernehmen und so die Effizienz und Produktivität deutlich steigern. In Unternehmen würden typischerweise generative KI-Systeme wie Microsoft Copilot, ChatGPT und GitHub Copilot eingesetzt, die alle auf OpenAI basieren. Es gebe aber auch ein wachsendes Interesse an lokalen Lösungen, die ohne externe Clouds oder APIs auskommen, auch wenn diese derzeit noch vergleichsweise wenig verbreitet seien. Der Integrationsprozess sei je nach Anwendungsfall sehr unterschiedlich. Ein allge-

meiner Chatbot mit „RAG“ könne oft von einem Drittanbieter gehostet und über ein Java-Script-Snippet in die Website des Unternehmens integriert werden. Für komplexere Anwendungen, wie automatisierte Agenten, sei eine sorgfältige Integration in die Unternehmenssysteme notwendig ([Fischer 2024](#)).

5.2 Anwendungsfall: Einfache Chatbots für die Generierung von Website-Texten

Da laut Jan Fischer der Integrationsprozess von KI-Systemen in Unternehmensprozesse stark vom Anwendungsfall abhängt (siehe Kapitel 5.1), wurde in einem Folgeinterview (siehe [Fischer 2024](#)) der Fokus auf praktische Anwendungsbeispiele gelegt. Dabei erläuterte der Experte verschiedene Aspekte der Implementierung und des Einsatzes von einfachen Chatbots zur Generierung von Website-Texten. Ein solcher Chatbot werde in der Regel auf Basis der eigenen Website oder eines Produktkatalogs trainiert und dann in die Website eines kleinen oder mittleren Unternehmens (KMU) integriert. So werde es den Nutzenden ermöglicht, über eine Chatblase auf der Website Fragen zu Produkten, Terminverfügbarkeiten oder anderen Inhalten zu stellen. Fischer vertieft ebenfalls die technische Umsetzung eines solchen Chatbots, die durch die Digitalisierung und Tokenisierung relevanter Daten wie Produktdatenblätter erfolgt: Die Daten werden einmalig in Tokens umgewandelt und bei jeder Nutzendenanfrage mit den Tokens der Anfrage verglichen (siehe Kapitel 3.1). Dabei wird ein Ähnlichkeitsmaß verwendet, um passende Textblöcke zu identifizieren. Kommerzielle Chatbots, wie beispielsweise die von OpenAI, nutzen diese Kontextinformationen und bieten den Nutzenden entsprechende Antworten an ([OpenAI 2022](#)). Der Vorteil dieser Methode liege in ihrer Kosteneffizienz, da nicht jedes Mal umfangreiche PDF-Dokumente verarbeitet werden müssen. Fischer ([2024](#)) erklärt: „Das hat den großen Vorteil, dass man nur einmal die Daten digitalisieren oder in Token umwandeln muss und beim zweiten Mal die Kosten sparen kann.“

Ein weiterer wichtiger Aspekt, den Fischer ([2024](#)) hervorhebt, ist die Aufbereitung und Bereitstellung der Daten durch das Unternehmen. Es sei sinnvoll, im Vorfeld häufige Nutzendenanfragen zu analysieren, um die relevanten Informationen gezielt bereitstellen zu können. Dazu könnten beispielsweise Mailprotokolle oder Kontaktdaten ausgewertet werden. Ebenso sollten Produktdatenblätter optimiert werden, indem unnötige Informationen entfernt werden, um die Daten für den Chatbot nutzbar zu machen. Fischer ([2024](#)) dazu: „Da sind manchmal viele Tabellen oder Schemazeichnungen oder Ähnliches dabei. Das sind Informationen, die ein Chatbot in der Regel im ersten Schritt nicht verarbeiten kann.“

Für den laufenden Betrieb eines solchen Chatbots nannte Fischer ([2024](#)) mehrere Optimierungsmöglichkeiten. So könnten die Prompts des Chatbots angepasst werden, um spezifische Ziele wie Upselling zu fördern. Bei Produktänderungen müssten die Daten aktualisiert werden. Außerdem empfiehlt Fischer, regelmäßig zu überprüfen, welche Fragen häufig gestellt werden, um den Chatbot entsprechend anzupas-

sen:

„Ich kann mir angucken, welche Fragen gestellt werden. Ich kann mir angucken, welche Antworten er gibt. Ich kann den Chatbot auch so bauen, dass er stets auf Kontaktinformationen oder Ähnliches verweist, damit ich sehe, ob ich damit wirklich aktiv Leads generiere oder nicht.“ ([Fischer 2024](#))

5.3 Potenziale für Nachrichtendienste am Beispiel von Retresco-Tools

Nach der Betrachtung einfacher Chatbots wird im folgenden Abschnitt die Integration generativer KI-Systeme in die journalistische Praxis skizziert. Anhand des Beispiels der Retresco-Tools werden erste Potenziale und Herausforderungen identifiziert. Dieser Abschnitt bietet einen Ausblick auf die automatisierte Texterstellung für Nachrichtendienste und legt nahe, dass eine vertiefte Analyse dieses komplexen Anwendungsfeldes für zukünftige Untersuchungen von Interesse sein könnte.

Die Tools von Retresco basieren auf dem Sprachmodell GPT-3 und bieten vielfältige Anwendungen zur Textgenerierung. Die Anwendung „Textengine.io“ von Retresco nutzt das GPT-3-Sprachmodell, um automatisiert qualitativ hochwertige Texte zu generieren. Diese Technologie ermöglicht es Nachrichtendiensten, große Mengen an Artikeln effizient zu erstellen, ohne dabei auf menschliche Autoren angewiesen zu sein. Durch den hybriden Einsatz von NLG und GPT-3 kann „Textengine.io“ ein breites Spektrum an Texten in verschiedenen Tonalitäten und Stilen erzeugen ([Linnik 2022](#)). Dies bietet Nachrichtendiensten die Möglichkeit, Inhalte schneller und kostengünstiger zu produzieren und damit ihre Reichweite zu erhöhen.

Die Automatisierung von Texten durch die Lösungen von Retresco geht über die reine Textgenerierung hinaus. Wie auf der offiziellen Website von Retresco ([Retresco o.D.-b](#)) beschrieben wird, ermöglichen diese Tools die vollständige Automatisierung des redaktionellen Prozesses von der Datenerfassung bis zur Veröffentlichung. Dies ist besonders vorteilhaft für Nachrichtendienste, die täglich eine große Anzahl von Artikeln veröffentlichen und dabei eine einheitliche Qualität und Tonalität sicherstellen müssen. Ein weiteres Anwendungsbeispiel ist der Einsatz von generativer KI im Lokaljournalismus. So wird das Tool „Rewrite“ beispielsweise von einer süddeutschen Regionalzeitung eingesetzt ([Sommer 2024](#)). Besonders im regionalen Kontext, in dem häufig Ressourcenknappheit herrsche, könne die automatisierte Textgenerierung eine wertvolle Unterstützung sein, um zeitnah und umfassend über lokale Ereignisse zu berichten. Retresco bietet mit „Rewrite“ eine Lösung an, die Tonalität und den Stil von Texten gezielt anzupassen ([Retresco o.D.-a](#)). Dies ist besonders für Nachrichtendienste relevant, die ihre Inhalte an unterschiedliche Zielgruppen anpassen müssen. Im Vergleich zu anderen GPT-basierten Systemen soll „Rewrite“ in der Lage sein, sprachliche Feinheiten zu erfassen und damit Texte zu generieren, die sowohl inhaltlich präzise als auch stilistisch ansprechend sind.

6 Fazit und Ausblick

Zusammenfassend lässt sich sagen, dass die Untersuchung der Rolle generativer KI und ihres Einflusses auf die Tonalität der Nachrichtenverbreitung und -wahrnehmung sowohl Chancen als auch Herausforderungen aufzeigt.

Generative KI-Modelle, wie sie von OpenAI entwickelt werden, zeigen bemerkenswerte Fähigkeiten bei der Erstellung menschenähnlicher Texte, was zu verbesserten Interaktionen mit den Nutzenden und betrieblicher Effizienz führt. Mit diesen Fortschritten geht jedoch das Risiko der Verbreitung von Fehlinformationen einher. Die Gefahr, dass KI-generierte Inhalte Ungenauigkeiten verbreiten, erfordert rigorose Strategien zur Minderung dieser Probleme. Ansätze wie „RAG“, sorgfältiges „Fine-Tuning“ und „RLHF“ können die Zuverlässigkeit und Genauigkeit von KI-Ergebnissen gewährleisten. Auch die Tonalität von KI-generierten Texten spielt eine zentrale Rolle für die Wahrnehmung und Akzeptanz der Informationen. Eine gut abgestimmte Tonalität kann die Empathie und das Vertrauen in die angebotenen Inhalte erhöhen. Gleichzeitig besteht die Herausforderung, dass subtile Nuancen wie Ironie oder kulturelle Unterschiede möglicherweise nicht korrekt erfasst werden, was zu Missverständnissen führen kann. Die Steuerung der Tonalität durch KI bietet das Potenzial, den Stil, den Kontext und die emotionale Färbung des Outputs an die menschliche Sprache und spezifische Aufgabenbereiche anzupassen.

Die Integration von generativer KI in die Medienlandschaft verändert journalistische Praktiken und erweitert die Grenzen der traditionellen Nachrichtenproduktion. Der „automatisierte Journalismus“ bietet Effizienz und Skalierbarkeit, wirft jedoch ethische Fragen und Herausforderungen bei der Qualitätskontrolle auf. Ein ausgewogener und verantwortungsvoller Einsatz von KI-Technologien ist entscheidend, um das Vertrauen der Öffentlichkeit und die Integrität der Informationen zu wahren. Einige Nachrichtenorganisationen arbeiten bereits daran, KI auf ihre ethischen und redaktionellen Richtlinien abzustimmen. Eine vielversprechende Lösung könnte der „erweiterte Journalismus“ sein, bei dem menschliche redaktionelle Einsichten mit KI kombiniert werden ([Beckett 2019](#), S. 58f.). Dies impliziert auch, dass KI-Anwendungen oder entsprechende Datensätze unter Umständen nicht für redaktionelle Zwecke eingesetzt werden sollten, wenn sie unzuverlässig sind und Verzerrungen nicht identifiziert oder charakterisiert werden können. „Voreingenommenheit ist allen Geschichten inhärent. KI ist nur so gut wie ihre Eingaben, und da sie von Menschen gemacht wird, kann sie aufgrund der Voreingenommenheit ihres Schöpfers voreingenommen sein“ ([Beckett 2019](#), S. 59).

In Zukunft wird die kontinuierliche Verbesserung der KI-Algorithmen in Kombination mit einer robusten menschlichen Aufsicht von entscheidender Bedeutung sein, um die Vorteile der generativen KI zu nutzen und gleichzeitig ihre Nachteile zu minimieren. Die Zukunft der Informationsverbreitung liegt in einer symbiotischen Beziehung

zwischen hochentwickelten KI-Systemen und menschlicher Expertise, um sicherzustellen, dass die Technologie dazu dient, die Qualität und Zuverlässigkeit von Informationen in unserer Gesellschaft zu verbessern, anstatt sie zu gefährden. Zukünftige Forschung sollte neben dem Journalismus auch den wissenschaftlichen Kontext berücksichtigen, der ebenfalls bei der Informationsbereitstellung spezifische Qualitätsanforderungen verfolgt und somit als wichtige Institution in einer zunehmend fragmentierten Gesellschaft bewertet werden kann. Auch der wissenschaftliche Diskurs könnte von den Vorteilen und Herausforderungen unterstützender KI-Tools betroffen sein.

Literatur

BAI, Zechen, et al., 2024. Large Language Models as Optimizers. In: *arXiv* [online]. Preprint abs/2404.18930 [Zugriff am: 29.10.2024]. arXiv.org e-print archive. Verfügbar unter: DOI: [10.48550/arxiv.2404.18930](https://doi.org/10.48550/arxiv.2404.18930)

BARTL, Marc, 2019. *Studie: Nicht einmal jeder Zweite erkennt einen Text von einer KI* [online]. Eugendorf: Medienfachverlag Johann Oberauer, 14.08.2019 [Zugriff am: 04.11.2024]. Verfügbar unter: <https://kress.de/news/beitrag/137855-studie-nicht-einmal-jeder-zweite-erkennt-einen-text-von-einer-ki.html>

BECKETT, Charlie, 2019. *The Journalism AI report: New powers, new responsibilities* [online]. *A global survey of journalism and artificial intelligence*. London: LSE Press [Zugriff am: 04.11.2024]. PDF E-Book. Verfügbar unter: <https://drive.google.com/file/d/1utmAMCmd4rfJHrUfLLfSJ-clpFTjyef1/view>

BROWN, Tom B., et al., 2020. Language Models are Few-Shot Learners. In: *arXiv* [online]. Preprint abs/2005.14165 [Zugriff am: 04.11.2024]. arXiv.org e-print archive. Verfügbar unter: DOI: [10.48550/arxiv.2005.14165](https://doi.org/10.48550/arxiv.2005.14165)

DEUTSCHER BUNDESTAG, 2020. *Unterrichtung der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale. Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale. Drucksache 19/23700*. Berlin: Bundesregierung. Verfügbar unter: <https://dserver.bundestag.de/btd/19/237/1923700.pdf>

DEUTSCHER ETHIKRAT, 2023. *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz* [online]. *Stellungnahme*. Berlin: Deutscher Ethikrat [Zugriff am: 05.11.2024]. PDF E-Book. Verfügbar unter: <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf>

DÖRR, Konstantin N., 2015. Mapping the field of Algorithmic Journalism. In: *Digital Journalism* [online]. 4(6), S. 700–722 [Zugriff am: 04.11.2024]. Taylor and Francis Online. DOI: 10.1080/21670811.2015.1096748

FISCHER, Jan, 2024. *Experten-Interview mit Jan Fischer*. Hamburg: SAIL HAW Hamburg. [Zugriff am 02.12.2024]. Verfügbar unter: <https://hdl.handle.net/20.500.12738/16574>

GAO, Leo, SCHULMAN, Jan und HILTON, Jacob, 2022. Scaling Laws for Reward Model Overoptimization. In: *arXiv* [online]. Preprint abs/2210.10760 [Zugriff am: 05.11.2024] arXiv.org e-Print archive. Verfügbar unter: DOI: [10.48550/arxiv.2210.10760](https://doi.org/10.48550/arxiv.2210.10760)

HAIM, Mario und GRAEFE, Andreas, 2018. Automatisierter Journalismus: Anwendungsbereiche, Formen und Qualität. In: NEUBERGER, Christoph und NUERNBERGK, Christian, Hrsg. *Journalismus im Internet: Profession – Partizipation – Technisierung* [online]. 2. Auflage. Wiesbaden: Springer VS, S. 139-160 [Zugriff am: 04.11.2024]. PDF E-Book. ISBN 978-3-531-93284-2. Verfügbar unter: DOI: 10.1007/978-3-531-93284-2_5

KÖRNER, Theresa, 2023. *Generalisiertes Vertrauen in automatisierten Journalismus*. Wiesbaden: Springer VS [Zugriff am: 04.12.2024]. Verfügbar unter: DOI: [10.1007/978-3-658-42735-1](https://doi.org/10.1007/978-3-658-42735-1)

LERNENDE SYSTEME - DIE PLATTFORM FÜR KÜNSTLICHE INTELLIGENZ, 2023. *Künstliche Intelligenz im Journalismus* [online]. *Potenziale und Herausforderungen für Medienschaffende*. München: Lernende Systeme - Die Plattform für Künstliche Intelligenz, 01.2023 [Zugriff am: 29.10.2024]. Verfügbar unter: <https://www.acatech.de/publikation/kuenstliche-intelligenz-im-journalismus/download-pdf?lang=de>

LINNIK, Anastasia, 2022. *NLG der Zukunft? Von datenbasierter Content-Erstellung über End-to-End-Textgenerierung mit GPT-3 zur hybriden Natural Language Generation* [online]. Berlin: Retresco, 28.09.2022 [Zugriff am: 05.11.2024]. Verfügbar unter: <https://www.retresco.de/blog/hybrid-nlg-gpt3-sprachmodell-textgenerierung>

LOVRINOVIC, Colin und VOLLAND, Holger, 2019. *The Future Impact of Artificial Intelligence on the Publishing Industry* [online]. Berlin: Gould Finch/Frankfurter Buchmesse [Zugriff am: 04.11.2024]. PDF E-Book. Verfügbar unter: <https://www.kidstarrapp.com/storage/app/public/papers/6IPGoLaDzt0aOrMnAFq5jphalu1CgTBTIErkCNNg.pdf>

MARIĆ, Nedim, 2024. *What Is API security? The Complete Guide* [online]. San Rafael: Bright Security, 23.01.2024 [Zugriff am: 05.11.2024]. Verfügbar unter: <https://brightsec.com/blog/api-security/>

NEUBERGER, Christoph, 2022. Journalism and Platforms as Mediating Third Parties in the Digital Public Sphere. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* [online]. 74(S1), S. 159-181 [Zugriff am: 04.11.2024]. Springer Nature Link. Verfügbar unter: DOI: [10.1007/s11577-022-00832-9](https://doi.org/10.1007/s11577-022-00832-9)

OPENAI, 2024. *Alles über ChatGPT* [online]. San Fransisco: OpenAI, 21.06.2024 [Zugriff am: 05.11.2024]. Verfügbar unter: <https://chatopenai.de/>

OPENAI, 2023. *Introducing GPT* [online]. San Fransisco: OpenAI, 6.11.2023 [Zugriff am: 05.11.2024]. Verfügbar unter: <https://openai.com/index/introducing-gpts/>

OPENAI, 2022. *Introducing ChatGPT* [online]. San Fransisco: OpenAI, 30.11.2022 [Zugriff am: 05.11.2024]. Verfügbar unter: <https://openai.com/index/chatgpt/>

PAAß, Gerhard und GIESSELBACH, Sven, 2023. *Foundation Models for Natural Language Processing* [online]. *Pre-trained Language Models Integrating*. Chan: Springer Nature Switzerland [Zugriff am: 05.11.2024]. PDF E-Book. Verfügbar unter: DOI: 10.1007/978-3-031-23190-2

REIFSCHNEIDER, Florian, 2024. *Was ist RAG?* [online]. *Wie Retrieval Augmented Generation Ihr Unternehmen voranbringen kann*. Frankfurt: Omnifact, 25.01.2024 [Zugriff am: 05.11.2024]. Verfügbar unter: <https://omnifact.ai/de/blog/what-is-rag-how-retrieval-augmented-generation-benefits-business>

RETRESCO, o.D.-a. *Rewrite – Modifiziere Texte für deine Zwecke und Zielgruppen* [online]. Berlin: Retresco [Zugriff am: 05.11.2024]. Verfügbar unter: <https://www.retresco.de/rewrite>

RETRESCO, o.D.-b. *Text-Automation für mehr Effizienz in der Content-Produktion*. Berlin: Retresco [Zugriff am: 05.11.2024]. Verfügbar unter: <https://www.retresco.de/loesungen/text-automation>

SCHÖBEL, Sofia et al., 2024. Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers. In: *Information Systems Frontiers* [online]. 26(2), S. 729-754 [Zugriff am: 30.12.2024]. Springer. Verfügbar unter: DOI: [10.1007/s10796-023-10375-9](https://doi.org/10.1007/s10796-023-10375-9)

SOMMER, Johannes, 2019. *Generative KI im Lokaljournalismus* [online]. *5 Best Practice Beispiele aus den Redaktionen*. Berlin: Retresco, 12.04.2019 [Zugriff am: 05.11.2024]. Verfügbar unter: <https://www.retresco.de/blog/generative-ki-im-lokaljournalismus-5-best-practice-beispiele>

VON OELFFEN, Sabine, 2020. Ausblick: Vorhaben und Handlungsfelder der EU mit Bezug zu KI. In: BALLESTREM, Johannes G. et al., Hrsg. *Künstliche Intelligenz: Rechtsgrundlagen und Strategien in der Praxis* [online]. Wiesbaden: Springer Gabler, S. 166-190 [Zugriff am: 04.11.2024]. PDF E-Book. Verfügbar unter: DOI: 10.1007/978-3-658-30506-2

VASWANI, Ashish, et al., 2017. Attention is all you need. In: *arXiv* [online]. Preprint abs/1706.03762 [Zugriff am: 05.11.2024]. arXiv.org e-print archive. Verfügbar unter: DOI: [10.48550/arxiv.1706.03762](https://doi.org/10.48550/arxiv.1706.03762)

ZUMBÜHL, Thomas, 2023. *Content Produktion: Tonalität in ChatGPT & Co* [online]. Aadorf: Hutter Consult, 25.04.2024 [Zugriff am: 05.11.2024]. Verfügbar unter: <https://www.thomashutter.com/content-tonalitaet-in-chatgpt-co/>