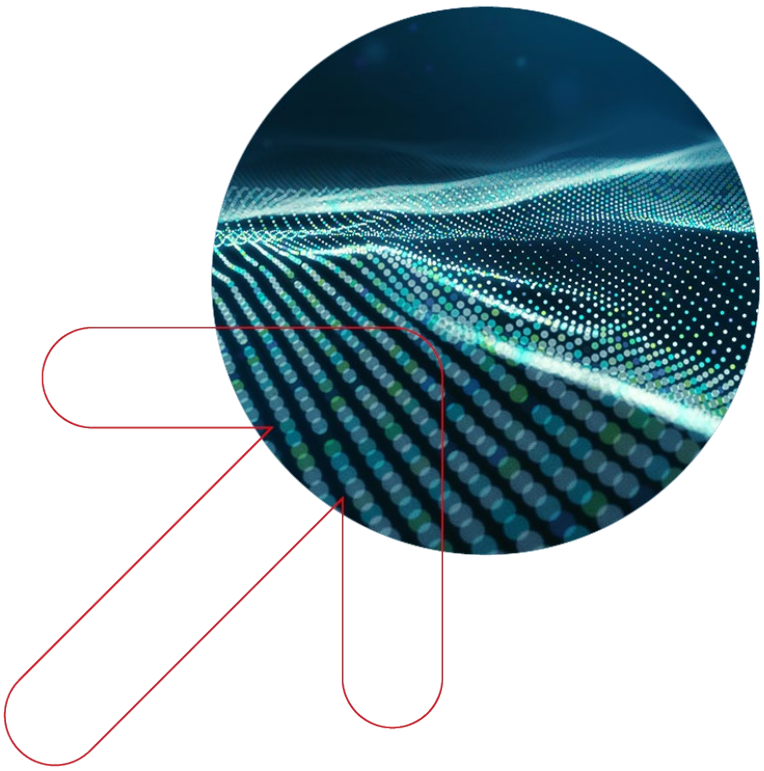
WIK • Kurzstudie



Generative KI – Zwei verbraucherseitige Untersuchungen

Autoren: Serpil Taş, Andrea Liebe, Lukas Wiewiorra

Bad Honnef, 27.12.2024

WIK
Wissenschaftliches Institut
für Infrastruktur und
Kommunikationsdienste

Impressum

WIK Wissenschaftliches Institut für
Infrastruktur und Kommunikationsdienste GmbH
Rhöndorfer Str. 68
53604 Bad Honnef
Deutschland
Tel.: +49 2224 9225-0
Fax: +49 2224 9225-63
E-Mail: info@wik.org
www.wik.org

Vertretungs- und zeichnungsberechtigte Personen

Geschäftsführerin und Direktorin	Dr. Cara Schwarz-Schilling
Direktor, Verwaltungs- und Abteilungsleiter	Alex Kalevi Dieke
Direktor, Abteilungsleiter	Prof. Dr. Bernd Sörries
Abteilungsleiter	Dr. Christian Wernick
Abteilungsleiter	Dr. Lukas Wiewiorra
Vorsitzender des Aufsichtsrates	Dr. Thomas Solbach
Handelsregister	Amtsgericht Siegburg, HRB 7225
Steuer-Nr.	222/5751/0722
Umsatzsteueridentifikations-Nr.	DE 123 383 795

Stand: Januar 2024

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einleitung

Die Verbreitung von Künstlicher Intelligenz (KI) nimmt weltweit rapide zu und erfasst nahezu alle Wirtschafts- und Gesellschaftsbereiche. Fortschritte in Hardwareleistung, Datenverfügbarkeit und Algorithmik haben die Implementierung von KI-Lösungen erheblich beschleunigt. Ein spezifischer Bereich, die generative KI, umfasst fortschrittliche Systeme, die eigenständig neue Inhalte wie Texte, Bilder oder Töne erzeugen können, indem sie aus bestehenden Datenmustern lernen.[1]

Generative KI bietet Organisationen und Unternehmen eine breite Palette an Einsatzmöglichkeiten. Im Bereich Marketing und Vertrieb kann sie genutzt werden, um Kundenkommunikation zu automatisieren, personalisierte Werbeinhalte zu erstellen und Kundendaten zu analysieren, um gezielte Kampagnen zu entwickeln. In Forschung und Entwicklung trägt generative KI zur Optimierung von Designprozessen oder zur Entwicklung neuer Produkte bei, beispielsweise durch Simulationen oder automatisierte Content-Erstellung. Weitere Anwendungsfelder sind das Supply Chain Management, wo sie Prognosen für Nachfrage und Bestände verbessern kann, sowie die IT, durch Automatisierung von Code-Generierung und Fehlersuche. Auch in der Unternehmensführung erleichtert generative KI strategische Entscheidungsprozesse, etwa durch datenbasierte Szenarien-Analysen.[2] In der Literatur wird der Einsatz von generativer KI als vorteilhaft beschrieben, da sie die Effizienz und Effektivität erhöht, Innovationen fördert, die Ressourcennutzung optimiert und langfristig zur Kostenreduktion beiträgt. Diese Eigenschaften machen KI zu einem zentralen Treiber der digitalen Transformation in Unternehmen.[3]

Neben den Potenzialen ist generative KI jedoch auch mit Risiken verbunden. Studien zeigen, dass KI-Systeme nicht immer korrekte oder sozial erwünschte Ergebnisse liefern. Besonders problematisch sind Verzerrungen, die durch fehlerhafte Daten, Trainingsmethoden oder Anwendungskontexte entstehen, was zu diskriminierenden Ergebnissen führen kann. Ein spezifisches Risiko ist die sogenannte „Halluzination“, bei der KI-Modelle plausible, aber faktisch falsche oder kontextuell irrelevante Ergebnisse generieren. Der GPT-4 Technical Report von OpenAI beschreibt diese Problematik als Folge inhärenter Verzerrungen, begrenzten Weltverständnisses und unvollständiger Trainingsdaten.[4] Diese Risiken verdeutlichen die Notwendigkeit robuster Kontrollmechanismen und einer verantwortungsvollen Entwicklung von KI-Systemen.

Die bewusste Nutzung von generativer KI zur Erstellung und Verbreitung von manipulativen oder gar falschen Informationen oder Deep Fakes ist ebenfalls ein zentrales Problem und kann zu gravierenden gesellschaftlichen Risiken führen. Obwohl das Konzept der Verbreitung von Falschinformationen im Internet nicht neu ist, verstärken KI-Techniken das Phänomen.[5] Im Vergleich zum Menschen kann sie schnellere sowie genauere und leichter verständliche Informationen produzieren, und damit auch überzeugendere Fehlinformationen generieren.[6] Angesichts der zahlreichen Anwendungsfälle, aber auch der Risiken, die mit dem Einsatz von KI verbunden sind, gibt es insbesondere zwei wesentliche

---

1 Brüns, J.D./ Meißner, M. (2024).
2 Deloitte (2023); McKinsey & Company (2023).
3 Deloitte (2023); Gillespie, N. et al. (2023); KPMG (2023).
4 Capgemini (2023).
5 Bontridder, N./ Poullet, Y. (2021).
6 Spitale, G. et al. (2023).

Rechtstexte, die Transparenz und Sicherheit gewährleisten sollen: der AI Act[7] und der Digital Markets Act (DSA)[8].

Der AI Act verfolgt das Ziel, ein ausgewogenes Verhältnis zwischen technologischer Entwicklung und dem Schutz der Gesellschaft zu schaffen. Er kombiniert die Förderung von Innovationen mit der Minimierung von Risiken und bietet Unternehmen sowie Organisationen ein rechtliches Rahmenwerk, um KI-Systeme sicher und effizient zu entwickeln und anzuwenden. Dabei soll das Vertrauen der Öffentlichkeit in KI gestärkt werden. Ein zentraler Ansatz des AI Act ist die risikobasierte Regulierung. KI-Anwendungen werden nach ihrem potenziellen Risiko für Individuen und die Gesellschaft kategorisiert – von minimalen bis hin zu inakzeptablen Risiken. Hochriskante Systeme unterliegen strengen Anforderungen, während Anwendungen mit niedrigem Risiko weniger reguliert werden. Darüber hinaus setzt der AI Act auf Transparenz, etwa durch die Kennzeichnung von KI-generierten Inhalten, und stellt sicher, dass Verantwortlichkeiten klar definiert sind, um Missbrauch und Fehlverhalten vorzubeugen.

Der DSA adressiert generative KI indirekt, indem er sich auf die systemischen Risiken konzentriert, die von digitalen Plattformen ausgehen, insbesondere wenn KI-Technologien integriert sind. Er verpflichtet sehr große Online-Plattformen (VLOPs) und sehr große Suchmaschinen (VLOSEs), systemische Risiken zu bewerten und zu minimieren. Der Text bezieht sich ausdrücklich auch auf Desinformationen, deren koordinierte Verbreitung oder Verstärkung zu systemischen Risiken führen kann.[9] In den Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und -Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 des DSA wird zudem empfohlen, KI-generierter Inhalte zu kennzeichnen.[10]

Vor diesem Hintergrund ergeben sich zwei zentrale Forschungsziele:

- Einheitliche Erfassung der Einstellungen der deutschen Verbraucher zu generativer KI und deren Nutzung sowie die Einschätzung des Risikos der Anwendung von generativer KI in verschiedenen Sektoren
- Analyse der Wahrnehmung und Verbreitung von Informationen im Internet.

Um die Einstellungen und Einschätzungen der Verbraucher sowie die Wahrnehmung und Verbreitung von Informationen im Internet zu erfassen, wurde eine repräsentative Befragung durchgeführt. Im Folgenden werden die Ergebnisse detailliert vorgestellt. Der Aufbau der Studie gestaltet sich wie folgt:

Nach dieser Einleitung folgt in Kapitel 2 eine Darstellung der Datenerhebung. Kapitel 3 und Kapitel 4 präsentieren die Ergebnisse der Befragung und rücken diese, soweit möglich, in den Kontext existierender Studien.

---

7 Verordnung (EU) 2024/1689 des europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz), Amtsblatt der Europäischen Union, Reihe L.

8 Verordnung (EU) 2022/2065 des europäischen Parlaments und des Rates vom 19. Oktober 2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste), Amtsblatt der Europäischen Union, L 277/1.

9 Siehe beispielsweise Erwägungsgründe 83 & 84 des DSA.

10 C/2024/3014 Mitteilung der Kommission: Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065, Amtsblatt der Europäischen Union, Reihe C.

Kapitel 3 geht zunächst auf die Nutzung von generativer KI ein, erörtert dann Kenntnisstand und Einstellungen zu generativer KI und widmet sich darauf aufbauend der Bewertung von KI durch die Befragten. Dabei wird der Einsatz von KI in verschiedenen Sektoren thematisiert und danach gefragt, inwiefern spezifische Maßnahmen des AI Acts geeignet erscheinen, mögliche Bedenken und Risiken abzumindern.

Kapitel 4 betrachtet insbesondere den Effekt eines KI-Labels auf die Bewertung der Korrektheit einer Information und die Verbreitung von Informationen.

Die Studie schließt mit einer Schlussbetrachtung.

2 Datenerhebung

Die im Folgenden untersuchten Daten wurden anhand einer Online-Verbraucherbefragung im November 2024 durch Computer Aided Web Interviews (CAWI) erhoben. Die Stichprobengröße lag bei 3.201 Befragten. Um eine Zusammenstellung der Stichprobe zu gewährleisten, die die deutsche Bevölkerung ab 18 Jahren angemessen abbildet, wurde die Ziehung einer Quotenstichprobe veranlasst. Die Ausstreuung der Stichprobe für diese Studie erfolgte hauptsächlich nach den Merkmalen Alter, Geschlecht und Region. Die Befragung bestand in erster Linie aus geschlossenen Fragen und wurde in deutscher Sprache durchgeführt. Die Befragung leitete mit einer kurzen Beschreibung von generativer KI angelehnt an Brüns & Meißner (2024) ein.[11]

Im ersten Teil der Befragung lag der Fokus auf der Nutzung von generativer KI durch die befragten Verbraucher sowohl im privaten als auch im beruflichen Kontext sowie ihre Einstellungen im Allgemeinen gegenüber generativer KI. Für letztere wurde auf bestehende Skalen aus der Literatur zurückgegriffen, auf die an der entsprechenden Stelle einzeln verwiesen wird. Darüber hinaus wurden die Bedenken der Verbraucher hinsichtlich des Einsatzes von Anwendungen der generativen KI in Unternehmen und Organisationen aus verschiedenen gesellschaftlichen Bereichen erhoben. Schließlich wurden die Befragten um eine Einschätzung gebeten, inwieweit die im AI Act vorgeschlagenen Maßnahmen zur Minderung möglicher Risiken durch KI die allgemeinen Bedenken gegenüber der Nutzung generativer KI ausräumen können.

Im zweiten Teil der Untersuchung wurde ein Experiment durchgeführt, um herauszufinden, ob ein KI-Label die Wahrnehmung und Verbreitung von Informationen beeinflusst und wie diese Wirkung mit dem Kontext zusammenhängt, in dem die Informationen präsentiert werden. Eine Beschreibung der Methodik findet sich in Kapitel 4.1. Zusätzlich wurden Fragen zur tatsächlichen Erkennung von Falschinformationen gestellt.

---

11 Brüns, J.D./ Meißner, M. (2024).

3 Einstellung zu und Nutzung von generativer KI in Deutschland

In diesem Kapitel werden die Ergebnisse der Befragung zum Umgang mit generativer KI und zur Risikowahrnehmung ihres Einsatzes in verschiedenen Branchen dargestellt.

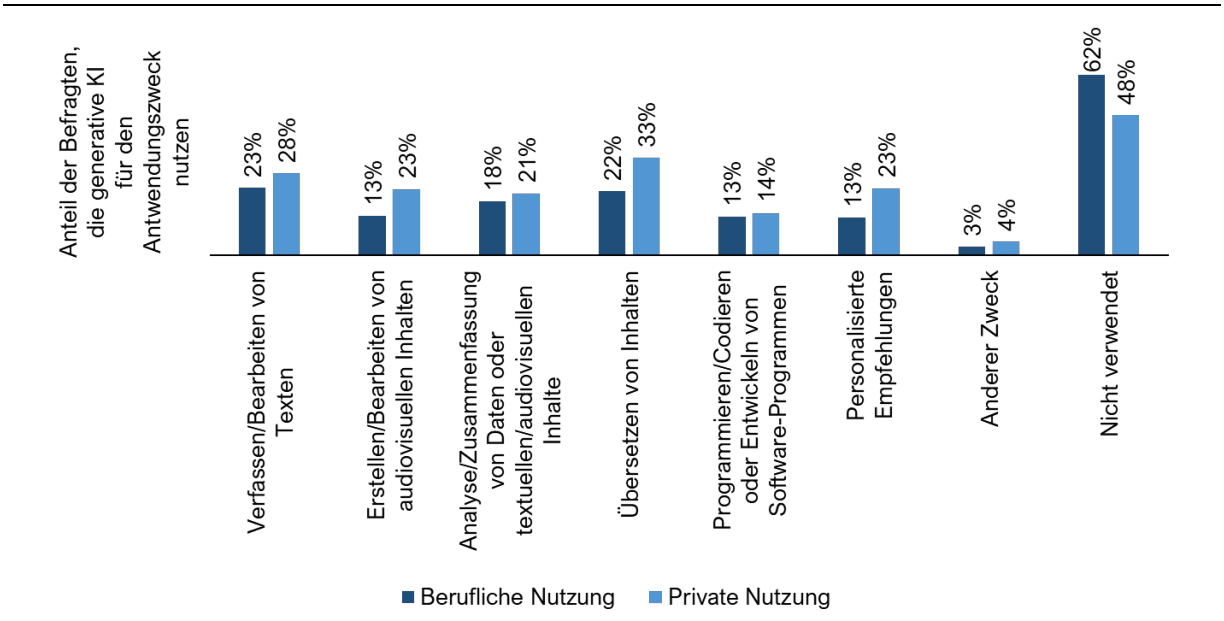
3.1 Nutzer und Nichtnutzer

Die Ergebnisse der Befragung zeigen, dass generative KI-Technologien bislang von einer signifikanten Anzahl von Personen nicht genutzt werden. 44 % der Befragten gaben an, generative KI überhaupt nicht zu verwenden, was auf eine gewisse Zurückhaltung oder möglicherweise mangelnde Bekanntheit solcher Technologien hinweisen könnte. 56 % der Befragten setzen hingegen generative KI entweder für berufliche oder private Zwecke ein. Der Großteil der Nutzer, nämlich 62 % verwenden diese Technologien sowohl im beruflichen als auch im privaten Kontext.

Das Bild, das sich bei der Betrachtung des Einsatzes von generativer KI nach Anwendungsbereichen ergibt, ist recht heterogen. Jedoch liegen die häufigsten Einsatzgebiete im Bereich des Verfassens und Bearbeitens von Texten sowie der Übersetzung von Inhalten.

Insgesamt verdeutlichen die Daten, dass die Nutzung generativer KI trotz ihres Potenzials noch nicht umfassend in allen Lebensbereichen angekommen ist.

Abbildung 3-1: Nutzung von generativer KI – berufliche vs. private Nutzung



Quelle: Befragung des WIK. N=3.201.

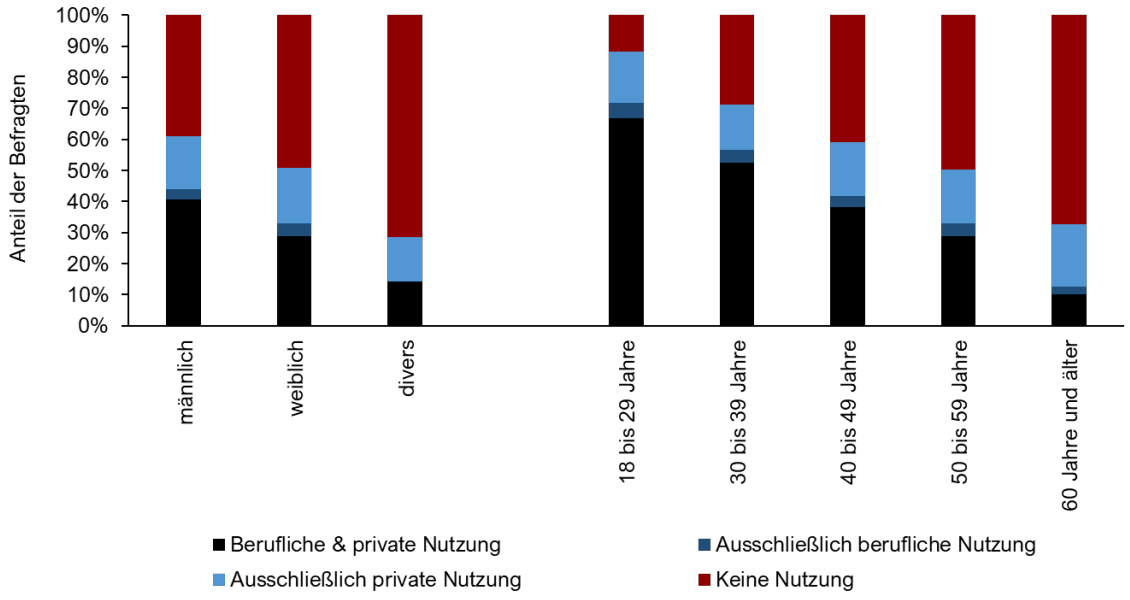
Die Daten zeigen weiter, dass deutliche Unterschiede in der Verbreitung der Nutzung generativer KI nach Geschlecht und Altersgruppen bestehen.

Männliche Befragte verwenden generative KI häufiger als weibliche Befragte. Dieses Geschlechtergefälle könnte auf verschiedene Faktoren zurückzuführen sein, darunter Unterschiede im Interesse an Technologie, in der Vertrautheit mit digitalen Anwendungen oder im Zugang zu entsprechenden Ressourcen. Die Ergebnisse könnten auch geschlechtsspezifische Unterschiede in der wahrgenommenen Nützlichkeit oder den Einsatzmöglichkeiten von generativer KI widerspiegeln. Untersuchungen, wie zum

Beispiel im Rahmen des Gender Equality Index 2020, weisen auf diese Unterschiede hinsichtlich neuer Technologien hin.[12]

Die Nutzung generativer KI nimmt mit steigendem Alter der Befragten ab. Dies deutet darauf hin, dass jüngere Altersgruppen technologieaffiner sind und möglicherweise weniger Barrieren beim Einsatz neuer digitaler Technologien verspüren. Der Rückgang der Nutzung mit steigendem Alter könnte auch durch den geringeren beruflichen Einsatz dieser Technologien erklärt werden, da ältere Befragte möglicherweise nicht mehr aktiv im Berufsleben stehen und KI weniger als relevant für ihren Alltag empfinden.

Abbildung 3-2: Nutzung von generativer KI – Demografische Verteilung

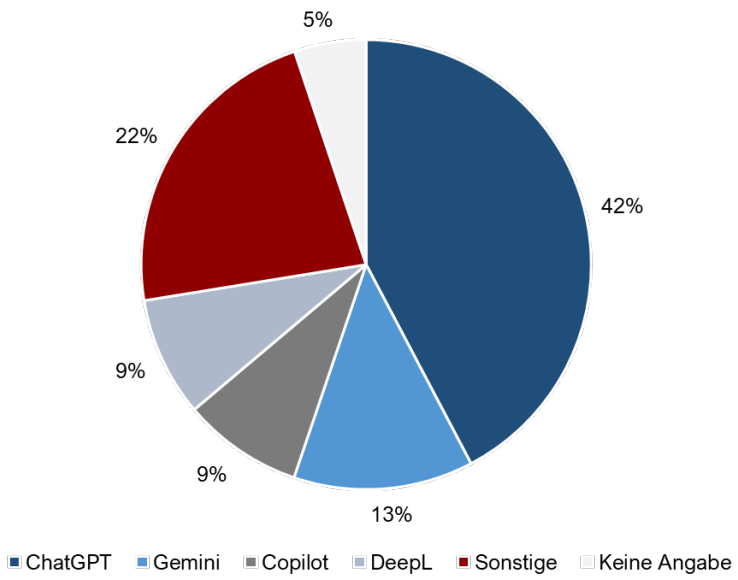


Quelle: Befragung des WIK. N=3.201.

Die Nutzer von generativer KI wurden im Anschluss daran dazu befragt, welche Tools sie im Einzelnen nutzen. Die Abbildung zeigt die Verteilung der Nutzung generativer KI-Modelle unter den Befragten. Im Durchschnitt geben Nutzer an, 1,5 Modelle zu verwenden. Die vier am häufigsten genutzten Modelle sind ChatGPT (42 %), gefolgt von Gemini (13 %), Copilot (9 %), und DeepL (9 %). 22 % der Befragten haben angegeben, andere Modelle zu nutzen, während 5 % keine Angabe gemacht haben.

12 European Institute for Gender Equality (2020).

Abbildung 3-3: Nutzung von generativer KI – Meistgenutzte Tools



Quelle: Befragung des WIK. N=1.414.

Die hohe Präferenz für ChatGPT deutet darauf hin, dass dieses Modell derzeit eine zentrale Rolle in der generativen KI-Nutzung einnimmt, möglicherweise aufgrund seiner Zugänglichkeit, Funktionalität oder Bekanntheit. Die Verteilung der restlichen Modelle zeigt eine Diversität in den Anwendungen, wobei spezialisierte Tools wie Copilot (vor allem in der Softwareentwicklung) und DeepL (für Übersetzungen) wichtige Nischen bedienen. Die Angabe von „Sonstige" unterstreicht zudem, dass es weitere Modelle gibt, die von spezifischen Nutzergruppen eingesetzt werden, jedoch weniger verbreitet sind.

3.2 Kenntnisstand und Einstellungen

Zur Messung des Wissens über und zur Anwendbarkeit von generativer Künstlicher Intelligenz (KI) wurden zwei Subskalen des Konzepts „Cognitive Learning" nach Ng et al. (2024) herangezogen und an den spezifischen Kontext dieser Studie angepasst. Die beiden Subskalen „Kennen und Verstehen" sowie „Anwenden, Bewerten und Gestalten" umfassen in Summe sechs Items und wurden an den Zweck dieser Studie angepasst.[13] Die Items wurden auf einer fünfstufigen Likert-Skala bewertet, die von „Stimme voll und ganz zu" bis „Stimme überhaupt nicht zu" reicht.

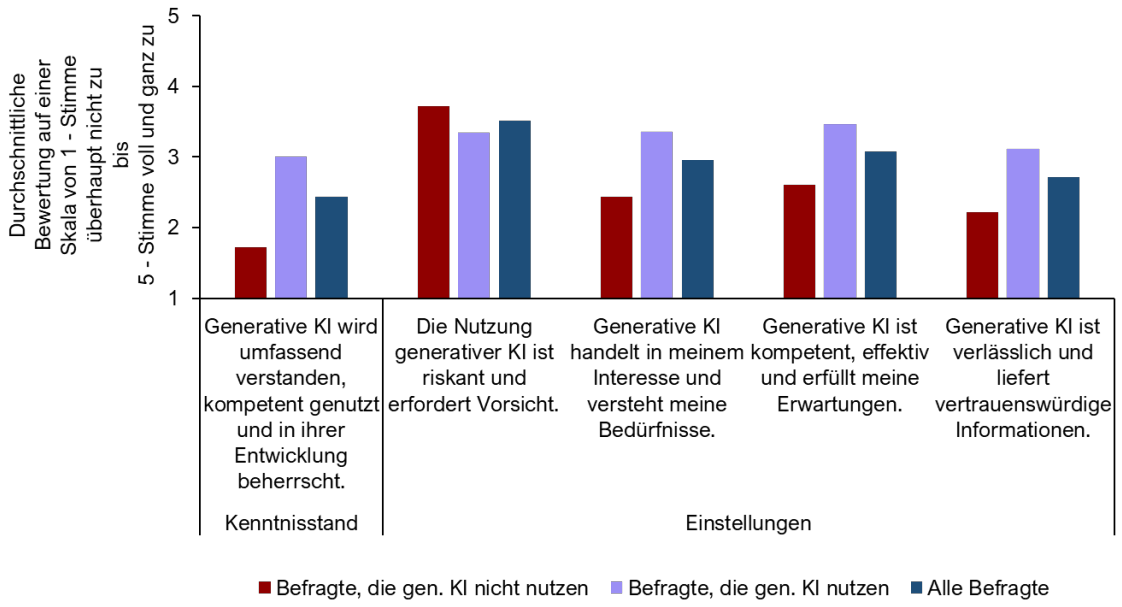
Die Befragten schätzten ihren Kenntnisstand[14] zu generativer KI insgesamt als gering ein. Es zeigte sich jedoch, dass Nutzer von generativer KI im Vergleich zu Nichtnutzern einen höheren Wissensstand angegeben haben. Dies ist ein Hinweis darauf, dass die praktische Anwendung von KI-Anwendungen mit einer Zunahme des selbst eingeschätzten Verständnisses und Wissens über KI einhergeht.

---

13 Ng, D.T.K. (2024).

14 Für die Bewertung des Kenntnisstands wurden die sechs Items von Ng et al. (2024) herangezogen: Subskala 1: Kennen und Verstehen: Ich weiß, was generative KI ist und kenne die Definition von generativer KI. | „Ich weiß, wie man generative KI-Anwendungen nutzt. | Ich kenne die Unterschiede verschiedener Konzepte, die für generative KI-Anwendungen verwendet werden (z. B. Deep Learning, maschinelles Lernen, etc.). Subskala 2: Anwenden, Bewerten und Gestalten: Ich kann KI-Anwendungen zur Problemlösung einsetzen. | Ich kann selbst generative KI- Anwendungen (z. B. Chatbots, Robotik) entwickeln, die Probleme lösen. | Ich kann generative KI-Anwendungen und -Konzepte für verschiedene Anwendungssituationen bewerten.

Abbildung 3-4: Wissenstand und Einstellungen



Quelle: Befragung des WIK. N=3.201.

Zur Untersuchung der Einstellungen von Individuen gegenüber generativer Künstlicher Intelligenz (KI) wurde die von Gulati et al. (2019) entwickelte Skala zur Human-Computer-Interaction (HCI) herangezogen.[15] Die Skala umfasst 12 Items, die auf vier Subskalen basieren und spezifische Dimensionen der Wahrnehmung und Bewertung von KI adressieren. „Wohlwollen/Unterstützung" bezieht sich auf die wahrgenommene Absicht der KI, dem Nutzer zu helfen.[16] Die Skala „Wahrnehmung Risiko" misst die Risikowahrnehmung und die Bedenken bei der Interaktion mit der KI.[17] „Kompetenz" bewertet, wie effektiv und kompetent die KI in ihrer Rolle wahrgenommen wird und „Vertrauen" schließlich bezieht sich auf das Vertrauen der Nutzer in die KI und die Zuverlässigkeit der von ihr gelieferten Informationen.[18] Die Items wurden auf einer fünfstufigen Likert-Skala bewertet, die von „Stimme voll und ganz zu" bis „Stimme überhaupt nicht zu" reichte.

Nutzer generativer KI zeigten insgesamt positivere Einstellungen gegenüber der Technologie als Nichtnutzer. Sie schätzten insbesondere die Benutzerfreundlichkeit und die Vorteile der KI höher ein als Nichtnutzer. Hinsichtlich der Wahrnehmung potenzieller Risiken und Gefahren gab es nur geringe Unterschiede zwischen Nutzern und Nichtnutzern. Beide Gruppen bewerteten Risiken ähnlich, was darauf hindeutet, dass die Risikowahrnehmung unabhängig von der Nutzungserfahrung relativ stabil ist.

---

15 Gulati, S. et al. (2019).

16 Dies wurde mit den folgenden drei Items gemessen: Ich glaube, dass generative KI in meinem besten Interesse handelt. | Ich glaube, dass generative KI ihr Bestes tut, um mir zu helfen. | Ich glaube, dass generative KI in der Lage ist, meine Bedürfnisse und Vorlieben zu verstehen.

17 Dies wurde mit den folgenden drei Items gemessen: Ich glaube, dass die Verwendung von generativer KI negative Folgen haben kann. | Ich habe das Gefühl, dass ich bei der Verwendung von generativer KI vorsichtig sein muss. | Es ist riskant, mit generativer KI zu interagieren.

18 Dies wurde mit den folgenden Items gemessen: Kompetenz: Ich glaube, dass generative KI kompetent und effektiv ist. | Ich denke, dass generative KI ihre Aufgaben sehr gut ausfüllt. | Ich glaube, dass generative KI alle Funktionen hat, die ich von ihr erwarte. Vertrauen: Wenn ich generative KI benutze, glaube ich, dass ich mich voll und ganz auf sie verlassen kann. | Ich kann mich immer auf generative KI und ihre Unterstützung verlassen. | Ich kann den Informationen vertrauen, die mir generativer KI liefert.

Die Ergebnisse legen nahe, dass die praktische Erfahrung mit generativer KI zu positiveren Einstellungen beiträgt, ohne jedoch die Risikowahrnehmung maßgeblich zu beeinflussen.

3.3 Bewertung des Einsatzes in verschiedenen Bereichen

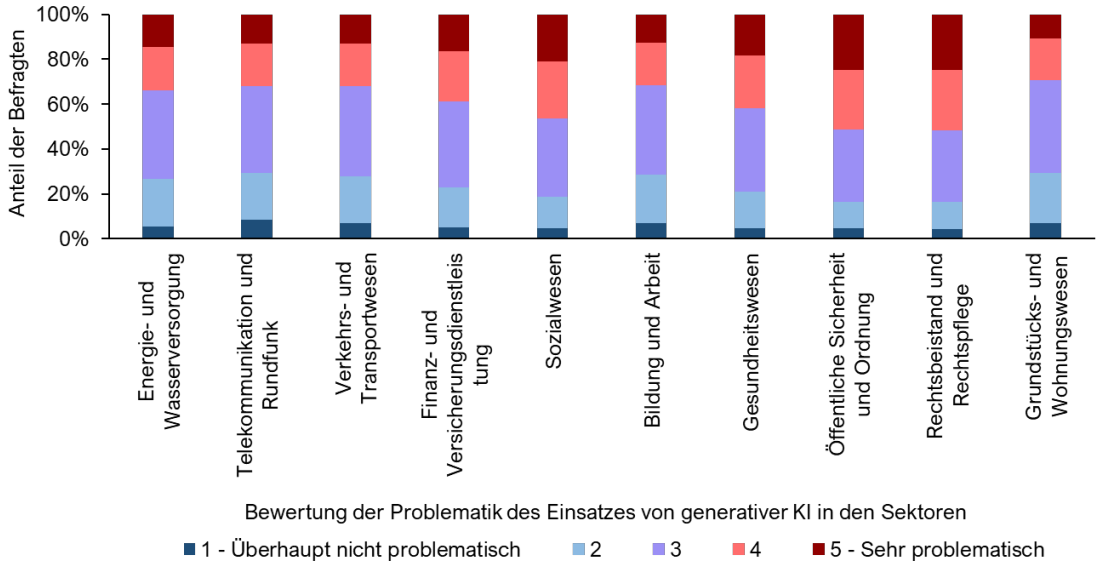
Im Anschluss wurde untersucht, welche Einstellungen die Befragten zum Einsatz generativer KI in verschiedenen Sektoren haben. Besonders kritisch sehen die Befragten den Einsatz in den Bereichen Sozialwesen, Gesundheitswesen, öffentliche Sicherheit und Ordnung sowie Rechtspflege. Diese Bereiche werden als besonders sensibel wahrgenommen, da hier oft hohe ethische und rechtliche Anforderungen bestehen und potenzielle Risiken schwerwiegende Folgen für Individuen und die Gesellschaft haben können. Auch King (2023) stellt in seiner Studie fest, dass die dort Befragten angeben, dass eine KI keine kritischen oder folgenschweren Entscheidungen treffen soll, insbesondere nicht in Situationen in denen es um Leben, Sicherheit und Freiheit geht.[19]

Eine Studie von KPMG bietet detailliertere Einsichten zur Bewertung von generativer KI im Arbeitsalltag. Der Einsatz von KI wird in verschiedenen Bereichen der Arbeitswelt unterschiedlich bewertet. Die Mehrheit akzeptiert KI in der Entscheidungsfindung im Management und bevorzugt sie gegenüber rein menschlichen Entscheidungen. Auch der Einsatz von KI zur Unterstützung und Automatisierung von organisationsbezogenen Aufgaben wie Sicherheitsüberwachung, administrativen Tätigkeiten oder Analysen wird positiv gesehen. Weniger Zustimmung gibt es jedoch, wenn KI organisatorische Entscheidungsprozesse beeinflusst. Besonders kritisch sehen die Befragten den Einsatz von KI im Personalmanagement, etwa zur Überwachung, Bewertung oder bei Rekrutierungsprozessen. Hingegen wird der Einsatz von KI zur Unterstützung und Leistungssteigerung von Mitarbeitenden, z. B. durch Feedback oder Entscheidungshilfen, mehrheitlich akzeptiert.[20]

---

19 King, S. (2023).
20 KPMG (2023).

Abbildung 3-5: Bewertung des Einsatzes von generativer KI in verschiedenen Sektoren



Quelle: Befragung des WIK. N=3.201.

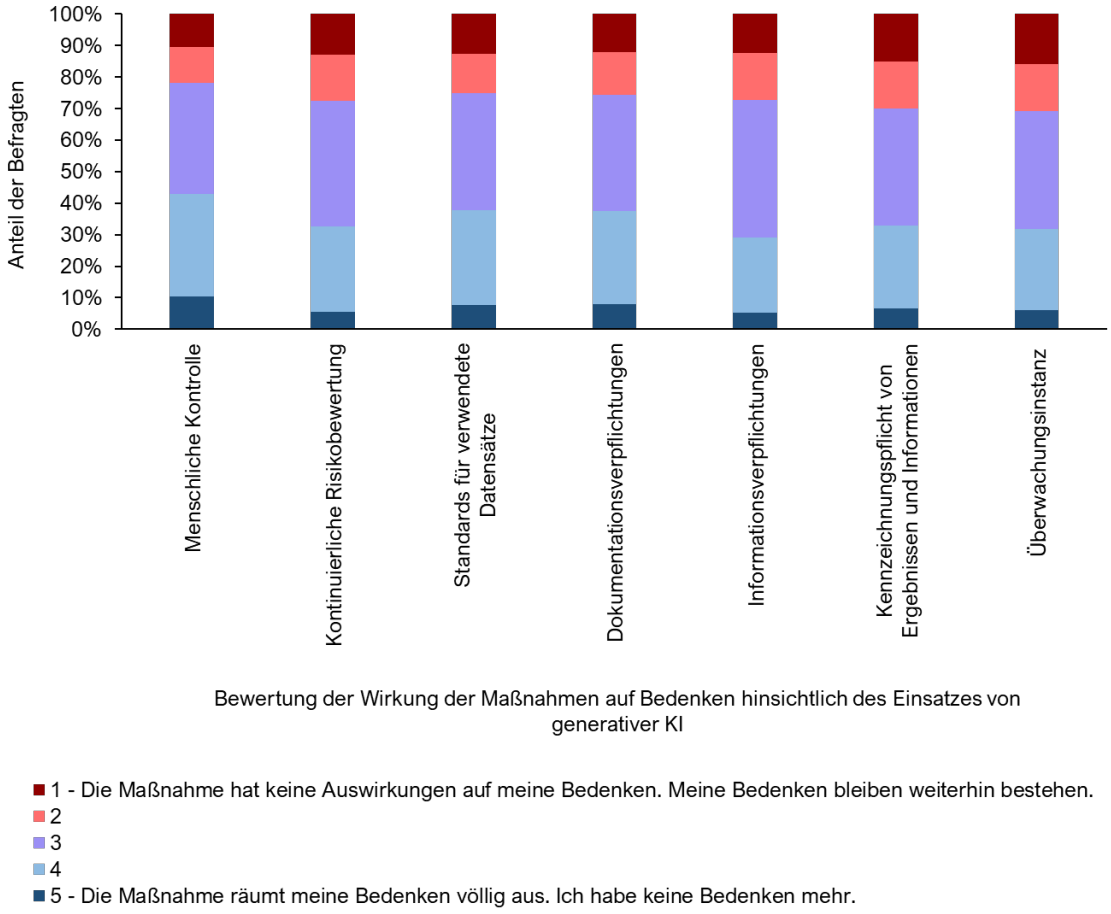
Lediglich 2 % der Befragten empfinden den Einsatz von generativer KI grundsätzlich und unabhängig von ihrem Anwendungsbereich als unproblematisch. Dies zeigt, dass die Mehrheit der Befragten differenziert zwischen den Anwendungsfeldern unterscheidet und die Risiken sowie möglichen Herausforderungen des KI-Einsatzes erkennt.

Anschließend wurden die Befragten gebeten, Einschätzungen dazu abzugeben, inwieweit spezifische Abhilfemaßnahmen ihre Bedenken hinsichtlich des Einsatzes von generativer KI ausräumen können. Grundlage hierfür bildeten die folgenden, grundsätzlich an den AI Act angelehnten Maßnahmen:

- Die Genauigkeit und Zuverlässigkeit der entsprechenden KI-Anwendungen wird durch menschliche Kontrolle sichergestellt. (angelehnt an Art. 14 AI Act)
- KI-Anwendungen unterliegen einer kontinuierlichen Risikobewertung. Dabei werden bekannte und vorhersehbare Risiken identifiziert und analysiert. Ziel ist es, ihr Auftreten durch geeignete Maßnahmen zu reduzieren oder zu kontrollieren. (angelehnt an Art. 9 AI Act)
- Die Datensätze, auf denen die KI-Anwendungen basieren, müssen hohe Qualitätsstandards einhalten, um falsche und diskriminierende Ergebnisse zu vermeiden. (angelehnt an Art. 10 AI Act)
- KI-Anwendungen dokumentieren sämtliche durchgeführten Prozesse und Ergebnisse, sodass eine lückenlose Nachvollziehbarkeit zu jedem Zeitpunkt gewährleistet ist. (angelehnt an Art. 12 AI Act)
- KI-Anwendungen stellen ausführliche und klar formulierte Unterlagen bereit, die alle erforderlichen Informationen über das System und seinen Zweck enthalten. Dabei werden die geltenden Standards zur Bereitstellung solcher Unterlagen eingehalten. (angelehnt an Art. 13 AI Act)
- Betroffene erhalten den Hinweis, dass eine Entscheidung von einer KI getroffen wurde bzw. die zur Verfügung gestellten Informationen von einer KI stammen. (angelehnt an Art. 50 AI Act)
- Die Überwachung der KI-Anwendungen erfolgt durch eine höhere behördliche Instanz, beispielsweise die der Europäischen Kommission. Die Einhaltung aller bisher genannten Maßnahmen wird zertifiziert.

Die Auswirkungen der Maßnahmen auf die Bedenken wurden auf einer Skala von 1 (keine Auswirkung auf meine Bedenken) bis 5 (meine Bedenken sind völlig ausgeräumt) bewertet.

Abbildung 3-6: Abhilfemaßnahmen



Quelle: Befragung des WIK. N=3.162.

Laut der Grafik haben die vorgeschlagenen Maßnahmen aus Sicht der Befragten tendenziell nur eine geringe Wirkung, um Bedenken vollständig auszuräumen. Ein großer Anteil der Befragten bewertet die Maßnahmen mit 1, 2 oder 3, was auf eine geringe bis moderate Wirksamkeit hindeutet. Nur ein kleiner Prozentsatz sieht die Maßnahmen als sehr wirksam (4 oder 5).

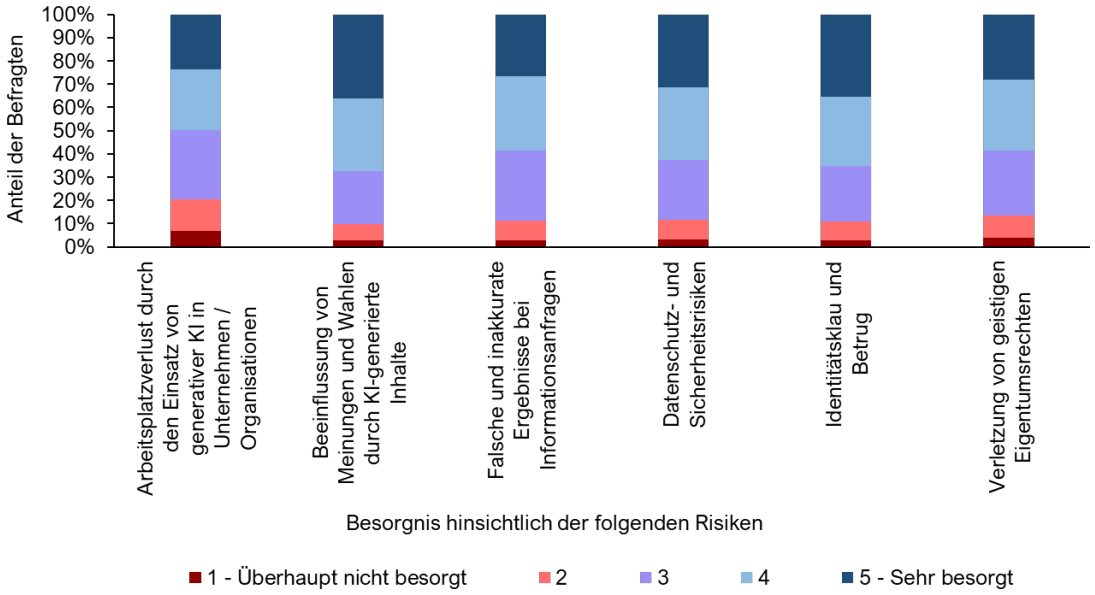
Lediglich die erste Maßnahme „Die Zuverlässigkeit und Zweckmäßigkeit der Anwendungen wird durch menschliche Kontrolle sichergestellt“ hat eine leicht höhere Bewertung im Vergleich zu anderen Maßnahmen, in Summe erhält sie aber immer noch viele niedrige Bewertungen. Auch in King (2023) sagen die Befragten aus, dass sie in weiteren Teilen menschlichen Entscheidungen mehr vertrauen als den Entscheidungen einer KI.[21]

Abschließend wurde die Wahrnehmung spezifischer Risiken, die durch den Einsatz von generativer Künstlicher Intelligenz (KI) entstehen können, erfragt. Die Befragten konnten den Umfang ihrer Besorgnis auf einer Skala von 1 (überhaupt nicht besorgt) bis 5 (sehr besorgt) angeben.

---

21 King, S. (2023).

Abbildung 3-7: Spezifische Risiken



Quelle: Befragung des WIK. N=3.201.

Der Arbeitsplatzverlust durch den Einsatz von generativer KI wird von den Befragten als das am wenigsten besorgniserregende Risiko bewertet. Die Mehrheit der Befragten (hoher Anteil in den Kategorien 1 und 2) scheint dies als ein eher unwahrscheinliches oder weniger relevantes Problem zu betrachten. Dies könnte darauf hindeuten, dass entweder ein Vertrauen in den Umgang mit Automatisierung besteht oder der Zusammenhang zwischen generativer KI und Arbeitsplatzverlust als weniger direkt wahrgenommen wird.

Die Beeinflussung von Meinungen und Wahlen durch KI-generierte Inhalte ist eines der Hauptanliegen der Befragten. Ein signifikanter Anteil bewertet dieses Risiko mit den höchsten Kategorien (4 und 5), was auf eine weitverbreitete Sorge vor Manipulation und gesellschaftlicher Destabilisierung hinweist. Ähnlich stark sind die Bedenken bezüglich Identitätsdiebstahl und Betrug, die ebenfalls häufig mit den höchsten Bewertungen versehen wurden. Diese Ergebnisse spiegeln eine ausgeprägte Skepsis gegenüber den Sicherheits- und Datenschutzrisiken wider, die mit generativer KI verbunden sind.

Falsche und inkorrekte Ergebnisse bei Informationsanfragen sowie Datenschutz- und Sicherheitsrisiken befinden sich im mittleren Bereich der Risikowahrnehmung. Die Verletzung von geistigen Eigentumsrechten ist ebenfalls ein relevantes Thema, zeigt aber im Vergleich zu den beiden Hauptsorgen (Manipulation und Betrug) eine etwas geringere Dringlichkeit.

Die Ergebnisse zeigen, dass die Verbraucher ihre Bedenken in erster Linie auf direkte Gefahren für die Gesellschaft (z. B. Manipulation von Meinungen) und die persönliche Sicherheit (z. B. Identitätsdiebstahl) konzentrieren. Die hohe Besorgnis in Bereichen wie Wahlmanipulation und Betrug deutet auf ein ausgeprägtes Bewusstsein für potenzielle Missbrauchsmöglichkeiten generativer KI hin. Gleichzeitig wird Vertrauen in die Handhabbarkeit technischer Herausforderungen wie Arbeitsplatzverlust oder ungenaue Ergebnisse sichtbar.

4 Wahrnehmung und Verbreitung von Informationen im Internet

Die Fähigkeit einer generativen KI, Informationen schnell und verständlich darzustellen, kann dazu genutzt werden, Falschinformationen in größerem Umfang zu generieren und zu verbreiten.[22] Insbesondere Informationen, die von erheblichem öffentlichen Interesse sind und das Verhalten sowie die Denkweisen von Individuen beeinflussen können, bergen gesamtgesellschaftliche Risiken, wenn sie manipuliert oder verfälscht werden. Darüber hinaus zeigen Studien, dass Menschen oft Schwierigkeiten haben, KI-generierte oder manipulierte Inhalte von nicht KI-generierten oder nicht manipulierten Inhalten zu unterscheiden.[23]

Der AI Act sieht unter anderem vor, dass „KI-generierte Texte, die mit dem Ziel veröffentlicht werden, die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse zu informieren, als künstlich erzeugt gekennzeichnet werden."[24] In den Leitlinien der Kommission für VLOPs und VLOSEs zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 des DSA, wird ebenfalls empfohlen KI-generierte Inhalte zu kennzeichnen.[25]

Im Rahmen dieser Studie wurde unter anderem ein Experiment durchgeführt, um den Einfluss eines KI-Labels auf die Wahrnehmung von Informationen und die Wahrscheinlichkeit ihrer Verbreitung zu untersuchen. Der methodische Ansatz und die Ergebnisse werden in den nachfolgenden Abschnitten näher erläutert.

4.1 Methodisches Vorgehen

Für das Experiment wurden die Befragten gebeten, jeweils zehn Beiträge hinsichtlich der von ihnen wahrgenommen Korrektheit der enthaltenen Information (auf einer Skala von „völlig falsch" [1] bis „völlig richtig" [6]) zu bewerten und anzugeben, inwiefern sie bereit wären, den Beitrag in ihren eigenen sozialen Netzwerken zu teilen (auf einer Skala von „sehr unwahrscheinlich" [1] bis „sehr wahrscheinlich" [6]). Die erste Angabe dient dazu, die Wahrnehmung der Beiträge zu bewerten, und die zweite dazu, das Verbreitungspotenzial der einzelnen Beiträge zu ermitteln.

Die Informationen in den Beiträgen unterscheiden sich ihrem Wahrheitsgehalt. Die Hälfte der zehn Beiträge gibt wahre Informationen wieder, die andere Hälfte besteht aus falschen Informationen.

Die Informationen beruhen auf bestehenden Meldungen, die online recherchiert wurden. Die wahren Informationen stammen aus Meldungen, die sich auf den Online-Präsenzen von Zeitungen wie z. B. „Der Spiegel"[26] und „Berliner Zeitung"[27] finden ließen. Informationen zu belegten Falschmeldungen wurden Faktencheck-Seiten wie dem „Correctiv"[28] entnommen. Bei der Auswahl der Meldungen wurde darauf geachtet, dass sie eine Reihe von unterschiedlichen Themengebieten abdecken, wie Gesundheit, Umwelt, Politik usw. Anschließend wurden aus allen zehn Meldungen mithilfe von ChatGPT 4o Nachrichten- und Social-Media-Beiträge erstellt. Zu diesem Zweck wurde für jede Nachricht eine neue

---

22 Chesney, R./ Citron, D. K. (2019).
23 Spitale, G. et al. (2023); Frank, J. et al. (2023).
24 Europäische Kommission (2024).
25 Verordnung (EU) 2022/2065 des europäischen Parlaments und des Rates vom 19. Oktober 2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste), Amtsblatt der Europäischen Union, L 277/1.
26 https://www.spiegel.de/.
27 https://www.berliner-zeitung.de/
28 https://correctiv.org/

Konversation gestartet, um eine Interferenz mit vorherigen Meldungen zu vermeiden. Die jeweilige Meldung wurde dieser Konversation in Form eines Word-Dokuments beigefügt. Anschließend wurden die Teilnehmer gebeten, die relevanten Informationen aus dem beigefügten Dokument zu extrahieren. ChatGPT wurde dann gebeten, die Rolle eines durchschnittlichen deutschen Social-Media-Nutzers bzw. Journalisten einzunehmen und auf Basis der extrahierten relevanten Informationen einen repräsentativen Social-Media-Beitrag bzw. Nachrichtenartikel unter 280 Zeichen zu verfassen.

Die verwendeten Anweisungen basieren auf der Chain-of-Thought-Methode nach Jiang et al. (2024), die diese Methode zur Erstellung von Nachrichtenartikeln anwenden.[29] Bashardoust et al. (2024) und Altay et al. (2024) verwenden ebenfalls einen ähnlichen Ansatz, um KI-generierte Inhalte für ihre Studien zu erstellen.[30]

Die daraus resultierenden Social-Media-Beiträge und Nachrichtenbeiträge wurden ohne weitere Textänderungen in den Fragebogen übernommen. Sie wurden lediglich manuell in ein Design für Nachrichtenbeiträge und ein Design für soziale Medien visuell eingefasst. In beiden Fällen wurde versucht, das Design so neutral wie möglich zu halten, um keine zusätzlichen Markenwirkungen zu messen. Dies bedeutet zudem, dass alle Beiträge, die den Befragten zur Verfügung gestellt wurden, als von einer KI generiert angesehen werden können.

Neben der Darstellungsform wurde auch das Auftreten eines KI-Labels variiert. So sah ein Teil der Befragten entweder einen Social-Media-Beitrag oder einen Nachrichtenbeitrag mit Label, der andere Teil ohne Label. Bei dem Label wurde eine Kombination aus einem Logo und dem Zusatz „Erstellt mit KI“ verwendet, angelehnt an Altay et al (2024).[31] Die Platzierung des Labels erfolgte sowohl bei den Social-Media-Beiträgen als auch den Nachrichtenbeiträgen jeweils rechts unten.

Daraus ergibt sich ein 2x2x2-Design mit den Zwischensubjektfaktoren „Darstellungsform“ (Nachrichtenbeitrag vs. Social-Media-Beitrag) und „Label“ (mit Label vs. ohne Label) sowie dem Innersubjektfaktor „Wahrheitsgehalt der Information“ (wahr vs. falsch).

Zusätzlich zum Experiment wurden einige Fragen zum Auffinden und zur wahrgenommenen Beeinflussung durch Falschinformationen in der Vergangenheit gestellt. Darüber hinaus gaben die Befragten eine Einschätzung dazu ab, ob sie KI-generierte Inhalte erkennen würden.

Die Ergebnisse des Experiments und der zusätzlichen Fragen werden in den folgenden zwei Abschnitten dargestellt.

4.2 Erkennen von Falschinformationen und KI-generierten Inhalten

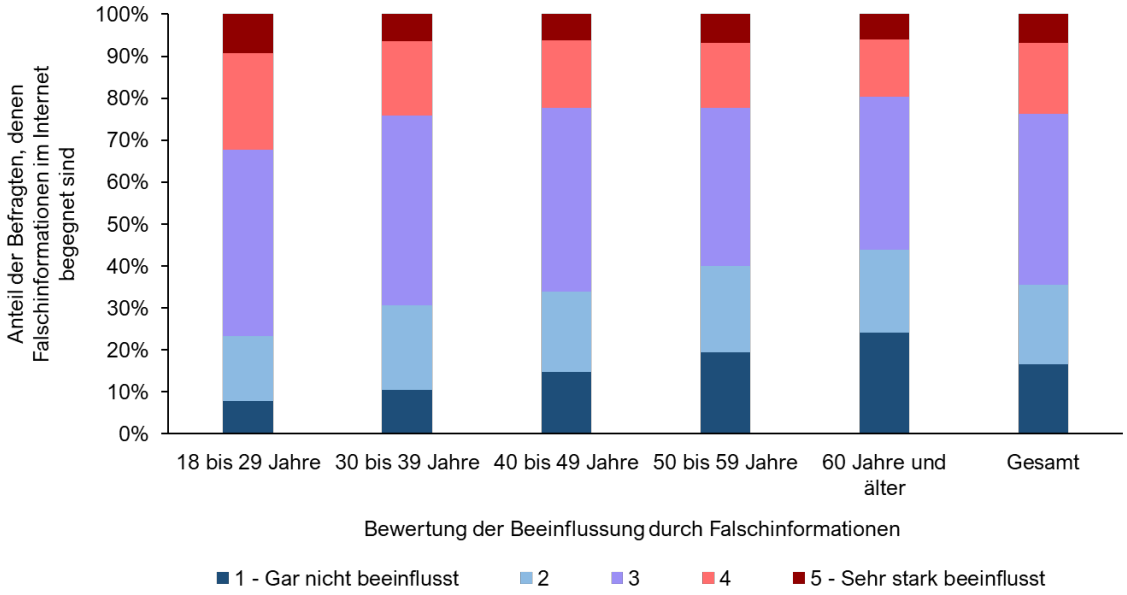
Etwa 78 % der Befragten gaben an, im vergangenen Jahr mindestens einmal auf Informationen oder Inhalte in Text-, Bild-, Audio- oder Videoform hereingefallen zu sein, die sie zunächst für korrekt hielten, sich jedoch später als falsch herausstellten.[32] Im Durchschnitt traf dies etwa auf 20 Informationen bzw. Inhalte zu.[33] Von den Befragten gaben 24 % an, dass die Informationen, die sie für falsch hielten, einen erheblichen Einfluss auf ihr Denken und ihre Entscheidungen hatten. Weitere 41 % gaben an, dass sie

---

29 Jiang, B. et al. (2024).
30 Bashardoust, A. et al. (2024); Altay,S./ Gilardi, F. (2024).
31 Altay,S./ Gilardi, F. (2024).
32 N=3.201.
33 N=3.180. Nach Bereinigung von Ausreißern anhand der Perzentil-Methode (oberes und unteres 1%-Perzentil). Methode wurde verwendet, da Daten asymmetrisch und nicht normal verteilt sind

zumindest teilweise beeinflusst wurden. Nur ein Drittel erklärte, eine geringe Beeinflussung erfahren zu haben (siehe Abbildung 4-1). Dabei verdeutlichen die Daten, dass sich insbesondere die jüngeren genauso wie die weiblichen Befragten stärker beeinflussen ließen.

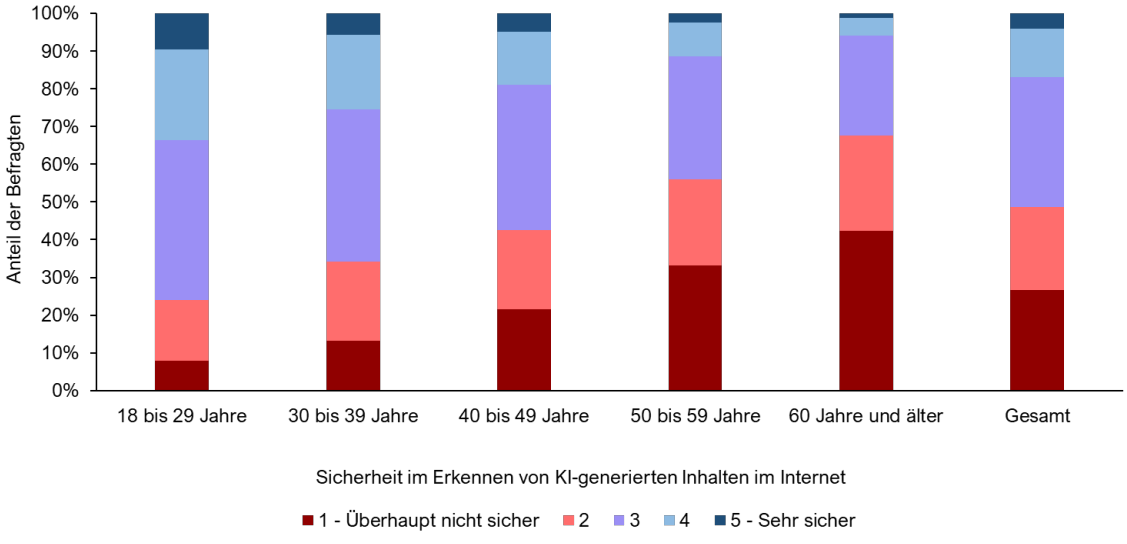
Abbildung 4-1: Einfluss von Falschinformationen auf Entscheidungen und Denkweisen von Individuen



Quelle: Befragung des WIK. N=2.486.

Zudem gaben im Rahmen dieser Befragung knapp die Hälfte der Befragten zu, KI-generierte Inhalte nicht erkennen zu können (siehe Abbildung 4-2). Insbesondere ältere Befragte und Frauen äußerten, dass sie sich nicht sicher sind, ob sie KI-generierte Inhalte richtig identifizieren können.

Abbildung 4-2: Sicherheit im Erkennen von KI-generierten Inhalten



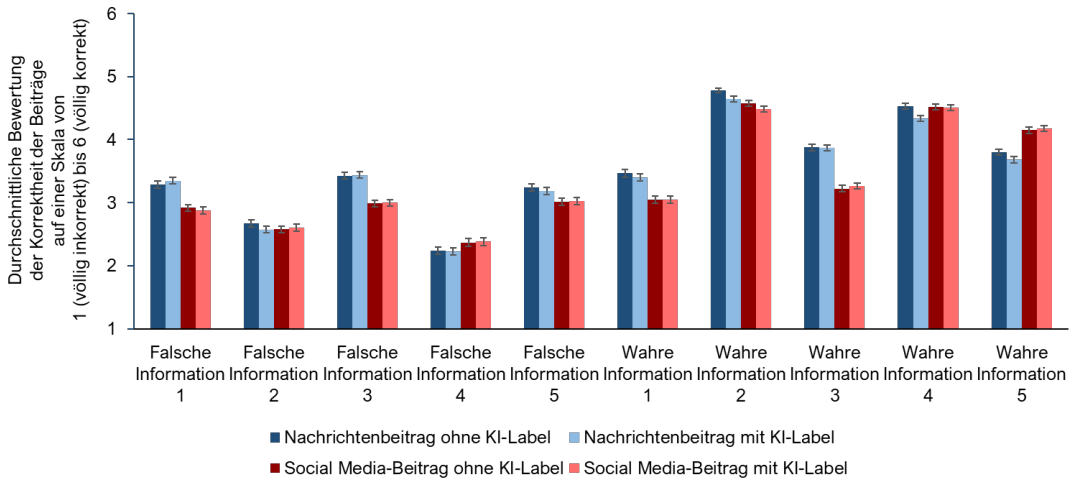
Quelle: Befragung des WIK. N=3.201.

4.3 Wirkung eines KI-Labels

Alle Befragten wurden zufällig einer der vier oben beschriebenen experimentellen Gruppen zugeordnet und sahen entweder zehn „Nachrichtenbeiträge ohne KI-Label“, „Nachrichtenbeiträge mit KI-Label“, „Social-Media-Beiträge ohne KI-Label“ oder „Social-Media-Beiträge mit KI-Label“. Für jeden der Beiträge bewerteten die Befragten die Korrektheit der Information sowie die Wahrscheinlichkeit, diese zu teilen. Die Stichprobengröße in den einzelnen experimentellen Gruppen variierte zwischen 697 Teilnehmern in der Gruppe „Social-Media-Beitrag mit KI-Label“ und 711 Teilnehmern in der Gruppe „Nachrichtenbeitrag ohne KI-Label“.[34]

Die Abbildung *4-3* visualisiert die durchschnittliche Bewertung der Korrektheit innerhalb der jeweiligen vier Gruppen für jede der fünf wahren und fünf falschen Informationen. Generell zeigt sich, dass die Beiträge mit den wahren Informationen durchgehend als korrekter bewertet werden als die Beiträge, die falsche Information beinhalten. Dies trifft unabhängig von der Darstellungsform oder dem Label zu. Darüber hinaus werden Social-Media-Beiträge tendenziell als weniger korrekt bewertet als Nachrichtenbeiträge. Das Label hingegen wirkt sich kaum merkbar auf die Bewertung der Korrektheit aus und spielt möglicherweise keine wesentliche Rolle bei der Wahrnehmung der Beiträge.

Abbildung 4-3: Mittelwerte und Standardabweichung – Bewertung der Korrektheit



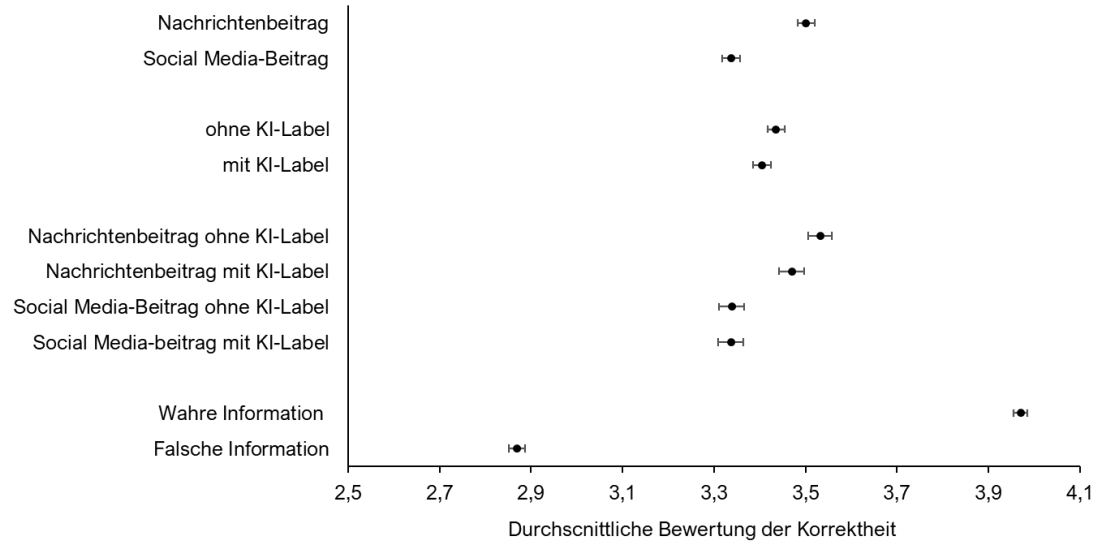
Quelle: Befragung des WIK. N=2.828.

Sehr deutlich werden die beobachtbaren Tendenzen in der nachfolgenden Abbildung. Hier zeigt sich, dass die durchschnittliche Bewertung zwischen den wahren und falschen Informationen um 1,1 Punkte auf der verwendeten Skala abweicht. Bei der Darstellungsform werden Nachrichtenbeiträge hinsichtlich

34 Von der repräsentativen Gesamtstichprobe von 3.201 Befragten wurden für die Auswertung des Experiments die Antworten von 2.828 Teilnehmern berücksichtigt. Befragte, die bei zwei Kontrollfragen angaben, die Fragen zu den Beiträgen zufällig beantwortet oder während der Befragung im Internet recherchiert zu haben, wurden bei der Analyse ausgeschlossen.

der wahrgenommenen Korrektheit um etwa 0,2 Punkte besser bewertet. Beim Label ist kaum ein Unterschied vorhanden (0,03 Punkte).

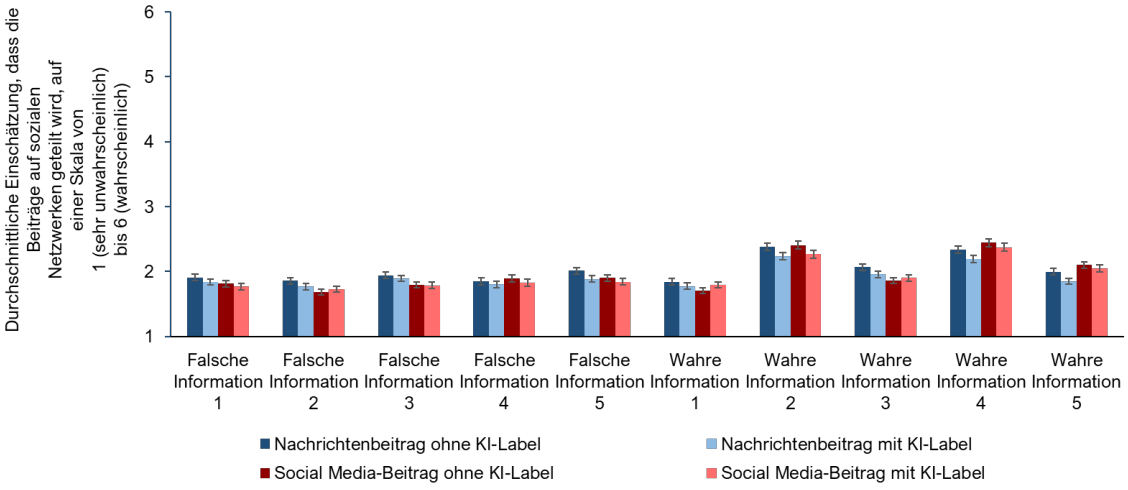
Abbildung 4-4: Durchschnittliche Bewertung der Korrektheit



Quelle: Befragung des WIK. N=2.828.

Abbildung *4-5* veranschaulicht die Wahrscheinlichkeit, dass die einzelnen Beiträge geteilt werden. Die Wahrscheinlichkeit, dass die präsentierten Beiträge geteilt werden, ist grundsätzlich sehr gering. Insgesamt zeigt sich jedoch, dass die Befragten dazu tendieren, Beiträge mit wahren Informationen wahrscheinlicher zu teilen als Beiträge mit falschen Informationen. Dieses gilt unabhängig von der Darstellungsform oder dem Label. Im Gegensatz zur wahrgenommenen Korrektheit erscheint der Einfluss sowohl der Darstellungsform als auch des Labels in diesem Zusammenhang gering.

Abbildung 4-5: Mittelwerte und Standardabweichung – Wahrscheinlichkeit des Teilens

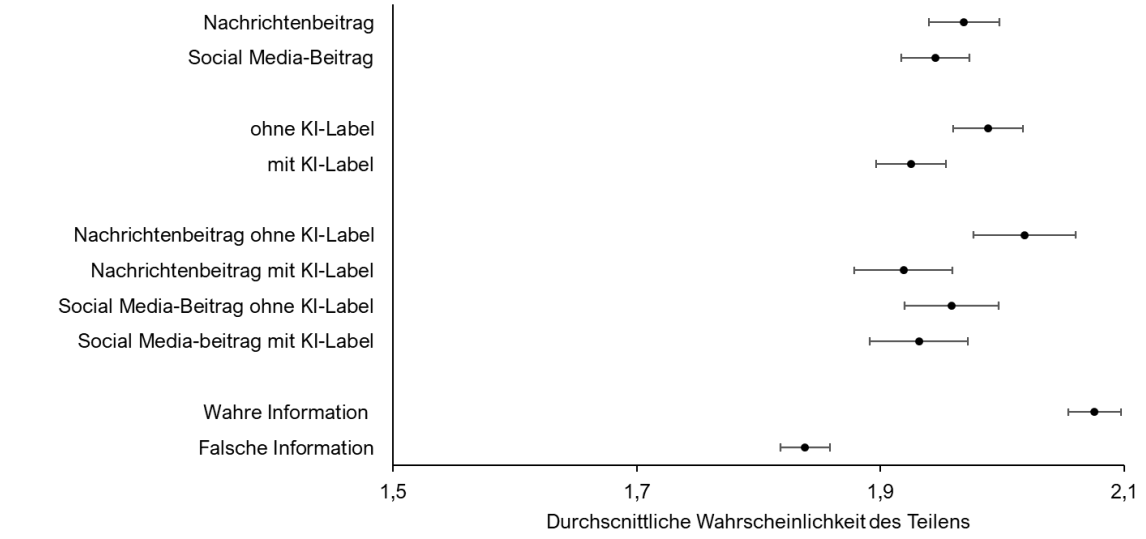


Quelle: Befragung des WIK. N=2.828.

Die durchschnittlichen Bewertungen bestätigen diese Beobachtungen. Der Unterschied in der Wahrscheinlichkeit, die Beiträge zu teilen, beträgt zwischen wahrer und falscher Information auf der

verwendeten 6-Punkt-Skala 0,2 Punkte. Die Unterschiede in der Darstellungsform und im Label sind wesentlich geringer und liegen bei 0,02 bzw. 0,06 Punkten.

Abbildung 4-6: Durchschnittliche Bewertung der Wahrscheinlichkeit des Teilens



Quelle: Befragung des WIK. N=2.828.

Zusammenfassend zeigen die Ergebnisse, dass insbesondere der Wahrheitsgehalt der Information die Bewertung der Korrektheit der Beiträge und die Wahrscheinlichkeit diese zu teilen beeinflussen kann. Die Darstellungsform beeinflusst ebenfalls spürbar die wahrgenommene Korrektheit eines Beitrags. Die Unterschiede in Hinblick auf das Label sind nur sehr gering.

Ein ähnliches Phänomen beobachten auch Tandoc Jr. et al. (2020). Im Rahmen ihres Experiments finden sich keine Hinweise auf Unterschiede in der wahrgenommenen Glaubwürdigkeit von Nachrichten, wenn als Quelle Algorithmen, Menschen oder eine Kombination aus beidem angegeben werden.[35]

Es gibt jedoch Studien, die eine Wirkung des Labels feststellen. So zeigt Waddell (2017), dass Nachrichten, deren Erstellung einer Maschine zugeschrieben wird, als weniger glaubwürdig wahrgenommen werden als Nachrichten, die einem Menschen zugeschrieben werden.[36] Zu einer ähnlichen Beobachtung kommen auch Graefe et al. (2016). Die Autoren folgern aus ihrem Experiment, dass Artikel stets als glaubwürdiger bewertet werden, wenn sie als von einem Journalisten verfasst deklariert sind und nicht als von einem Computer geschrieben. Sie weisen jedoch auch darauf hin, dass die Unterschiede sehr gering und daher statistisch nicht signifikant sind.[37] Des Weiteren stellen Toff & Simon (2023) im Rahmen des von ihnen durchgeführten Experiments fest, dass Nachrichtenartikel, die als KI-generiert gekennzeichnet sind, im Durchschnitt als weniger vertrauenswürdig wahrgenommen werden als solche ohne diese Kennzeichnung. Sie konnten jedoch nicht belegen, dass die Kennzeichnung auch einen Effekt auf die wahrgenommene Richtigkeit der Nachricht hat.[38]

---

35 Tandoc Jr, E. C. et al. (2020).
36 Waddell, T. F. (2017).
37 Graefe, A. (2018).
38 Toff, B./ Simon, F. M. (2023).

Gänzlich konträre Ergebnisse erzielt die Studie von Wu (2019). Der Autor zeigt, dass automatisiert erstellte Inhalte glaubwürdiger und weniger voreingenommen wahrgenommen werden als menschlich verfasste Nachrichten.[39] Hofeditz et al. (2021) stellten ebenfalls fest, dass Nachrichten, bei denen der Einsatz von KI offengelegt wird, tendenziell als glaubwürdiger wahrgenommen werden als solche, bei denen dies nicht der Fall ist. Dieser Unterschied ist jedoch statistisch nicht signifikant.[40] Wölker et al. (2018) zeigen in ihrer Studie, dass die Wahrnehmung der Glaubwürdigkeit von automatisiert verfassten Nachrichten höher ist als bei Nachrichten, die von Menschen verfasst sind. Dieses Ergebnis ist jedoch vom Inhalt der Nachricht abhängig. Die Autoren belegen den Effekt für Sportartikel aber nicht für Finanzartikel.[41]

Einige aktuelle Studien haben neben der Kennzeichnung von KI-generierten Nachrichten auch den Wahrheitsgehalt der Nachrichten berücksichtigt. In diesem Forschungsstrang zeigen sich ebenfalls Unterschiede in den Ergebnissen. Die Arbeiten von Bashardoust et al. (2024), Altay und Gilardi (2024) sowie Longoni et al. (2022) verdeutlichen, dass eine KI-Kennzeichnung die Wahrnehmung von Nachrichten negativ beeinflussen kann. Bashardoust et al. (2024) untersuchen speziell Falschinformationen zur COVID-19-Pandemie und zeigen, dass KI-generierte Falschinformation als weniger korrekt wahrgenommen werden als solche, die von Menschen erstellt worden sind.[42] In der umfassender angelegten Studie von Altay und Gilardi (2024) wird deutlich, dass die Teilnehmer zwar „KI-generiert“ nicht automatisch mit „falsch“ gleichsetzen, die Kennzeichnung von Schlagzeilen als KI-generiert jedoch ihre wahrgenommene Genauigkeit mindert. Dies gilt unabhängig davon, ob die Schlagzeilen tatsächlich wahr oder falsch sind oder ob sie von einem Menschen oder einer KI erstellt wurden.[43] Ähnlich weisen Longoni et al. (2022) darauf hin, dass Nachrichten, die von KI verfasst wurden, im Vergleich zu von Menschen geschriebenen Nachrichten als weniger akkurat eingeschätzt werden. Zudem zeigen ihre Ergebnisse, dass Menschen dazu neigen, wahre KI-generierte Nachrichten fälschlicherweise als inkorrekt einzustufen, während sie falsche KI-generierte Nachrichten korrekterweise als inkorrekt bewerten.[44] Die Studien von Altay und Gilardi (2024) und Longoni et al. (2022) zeigen somit auch, dass ein Label nicht nur auf tatsächlich falsche Informationen wirkt, sondern auch die Wahrnehmung von tatsächlich richtigen Informationen beeinflussen kann. Altay und Gilardi (2024) schlagen daher vor, dass die Kennzeichnung von KI-generierten Inhalten mit Vorsicht angegangen werden sollte, um unbeabsichtigte negative Auswirkungen auf harmlose oder sogar nützliche KI-generierte Inhalte zu vermeiden.[45]

Ein etwas anderes Ergebnis erzielt die Studie von Spitale et al. (2023). In ihrem Experiment erkennen die Teilnehmer nutzergenerierte falsche Beiträge auf der Social-Media-Plattform Twitter besser als falsche Beiträge, die von GPT-3 erstellt wurden. Die Autoren folgern daraus, dass von GPT-3 erstellte Falschinformationen ein höheres Täuschungspotenzial besitzen als nutzergenerierte Falschinformationen.[46]

---

39 Wu, Y. (2019).
40 Hofeditz, L. (2021).
41 Wölker, A./ Powell, T. E. (2018).
42 Bashardoust, A. et al. (2024).
43 Altay,S.// Gilardi, F. (2024).
44 Longoni, C. et al. (2022).
45 Altay,S.// Gilardi, F. (2024).
46 Spitale, G. et al. (2023)

Hinsichtlich der Verbreitung von Informationen, kommen Forscher ebenfalls zu unterschiedlichen Ergebnissen zum Einfluss einer Kennzeichnung. Während Altay und Gilardi (2024) feststellen, dass ein KI-Label die Intention zu teilen reduziert, finden Bashardoust et al. (2024) keinen Einfluss.[47]

47 Bashardoust, A. et al. (2024); Altay,S. // Gilardi, F. (2024).

5 Schlussbetrachtung

Die vorliegende Untersuchung fokussiert auf zwei Aspekte generativer KI: (i) die Einstellung zu und Nutzung von generative KI durch Verbraucher sowie (ii) die Wahrnehmung und Verbreitung von KI-generieten Inhalten.

Einstellung zu und Nutzung von generativer KI

Die Nutzung generativer KI ist derzeit noch stark heterogen und trotz ihres großen Potenzials ist ihre Adaption aktuell noch nicht weit verbreitet. Die Ergebnisse zeigen, dass generative KI von knapp der Hälfte der Verbraucher in Deutschland nicht genutzt wird, weder im beruflichen noch im privaten Kontext. Der allgemeine Kenntnisstand zur generativen KI wird entsprechend insgesamt als eher gering eingeschätzt, wobei Nutzer der Technologie ihren Wissensstand häufig höher einschätzen als Nichtnutzer. Nutzer generativer KI zeigen insgesamt positivere Einstellungen gegenüber der KI im Vergleich zu denjenigen, die sie nicht verwenden.

Sektoren und Anwendungsfälle, die direkt mit Verbrauchern in Kontakt stehen, werden tendenziell kritischer bewertet als solche mit eher gesellschaftlichen Auswirkungen. Die Bedenken der Verbraucher hinsichtlich generativer KI lassen sich aktuell nicht vollständig ausräumen. Die vorgeschlagenen Maßnahmen zur Minderung dieser Bedenken haben in der Befragung nur eine geringe Wirkung gezeigt.

Einfluss einer KI-Labels auf die Wahrnehmung und Verbreitung KI-generierter Inhalte

In Rahmen dieser Studie wird festgestellt, dass die wahrgenommene Korrektheit eines Beitrags und die Wahrscheinlichkeit, diesen weiterzuverbreiten in erster Linie vom Wahrheitsgehalt der dargestellten Information abzuhängen scheint. Ein weiterer entscheidender Faktor für die Bewertung der Korrektheit ist die Darstellungsform. So werden Social-Media-Beiträge womöglich mit mehr Skepsis betrachtet und erhalten niedrigere Bewertungen im Vergleich zu den entsprechenden Nachrichtenbeiträgen. Der tendenziell geringe bis kaum vorhandene Effekt eines Labels auf die Verbreitung und/oder Wahrnehmung von Beiträgen wird von einigen Studien geteilt. Andere Studien berichten hingegen von negativen oder sogar positive Auswirkungen. Im letzteren Fall werden computergenerierte, KI-generierte oder automatisierte Nachrichten sogar als potenziell glaubwürdiger oder genauer angesehen als solche Nachrichten, die menschlichen Autoren zugeschrieben werden.

Darüber hinaus zeigen einige Studien, dass die Wahrnehmung und Verbreitung wahrer Informationen durch ein KI-Label negativ beeinflusst werden kann.

Der AI Act und die Leitlinien der Kommission für VLOPs und VLOSEs zur Minderung systemischer Risiken bei Wahlprozessen gemäß DSA zielen auf eine Kennzeichnung von KI-generierten Inhalten ab. Die verschiedenen Auswirkungen sollten bei der Bewertung der Implikationen einer Kennzeichnung von KI-generierten Inhalten berücksichtigt werden.

Literaturverzeichnis

Altay, S.; & Gilardi, F. (10 2024). People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation. *PNAS Nexus*, *3*(10), pgae403. Online verfügbar unter: https://doi.org/10.1093/pnasnexus/pgae403 [Letzter Abruf 18.12.2024].

Araujo, T.; Helberger, N.; Kruikemeier, S.; & de Vreese, C. H.. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence, *AI & Society*, 35(3), p. 611-623. Online verfügbar unter: https://doi.org/10.1007/s00146-019-00931-w [Letzter Abruf 18.12.2024].

Bashardoust, A.; Feuerriegel, S.; & Shrestha, Y. R. (2024). Comparing the willingness to share for human-generated vs. AI-generated fake news. *arXiv*. Online verfügbar unter: https://doi.org/10.48550/arXiv.2402.07395 [Letzter Abruf 18.12.2024].

Bontridder, N.; & Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy*, *3*, e32. Online verfügbar unter: https://doi.org/10.1017/dap.2021.20 [Letzter Abruf 18.12.2024].

Brüns, J.D.; & Meißner, M. (2024): Do you create your content yourself? Using generative artificial intelligence for social media content creation diminishes perceived brand authenticity, *Journal on Retailing and Consumer Services*, Online verfügbar unter: https://doi.org/10.1016/j.jretconser.2024.103790 [Letzter Abruf 10.12.2024].

Capgemini (2023): Why consumers love generative AI, Online verfügbar unter: https://prod.ucwe.capgemini.com/wp-content/uploads/2023/05/Final-Web-Version-Report-Creative-Gen-AI.pdf [Letzter Abruf 10.12.2024].

Chesney, R.; & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1753–1820, Online verfügbar unter: https://lawcat.berkeley.edu/record/1136469?v=pdf [Letzter Abruf 18.12.2024].

Deloitte (2023): The Generative AI Dossier- A seleciton of high-impact usecases across six major industries, Online verfügbar unter: https://www2.deloitte.com/us/en/pages/consulting/articles/gen-ai-use-cases.html [Letzter Abruf 10.12.2024).

European Institute for Gender Equality (2020): Gender Equality Index 2020: Digitalisation and the future of work, Online verfügbar unter: https://eige.europa.eu/publications-resources/toolkits-guides/gender-equality-index-2020-report/gendered-patterns-use-new-technologies?language_content_entity=en [Letzter Abruf 18.12.2024].

Europäische Kommission. (2024). KI-Gesetz. Online verfügbar unter: https://digital-strategy.ec.europa.eu/de/policies/regulatory-framework-ai [Letzter Abruf 18.12.2024].

Europäisches Parlament; & Europäischer Rat (2024): Verordnung (EU) 2024/1689 vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz, „AI Act“, Online verfügbar unter: https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L_202401689 [Letzter Abruf 10.12.2024].

Frank, J.; Herbert, F.; Ricker, J.; Schönherr, L.; Eisenhofer, T.; Fischer, A.; Dürmuth, M.; & Holz, T. (2023). A Representative Study on Human Detection of Artificially Generated Media Across Countries. *arXiv*. Online verfügbar unter: https://doi.org/10.48550/arXiv.2312.05976 [Letzter Abruf 18.12.2024].

Gillespie, N.; Lockey, S.; Curtis, C.; Pool, J.; & Akbari, A. (2023). Trust in Artificial Intelligence: A Global Study. The University of Queensland and KPMG Australia. Online verfügbar unter: https://doi.org/10.14264/00d3c94 [Letzter Abruf 10.12.2024].

Graefe, A.; Haim, M.; Haarmann, B.; & Brosius, H.-B. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, *19*(5), 595-610. https://doi.org/10.1177/1464884916641269 [Letzter Abruf 18.12.2024].

Gulati, S.; Sousa, S.; & Lamas, D. (2019). Design, development and evaluation of a human-computer trust scale. Behaviour & Information Technology, 38(10), 1004–1015. https://doi.org/10.1080/0144929X.2019.1656779 [Letzter Abruf 18.12.2024].

Hofeditz, L.; Mirbabaie, M.; Holstein, J.; & Stieglitz, S. (2021). Do you trust an AI-journalist? A credibility analysis of news content with AI-authorship. *ECIS 2021 Research Papers*. 50. Online verfügbar unter: https://aisel.aisnet.org/ecis2021_rp/50 [Letzter Abruf 12.12.2024].

Jiang, B.; Tan, Z.; Nirmal, A.; & Liu, H. (2024). Disinformation Detection: An Evolving Challenge in the Age of LLMs. In *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)* (pp. 427–435). Online verfügbar unter: https://doi.org/10.1137/1.9781611978032.50 [Letzter Abruf 18.12.2024].

King, S. (2023): Krista 2023 AI Trust Survey, Online verfügbar unter: https://krista.ai/ai-trust-survey-2023/ [Letzter Abruf 10.12.2024].

Kleine, F. (2022): Perception of Deepfake Technology – The Influence of the Recipient's Affinity for Technology on the Perception of Deepfakes, Darmstadt, Online verfügbar unter: https://www.researchgate.net/publication/364254498_Perception_of_Deepfake_Technology_-_The_Influence_of_the_Recipients'_Affinity_for_Technology_on_the_Perception_of_Deepfakes [Letzter Abruf 09.12.2024].

KPMG (2023): Generative AI: From buzz to business value, Online verfügbar unter: https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2023/generative-ai-survey.pdf [Letzter Abruf 10.12.2024].

Longoni, C.; Fradkin, A.; Cian, L.; & Pennycook, G. (2022). News from Generative Artificial Intelligence Is Believed Less. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 97–106. Seoul, Republic of Korea. Online verfügbar unter: https://doi.org/10.1145/3531146.3533077 [Letzter Abruf 10.12.2024].

McKinsey & Company (2023): What's the future of generative AI? An early view in 15 charts, Online verfügbar unter: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/mckinsey%20explainers/whats%20the%20future%20of%20generative%20ai%20an%20early%20view%20in%2015%20charts/whats-the-future-of-generative-ai-an-early-view-in-15-charts.pdf [Letzter Abruf 10.12.2024].

Ng, D. T. K.; Wu, W.; Leung, J. K. L.; Chiu, T. K. F.; & Chu, S. K. W. (2024): Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach, *British Journal of Educational Technology*, 55, p. 1082-1104. Online verfügbar unter: https://doi.org/10.1111/bjet.13411 [Letzter Abruf 18.12.2024].

Spitale, G.; Biller-Andorno, N.; & Germani, Federico (2023): AI model GPT-3 (dis)informs us better than humans, *Science Advances*, 9(26). Online verfügbar unter: https://doi.org/10.1126/sciadv.adh1850 [Letzter Abruf 18.12.2024]

Tandoc Jr., E. C.; Yao, L. J.; & Wu, S. (2020). Man vs. Machine? The Impact of Algorithm Authorship on News Credibility. *Digital Journalism*, *8*(4), 548–562. Online verfügbar unter: https://doi.org/10.1080/21670811.2020.1762102 [Letzter Abruf 18.12.2024].

Toff, B.; & Simon, F. M. (2023). "Or they could just not use it?": the paradox of AI disclosure for audience trust in news. Online verfügbar unter: https://osf.io/preprints/socarxiv/mdvak [Letzter Abruf 10.12.2024].

Waddell, T. F. (2017). A Robot Wrote This? How perceived machine authorship affects news credibility. *Digital Journalism*, *6*(2), 236–255. https://doi.org/10.1080/21670811.2017.1384319 [Letzter Abruf 18.12.2024].

Wu, Y. (2019). Is Automated Journalistic Writing Less Biased? An Experimental Test of Auto-Written and Human-Written News Stories. *Journalism Practice*, *14*(8), 1008–1028. https://doi.org/10.1080/17512786.2019.1682940 [Letzter Abruf 18.12.2024].

Wölker, A.; & Powell, T. E. (2021). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, *22*(1), 86-103. Online verfügbar unter: https://doi.org/10.1177/1464884918757072 [Letzter Abruf 18.12.2024].