

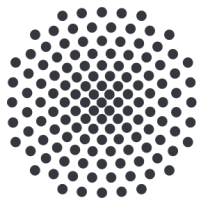
Wertschätzung menschlicher Arbeit in Zeiten generativer KI

BACHELORARBEIT

der Fakultät Wirtschafts- und Sozialwissenschaften der
UNIVERSITÄT HOHENHEIM
zur Erlangung des akademischen Grades
„Bachelor of Science“



UNIVERSITÄT
HOHENHEIM



Universität Stuttgart

Betreut von:

Prof. Dr. Henner Gimpel

Vorgelegt von:

Thomas Schaffrath, Sidney Strasser

3533095, 3532588

Brommerstraße 17A, 70563 Stuttgart

Taschenstraße 54, 70736 Fellbach

Abgabetermin der Arbeit: 29.01.2024

Zusammenfassung

ChatGPT ist seit der globalen Veröffentlichung im November 2022 in aller Munde. Generative KI verändert die Art und Weise, wie wir arbeiten und wie wir Medien konsumieren. Während viele Chancen für die Nutzung von generativer KI existieren, stellt die Monetarisierung von KI-generierten Erzeugnissen Unternehmen vor diverse Herausforderungen. Über die Zahlungsbereitschaft des Konsumierenden für die Erzeugnisse generativer KI existiert nur wenig Literatur. Diese Forschungslücke soll diese Arbeit adressieren. Die Arbeit befasst sich mit der Zahlungsbereitschaft von Konsumierenden für Texte von generativer KI und berücksichtigt dabei die mediiierenden Faktoren Vertrauen und Widerlegung von Annahmen. Das Forschungsmodell wurde quantitativ durch eine multiple lineare Regression geprüft. Die dafür benötigten Daten wurden mithilfe eines Online-Experimentes erhoben, bei welchem Teilnehmende Texte von generativer KI und Menschen erhielten und diese in einer Umfrage auf ihre Zahlungsbereitschaft, ihr Vertrauen und ihre Erwartungen bewerten mussten. Indem die Herkunft des Textes experimentell manipuliert wurde, konnten Aussagen über eine Aversion gegenüber generativer KI und wahrgenommene Qualität getroffen werden. Die Ergebnisse des Experimentes zeigen eine signifikant geringere Zahlungsbereitschaft für Texte, bei welchen die generative KI als Erzeuger angegeben wurde. Gleichzeitig konnte kein signifikanter Zusammenhang zwischen der Zahlungsbereitschaft und generativer KI als tatsächlichem Verfasser gezeigt werden. Die daraus entstehenden moralischen und juristischen Implikationen werden diskutiert. Während Vertrauensprobleme in generative KI existieren, konnte kein signifikanter Effekt auf die Zahlungsbereitschaft festgestellt werden.

Stichworte: Zahlungsbereitschaft, künstliche Intelligenz, Vertrauen, Widerlegung von Annahmen

Inhaltsverzeichnis

Zusammenfassung	I
Inhaltsverzeichnis	II
Abbildungsverzeichnis	IV
Tabellenverzeichnis	V
Abkürzungsverzeichnis	1
1 Einleitung	2
2 Theoretische Grundlagen und Hypothesenentwicklung	3
2.1 generative Künstliche Intelligenz	3
2.1.1 Erkennbarkeit von Künstliche Intelligenz (KI)-generiertem und Auto- renschaft	7
2.1.2 Einfluss auf Medien	9
2.2 Begriffliche Abgrenzungen im Kontext der Zahlungsbereitschaft	10
2.2.1 Zahlungsbereitschaft	10
2.2.2 Verkaufsbereitschaft	10
2.2.3 Zahlungsabsicht	10
2.2.4 Vorausgegangene Zahlungen	10
2.3 Zahlungsbereitschaft für Texte	11
2.4 Auswirkungen der Erwartungshaltung auf die Zahlungsbereitschaft für algo- rithmisch erzeugte Inhalte	12
2.5 Vertrauen	13
2.5.1 Vertrauen in Informationssysteme	14
2.5.2 Vertrauen im literarischen und journalistischen Kontext	16
2.5.3 Vertrauen und Zahlungsbereitschaft	19
3 Experimentdesign	19
3.1 Auswahl der Teilnehmenden	21
3.2 Gezieltes Falschinformieren	22
3.3 Auswahl der Texte	23
3.4 Datenerhebung	28
3.4.1 Vorabumfrage	28
3.4.2 Experiment	28
3.4.3 Nachbefragung	29

4	Ergebnisse	31
4.1	Grundlagen der multiplen linearen Regression	31
4.1.1	Einführung in die multiple lineare Regression	31
4.1.2	Annahmen der multiplen linearen Regression	31
4.1.3	Interpretation der Regressionskoeffizienten	32
4.1.4	Vergleich mit SEM und Begründung der Methodenwahl	33
4.2	Deskriptive Analysen und Datenbeschreibung	33
4.3	Ergebnisse der multiplen linearen Regression	36
5	Diskussion	38
5.1	Eine KI als angegebener Verfasser senkt die Zahlungsbereitschaft für den Text	38
5.2	Das Vertrauen in generative KI als angegebener Verfasser ist geringer als bei menschlichen Verfassern	39
5.3	Das Vertrauen in den Verfasser beeinflusst die Zahlungsbereitschaft nicht signifikant	40
5.4	Widerlegung von Annahmen als medierendes Konstrukt	41
5.5	Kontrollvariablen	41
6	Beitrag zur Forschung	42
7	Limitationen und offene Forschungsimpulse	42
8	Fazit	44
	Literaturverzeichnis	45
	Kapitelangaben und Erklärungen	57
	Kapitelangaben	57
	Erklärung zur Verwendung generativer KI-Systeme	58
	Eigenständigkeitserklärung	60
	Anhang	61

Abbildungsverzeichnis

1	Forschungsmodell	3
2	Formatives Vertrauen nach Mayer et al. (1995)	15
3	Experimentablauf	20
4	Entscheidungsbaum für das Experiment	26
5	Auswahl der Teilnehmenden	34
6	Geschätztes Modell	37

Tabellenverzeichnis

1	Ermittlung der benötigten Teilnehmeranzahl	21
2	Übersicht über die erhobenen Daten	30
3	Deskriptive Statistiken und CFA der Daten	35
4	Fragenkatalog	66
5	Alle Pfade des Regressionmodells	70

Abkürzungsverzeichnis

AV	Angegebener Verfasser
AVK	Angegebener Verfasser ist KI
AVM	Angegebener Verfasser ist Mensch
CFA	Confirmatory Factor Analysis
ChatGPT	Chat generative pre-trained transformer
WvA	Widerlegung von Annahmen
GPT	Generative Pre-trained Transformer
HITL	Human in the Loop
HOOTL	Human out of the Loop
IS	Informationssysteme
KI	Künstliche Intelligenz
LLM	Large Language Model
MLR	Maximum Likelihood Robust
NLP	Natural Language Processing
RNN	Recurrent Neural Network
TV	Tatsächlicher Verfasser
TVM	Tatsächlicher Verfasser ist Mensch
TVK	Tatsächlicher Verfasser ist KI
UrhG	Urheberrechtsgesetz
VB	Verkaufsbereitschaft
VZ	Vorausgegangenen Zahlungen
WGA	Writers Guild of America
ZA	Zahlungsabsicht
ZB	Zahlungsbereitschaft

1 Einleitung

Im Mai 2023 streikte die Gewerkschaft Writers Guild of America (WGA) für 148 Tage, der zweitlängste Streik in der Geschichte der Gewerkschaft. Die WGA ist ein Zusammenschluss von zwei kleineren Gewerkschaften, der „Writers Guild of America East“ und „Writers Guild of America West“. Sie vertritt Autoren und Autorinnen aus der Film- und Rundfunkindustrie in Tarifverhandlungen (Writers Guild of America East, 2024). Neben üblichen Themen wie Bezahlung und Arbeitsbedingungen gab es einen neuen Konfliktpunkt bei den diesjährigen Verhandlungen: die Nutzung von generativer künstlicher Intelligenz (KI), wie beispielsweise ChatGPT, im kreativen Prozess (Writers Guild of America, 2023). Die Bedenken der Gewerkschaft liegen auf der Hand: Drehbuchautoren und Drehbuchautorinnen könnten durch generative KI ihre Anstellung verlieren, wenn Drehbücher künftig automatisiert von generativer KI erstellt werden können (Coyle, 2023a). Obwohl die generelle Einstellung gegenüber KI weiterhin positiv sei, kamen Fast und Horvitz (2017) bereits vor sechs Jahren zu dem Schluss, dass die Angst, durch KI die Anstellung zu verlieren, in den letzten Jahren gestiegen ist. Seit OpenAI den Chatbot ChatGPT im November 2022 veröffentlicht hat, ist die Diskussion um den Einfluss von generativer KI auf den Arbeitsmarkt nun auch in den Vordergrund in vielen kreativen Berufen, wie auch dem Drehbuchschreiben, gerückt. Obwohl der Streik der WGA nun mit einem Abkommen zu Ende gekommen ist, in welchem die WGA weitreichende Zugeständnisse bei der KI-Frage seitens der Filmstudios erhielt (Coyle, 2023b), ist die Diskussion über (generative) KI im Unternehmenskontext in anderen kreativen Bereichen noch nicht zu Ende geführt. Die betriebswirtschaftlichen Vorteile für Unternehmen liegen auf der Hand: Wenn Angestellte durch Automatisierung ersetzt werden können, ergeben sich Lohnkosteneinsparungen bei erhoffter gleicher Leistung. Hui et al. (2023) untersuchten die Auswirkung von generativer KI auf den digitalen Arbeitsmarkt, indem sie die Nachfrage und Vergütung für selbstständige, kreative Arbeit auf der Plattform Upwork seit der Einführung von ChatGPT gemessen haben. Sie konnten dabei einen signifikanten Nachfrage- sowie Einkommensrückgang feststellen: Durch die Potenziale generativer KI nimmt die primäre Unternehmenskundschaft auf Upwork (Upwork Inc., n. D.) seltener die Leistung von freiberuflich arbeitenden Personen an. Da generative KI-Systeme von den Entwicklern mittlerweile auch im Unternehmenskontext monetarisiert werden (Midjourney, 2024; OpenAI, 2023) und Unternehmen potenziell auch einen monetären Nutzen aus der Verwendung der Systeme ziehen möchten, liegt die Vermutung nahe, dass sich aus den Systemen bald auch andere, neuartige Geschäftsmodelle erschließen lassen. So könnten einzigartige generative KI-Systeme entwickelt werden, deren Ausgaben die innehabenden Unternehmen verkaufen könnten. Was wir dabei im Rahmen dieser Arbeit untersuchen wollen, ist die Perspektive der Privatkundschaft. Sind Konsumierende bereit, für (kreative) Arbeit, die von Maschinen erledigt wurde, das Gleiche zu bezahlen, wie für menschliche Arbeit? Generative KI kann zwar bereits deutlich mehr als nur Texte

generieren. Die Anwendungsgebiete reichen von der Musikkomposition (Catak et al., 2021) bis zur Bildgenerierung (Ramesh et al., 2021) und weit darüber hinaus. Dennoch nehmen wir in dieser Arbeit eine Einschränkung auf Texte vor. Einerseits sind wir dadurch in der Lage, an bestehende Literatur über die Zahlungsbereitschaft (ZB) im Journalismusbereich anzuknüpfen (O'Brien et al., 2020), andererseits finden sich für texterzeugende Modelle viele Anwendungen in der Praxis (Raj et al., 2023). Dadurch ergibt sich folgende Forschungsfrage, die im Rahmen dieser Arbeit beantwortet werden soll: *FF1: Wie verhält sich die Zahlungsbereitschaft für Texte, je nachdem, ob der Verfasser ein Mensch oder eine generative KI ist?*

2 Theoretische Grundlagen und Hypothesenentwicklung

Um die ZB für Texte von menschlichen Verfassern und generativer KI zu untersuchen, gilt es vorerst, den theoretischen Hintergrund für generative KI, ZB sowie die diversen Einflussfaktoren auf die ZB zu erläutern. Im Zuge dieser Erläuterung werden dabei die Hypothesen des Forschungsmodells hergeleitet, welche die Forschungsfrage beantworten soll. Das Forschungsmodell ist in Abbildung 1 dargestellt:

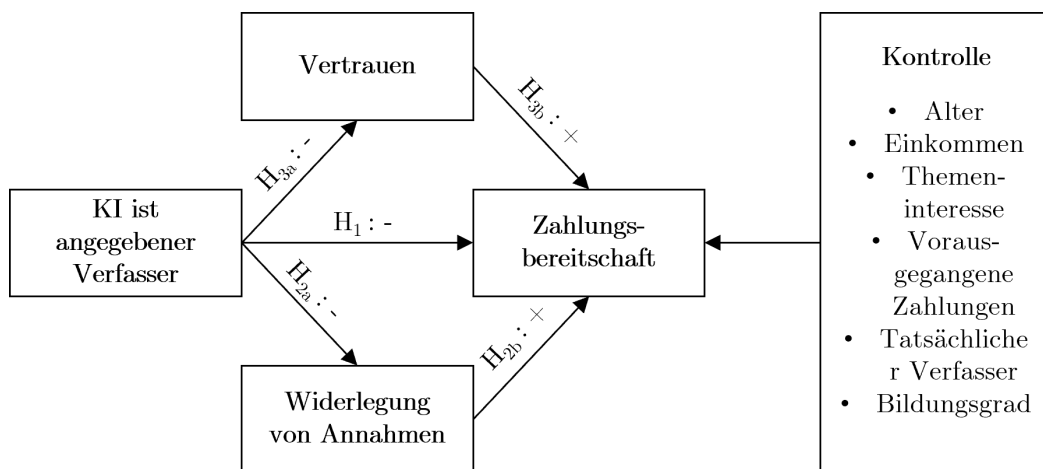


Abbildung 1: Forschungsmodell

2.1 generative Künstliche Intelligenz

Seit einiger Zeit ist KI ein in der Wissenschaft häufig diskutiertes Feld. Nach einer Analyse der Bitkom e.V. (n.d.) ist der Anteil der Unternehmen in Deutschland, die KI verwenden, innerhalb eines Jahres (2022 auf 2023) um sechs Prozentpunkte von 9% auf 15% gestiegen.

Zwei Drittel der befragten Unternehmen sehen KI als die wichtigste Zukunftstechnologie an, was verdeutlicht, dass KI auch zunehmend an strategischer Relevanz in Unternehmen gewinnt. Um was es sich bei KI allerdings genau handelt, lässt sich nicht in wenigen Sätzen erklären. Nach einem Bericht der MMC (2019), in dem 2830 KI-Startups analysiert wurden, gab es bei ca. 40 % der Unternehmen keine Evidenz für einen Einsatz von KI. Der Begriff wird also zunehmend von vor allem jungen Unternehmen als Marketinginstrument für neue Anwendungen und Geschäftsmodelle verwendet, obwohl diese keine Methoden der KI einsetzen. Nur wenige Menschen wissen dabei, was genau hinter den Anwendungen steckt und wie sie funktionieren. In diesem Kapitel sollen die grundlegenden Begriffe erklärt werden, die zum Einsatz kommen, wenn im Rahmen dieser Arbeit von (generativer) KI gesprochen wird.

Der Begriff Künstliche Intelligenz wurde 1956 erstmals im Zuge der Dartmouth Conference von John McCarthy in den USA erwähnt und gilt bis heute als der allgemeingültige Begriff, wenn es darum geht, menschliches Lernen, Denken und Handeln zu mechanisieren. Heute wird Künstliche Intelligenz als Begriff häufig im selben Zuge mit maschinellem Lernen, Deep Learning, Neuronalen Netzen und weiteren Begriffen genannt. Diese Mehrdeutigkeit erschwert dabei das Verständnis des Begriffs selbst. Was allerdings genau unter dem Begriff der künstlichen Intelligenz verstanden wird, ist nicht eindeutig und den Begriff klar zu definieren stellt sich als genauso schwierig dar, wie eine Definition von (menschlicher) Intelligenz selbst. Das Resultat ist eine Vielzahl an kursierenden Definitionen, die auf verschiedene Aspekte des Forschungsgebiets abzielen. So definierte McCarthy (2004) KI in einer populärwissenschaftlichen Veröffentlichung als „die Wissenschaft und Technik der Entwicklung intelligenter Maschinen, insbesondere intelligenter Computerprogramme“ (eigene Übersetzung aus McCarthy (2004), S.2), während Marvin Minsky den Begriff als „[...] die Wissenschaft, die Maschinen dazu bringt, Dinge zu tun, die Intelligenz erfordern würden, wenn sie von Menschen ausgeführt werden“ (eigene Übersetzung der Zitation aus Bolter, 1993, S. 193) beschreibt.

Nach Russell und Norvig (1995) lassen sich die meisten Definitionen für KI allerdings in vier grundlegende Kategorien einteilen:

- Systeme, die wie Menschen denken
- Systeme, die wie Menschen handeln
- Systeme, die rational denken
- Systeme, die rational handeln

Nach einer ausgiebigen Analyse verschiedener Definitionen von KI konnten Samoili et al. (2020) vier grundlegende Gemeinsamkeiten in den Definitionen feststellen, die sie als mögliche Hauptmerkmale von künstlicher Intelligenz identifizieren:

- **Wahrnehmung der Umwelt**, inklusive der Einbeziehung der Komplexität der realen Welt
- **Informationsverarbeitung**: Sammeln und Interpretieren von Inputs (in Form von Daten)
- **Entscheidungsfindung** (einschließlich logisches Denken und Lernen): Handeln, Ausführung von Aufgaben (einschließlich Anpassung, Reaktion auf Veränderungen in der Umwelt) mit einem gewissen Maß an Autonomie
- **Erreichen bestimmter Ziele**: Dies wird als der eigentliche Zweck von KI-Systemen angesehen.

Durch das Auftreten von generativen KI-Systemen wie ChatGPT, Midjourney und anderen Vertretern in jüngster Zeit erfährt das Thema rund um generative KI eine erhöhte mediale- und wissenschaftliche Aufmerksamkeit, was sich in einer erhöhten Anzahl an erschienenen Publikationen mit dem Stichwort und Web-Suchen zu künstlicher Intelligenz bemerkbar macht (Google Trends, [21.01.2024](#)). Laut einer Studie der Bitkom e.V. ([n.d.](#)) verwenden bereits 2 % der deutschen Unternehmen generative KI, weitere 13 % planen einen Einsatz der Technologie und rund ein Viertel (23 %) kann sich generell einen Einsatz von generativer KI vorstellen. Daraus lässt sich schließen, dass die Bedeutung generativer KI auch zunehmend im Unternehmenskontext steigt.

Generative künstliche Intelligenz ist dabei ein eigenständiges Teilgebiet der künstlichen Intelligenz, das wie folgt definiert werden kann: „[...] eine Technologie, die (i) Deep Learning-Modelle nutzt, um (ii) menschenähnliche Inhalte (z. B. Bilder, Wörter) als Reaktion auf (iii) komplexe und vielfältige Anfragen (z. B. Sprachen, Anweisungen, Fragen) zu erzeugen“ (eigene Übersetzung aus Lim et al., [2023](#), S.2). Die Definition von Lim et al. ([2023](#)) lässt sich in die Kategorisierung von Russell und Norvig ([1995](#)) einfügen und erfüllt die Hauptmerkmale nach Samoili et al. ([2020](#)), weshalb wir sie im Rahmen dieser Arbeit für den Begriff der generativen KI verwenden.

Da generative KI in dieser Arbeit für die Generierung von Texten verwendet wird, ist es wichtig, den Begriff „generative KI“ vom Begriff „dialogorientierte KI“ abzugrenzen. Unter *dialogorientierter KI* versteht man einen „[...] Teilbereich der künstlichen Intelligenz, der sich mit sprach- oder textbasierten KI-Systemen beschäftigt, die in der Lage sind, Gespräche und verbale Interaktionen zu simulieren und zu automatisieren.“ (eigene Übersetzung aus Kulkarni et al., [2019](#), S.1). Im Kundendienst werden beispielsweise häufig Chatbots eingesetzt, die auf bestimmte Schlüsselwörter achten und darauf basierend eine Antwort liefern. Als rein regelbasierte Systeme lassen sie sich zwar der dialogorientierten KI zuordnen, da sie auf die Konversation mit Menschen ausgerichtet sind, es kommen hierbei aber keine generativen Technologien zum Einsatz. Genau so sind GPT-3 und GPT-4 nicht zwangsläufig dialogorientierte KI-Systeme, da die Modelle auch ohne den Konversationsaspekt zum Generieren

neuartiger Texte verwendet werden können. Auf der Plattform ChatGPT können die Modelle allerdings mit einem Nutzer in einem Dialog interagieren, was durch eine Feinanpassung der Modelle ermöglicht wird (OpenAI, 2022). Somit ist nicht jede dialogorientierte KI eine generative KI und nicht jede generative KI eine dialogorientierte KI.

In dieser Arbeit wird die Plattform ChatGPT zur Erzeugung von Texten verwendet (OpenAI, 2022). Chat generative pre-trained transformer (ChatGPT) kann dabei als dialogorientierte generative KI klassifiziert werden, da sie auf einen Dialog mit Nutzenden ausgerichtet ist und dabei mithilfe eines speziellen künstlichen neuronalen Netzes auf Anfragen des Nutzers menschenähnliche Antworten generiert. Diese Art von System ist weitestgehend unter dem Begriff *Large Language Model (LLM)* bekannt, das zum Teilgebiet des *Natural Language Processing (NLP)* gehört, welches sich mit der Interaktion zwischen Computern und menschlicher Sprache befasst (Rana, 2023). LLMs sind spezialisierte KI-Systeme, die auf das Verstehen und Generieren von menschlicher Sprache ausgerichtet sind und dafür auf riesigen Textdatenbanken trainiert werden (Schwenk, 2007). Sie stellen somit eine Teilmenge von generativer KI dar. Dazu arbeiten LLMs meist mit sogenannten *Worteinbettungen*: Ein Satz wird dabei in mehrere *Marken* (engl. „Token“) zerlegt, die als Vektor in einem multidimensionalen Raum dargestellt werden können. Marken mit einer ähnlichen Bedeutung werden dabei nah beieinander in den Raum projiziert (Kucharavy et al., 2023). Um die menschliche Sprache sinngemäß verstehen und generieren zu können, ist es wichtig, dass die LLMs mit dem Kontext der Wörter in einem Text arbeiten und nicht nur die einzelnen Marken unabhängig voneinander erfassen. Es müssen also Sequenzen von Marken erfasst und generiert werden können, damit beispielsweise der Kontext eines Satzes bei einer Übersetzung nicht verloren geht (Kucharavy et al., 2023). Ein moderner Ansatz um dies zu ermöglichen ist die *Transformer-Architektur*. Die ursprünglich für die maschinelle Übersetzung von Texten vorgeschlagene Architektur (Vaswani et al., 2017) findet mittlerweile in vielen anderen NLP-Bereichen wie auch der Generierung von Sprache eine Anwendung (Ahmed et al., 2023). Stark vereinfacht lässt sich die Funktionsweise der Architektur wie folgt erklären: Für die Eingabesequenz wird eine Worteinbettung erstellt, die ein sog. „Encoder“ verarbeitet, damit der Kontext der Eingabe widergespiegelt wird. Die verarbeitete Sequenz wird daraufhin an einen sog. „Decoder“ weitergegeben, der basierend auf der erhaltenen Sequenz eine Reihe an Ausgabesequenzen erzeugt. Eine wichtige Komponente der Architektur ist der Aufmerksamkeitsmechanismus, der bestimmte Teile der Eingabe basierend auf ihrer Relevanz für die gesamte Eingabe gewichtet, wenn eine Ausgabesequenz generiert wird. So wird sichergestellt, dass der Kontext und die Bedeutung der Eingabe in der Ausgabe berücksichtigt wird. Schlussendlich wird eine Wahrscheinlichkeitsverteilung für die möglichen Ausgabesequenzen erstellt, auf deren Basis die tatsächliche Ausgabesequenz gewählt wird (Briganti, 2023). Die *Generative Pre-trained Transformer (GPT)* nutzen die Transformer-Architektur und sind dabei eine spezifische Reihe von LLMs, die von OpenAI entwickelt werden.

Wenn man die Technologie als anwendende Person nutzen möchte, ist das theoretische Hintergrundwissen über die Funktionsweise von geringer Bedeutung. Wichtiger ist die Art und Weise, wie man *Prompts*, also Anfragen an die KI, stellt. Aus dieser Herausforderung hat sich das Gebiet des *Prompt Engineering* entwickelt. Dabei geht es um die „Gestaltung des Inputs oder der 'Anfrage' in einer Art und Weise, die das Modell dazu bringt, den gewünschten Output zu produzieren“ (eigene Übersetzung aus Heston und Khun, 2023, S.2). In aktuellen Zeiten wird diese Disziplin immer wichtiger, da durch richtiges Stellen von Anfragen spezifischere und qualitativ hochwertigere Ausgaben nach den Bedürfnissen der anfragestellenden Person erzielt werden und so fehlerhafte oder irreführende Ausgaben vermieden werden können (OpenAI, 2022). Die Anfrage kann entweder vom Menschen selbst oder automatisiert erzeugt und verarbeitet werden. Dieser Unterschied wird in der wissenschaftlichen Literatur als *Human in the Loop (HITL)* und das Gegenstück als *Human out of the Loop (HOOTL)* bezeichnet (Schirner et al., 2013).

2.1.1 Erkennbarkeit von KI-generiertem und Autorenschaft

Da Systeme wie ChatGPT, Midjourney und Co aufgrund ihrer benutzerfreundlichen Gestaltung von immer mehr Konsumierenden genutzt werden und zunehmend auch Unternehmen auf die Potenziale der Technologien aufmerksam werden, ist das Resultat eine immer größer werdende Menge an KI-generierten Inhalten im Internet und in anderen Medien. Dadurch stellt sich die Frage der Erkennbarkeit KI-generierter Inhalte – Ist es möglich, KI-generiertes von menschengemachtem zu unterscheiden? Bereits im Jahre 1950 hat Alan Turing den *Turing-Test* (ursprünglich „Imitation Game“) vorgeschlagen, um festzustellen, ob eine Maschine ein gleichmächtiges Denkvermögen besitzt wie ein Mensch (Turing, 1950). Der Turing-Test zielt darauf ab, einen Menschen mit einer Maschine sprechen zu lassen, ohne dass der Mensch eine Kenntnis über seine Gegenpartei besitzt. Kann der Mensch nicht erkennen, dass es sich bei seinem Gegenüber um eine Maschine handelt, hat die Maschine den Test bestanden. Der Turing-Test kommt auch heute noch in modifizierten Varianten zum Einsatz, unter anderem um das Problem der Erkennbarkeit von KI-generierten Inhalten zu beantworten. Die rasante Entwicklung in diversen Bereichen der künstlichen Intelligenz macht es Menschen allerdings immer schwerer, KI-generierte Inhalte von Menschengemachten zu unterscheiden. Bereits vor dem Erscheinen von ChatGPT im November 2022 kamen Köbis und Mossink (2021) zu dem Schluss, dass Menschen nicht in der Lage waren, KI-generierte Poesie von menschengeschriebenen Gedichten zu unterscheiden. Die von dem GPT-3(.5)/4-Vorgänger GPT-2 generierten Gedichte konkurrierten sogar mit denen angesehener Dichter, was das Problem zusätzlich verdeutlicht. Diverse wissenschaftliche Arbeiten unterstützen diese Aussage (Hitsuwari et al., 2023, Kreps et al., 2022). Auch in anderen kreativen Bereichen wie der bildenden Kunst gibt es verschiedene Studien, die auf dieselbe Problemstellung hinweisen (Gangadharbatla, 2022). Durch den zunehmenden Einsatz von LLMs beschäftigen sich zahlreiche Werke mit

dem Problem der Erkennbarkeit von KI-generierten Inhalten. Es werden diverse Modelle vorgeschlagen, die (teil-)automatisiert erkennen können sollen, ob ein Text KI-generiert ist oder nicht (Gehrmann et al., 2019, Mitchell et al., 2023). Forschende sind weiterhin in der Lage, Systeme zur Erkennung generativer KI als ungenau darzustellen (Sadasivan et al., 2023). Dementsprechend bedarf es auf diesem Gebiet weiterer Forschung. Das Grundproblem der Erkennbarkeit bleibt also vorerst bestehen. Da man aktuell nicht zwischen den Inhalten von Mensch und Maschine unterscheiden kann, müssen vorerst rechtliche und ethische Fragestellungen beantwortet werden. Konkreter geht es vor allem um die Urheberschaft von KI-generierten Inhalten – Wer ist nun die urhebende Person eines KI-generierten Werkes? Der Mensch, der den Text von der KI angefordert hat? Die KI selbst?

Im deutschen Recht wird das Urheberrecht über das Urheberrechtsgesetz (UrhG) geregelt. Nach dem UrhG sind KI-Erzeugnisse weitestgehend schutzlos. Sie gelten nicht als persönliche geistige Schöpfung, da das Urheberrecht nur menschliches Schaffen anerkennt. Allerdings könnten einige KI-Erzeugnisse unter das Leistungsschutzrecht des Datenbankherstellers fallen, da dieses keine persönliche geistige Schöpfung voraussetzt, sondern auf Investitionsschutz abzielt (Legner, 2020). Trifft ein Mensch allerdings eine eigene gestalterische Entscheidung, die die Ausdrucksform des konkreten Erzeugnisses beeinflusst, ist urheberrechtlicher Schutz zu erwägen (Legner, 2020). Dass diese Abgrenzung allerdings schwierig sein kann, liegt auf der Hand. Es ist noch nicht geklärt, in welchem Umfang ein Mensch an dem Werkstück mitzuwirken hat, damit es urheberrechtlichem Schutz unterliegt. Das UrhG definiert die urhebende Person als „Schöpfer des Werkes“ („Urheberrechtsgesetz“, 2021), definiert jedoch die Begriffe „Verfasser“ und „Autor“ nicht direkt. Dennoch werden die Begriffe „Verfasser“ und „Autor“ als Synonyme für den Urheberbegriff verwendet („Urheberrechtsgesetz“, 2021, §65), („Urheberrechtsgesetz“, 2021, §61a, Anlage). Somit geht nach der deutschen Rechtsprechung mit der Autorenschaft und Verfasserschaft eines Werkes das Urheberrecht und Verwertungsrecht implizit einher. Es wird deshalb im Rahmen dieser Arbeit, wo möglich, vermieden, die Begriffe „Autor“, „Verfasser“ und „Urheber“ für generative KI zu verwenden, um die juristischen Implikationen bezüglich der Urheberrechte nach UrhG, die mit diesen Begriffen einhergehen, zu vermeiden. Da es im Sinne der Verständlichkeit herausfordernd ist, umschreibende Begriffe für die Tätigkeit einer generativen KI zu finden, wird die Benutzung der Begriffe „Autor“ und „Verfasser“ teilweise unvermeidlich sein. Die Benutzung dieser Begriffe soll nicht implizieren, dass generativer KI die gleichen Urheberrechte nach UrhG zustehen wie einem menschlichen Verfasser und dienen lediglich der Lesbarkeit und Verständlichkeit. Die juristische Einordnung von Erzeugnissen generativer KI ist in diversen Nachbarstaaten anders geregelt. Im Vereinigten Königreich ist es beispielsweise möglich, dass jene Person, die die notwendigen Vorkehrungen für die Erstellung eines computergenerierten Werks trifft, gemäß dem „Copyright, Designs and Patents Act“ urheberrechtlichen Schutz für dieses Werk erhält (UK Public General Acts,

1988). Da keine Erkennbarkeit von KI-generierten Inhalten gewährleistet werden kann und auch die rechtlichen Bedingungen global noch nicht standardisiert sind, treffen international anerkannte wissenschaftliche Zeitschriften wie „Nature“ bereits Vorgaben, wie Verfasser mit generativer KI umzugehen haben, sodass der Einsatz der Technologie möglichst transparent gestaltet wird (Nature, 2023). Solche gesetzlichen Transparenzmaßnahmen werden von großen Interessenverbänden der Journalisten und Journalistinnen bereits gefordert (ver.di, 2023) und auf europäischer Ebene im Rahmen der sog. KI-Verordnung¹ bereits für audiovisuelle Inhalte diskutiert (Füllsack, 2023).

2.1.2 Einfluss auf Medien

Das Feld der *Content Creation*, also die Erzeugung von Inhalten, spielt vor allem in den sozialen Medien eine wichtige Rolle. Große Plattformen wie YouTube, Twitter und Instagram sind auf die Inhalte der Nutzenden angewiesen, die wiederum von anderen Nutzenden konsumiert werden. Durch generative KI eröffnen sich den Erstellenden der Inhalte dabei neue Möglichkeiten, diese zu gestalten. KI kann dabei monotone Aufgaben übernehmen und bietet den Erstellenden die Möglichkeit, sich mehr auf die Ideen und Konzepte der Inhalte zu konzentrieren (Huang et al., 2023) und zudem wird einem breiteren Spektrum an Nutzenden ein Zugang zur Erzeugung von Inhalten mithilfe generativer KI ermöglicht (Ghosh und Fossas, 2022, Cao et al., 2023). Auch die Plattformen selbst haben das Potenzial der Technologie bereits erkannt und setzen generative KI ein. So gibt es z. B. auf TikTok einige Filter, die generative KI einsetzen, um Gesichter oder die Umgebung zu verändern. Neben den Vorteilen, die generative KI für die Erzeugung von Inhalten bietet, gibt es allerdings auch einige Probleme, die berücksichtigt werden müssen. So haben vor allem kleinere Content Creator Angst, durch generative KI ersetzt zu werden (Ooi et al., 2023, Gmyrek et al., 2023, Budhwar et al., 2023). Auch das in Absatz 2.1.1 angesprochene Problem des Urheberrechts spielt in den Medien eine große Rolle, da menschliche Werke zum Training der KI-Modelle eingesetzt werden. Vor allem letzteres Problem ist zurzeit in starker Diskussion. Am 27. Dezember 2023 verklagte die New York Times OpenAI und Microsoft wegen Urheberrechtsverletzungen, da Millionen ihrer veröffentlichten Artikel zum Training automatisierter Chatbots verwendet wurden. Diese Chatbots stehen nun in direkter Konkurrenz zu den Nachrichtenagenturen als verlässliche Informationsquellen (Metz, 2024). Da die Erzeugung von Inhalten mittlerweile ein eigenständiges Berufsfeld ist, stellt sich in diesem Zuge auch die Frage der Monetarisierung KI-generierter Inhalte. In diesem Kontext ist die ZB ein entscheidendes Konstrukt, mit dem es diese Frage zu evaluieren gilt.

¹Der vollständige Name der Verordnung lautet: Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union.

2.2 Begriffliche Abgrenzungen im Kontext der Zahlungsbereitschaft

Im Kontext dieser Arbeit ist der Begriff der Zahlungsbereitschaft (ZB) von entscheidender Bedeutung. In der Literatur gibt es allerdings viele verschiedene Begriffe, die oftmals gleichbedeutend genutzt werden (O'Brien et al., 2020) und die es von der ZB abzugrenzen gilt.

2.2.1 Zahlungsbereitschaft

Die *Zahlungsbereitschaft (ZB)* ist eine quantitative Kenngröße, die beschreibt, wie viel Geld eine Person maximal bereit ist, für ein Gut zu zahlen (Chyi und Yang, 2009, Goyanes, 2014). Als intervallskalierte Kenngröße gibt sie also die Obergrenze der monetären Wertschätzung eines (digitalen) Gutes oder einer Dienstleistung einer Person an. Die ZB kann entweder mittels direkter oder indirekter Erhebungen angegeben oder durch tatsächliche oder simulierte Daten offengelegt werden (Breidert et al., 2015).

2.2.2 Verkaufsbereitschaft

Anders als die ZB stellt die *Verkaufsbereitschaft (VB)* den Preis dar, zu dem eine Person ein bestimmtes Gut oder eine Dienstleistung verkaufen würde (Horowitz und McConnell, 2003). Sie ist somit auch eine intervallskalierte Kenngröße, stellt jedoch keine Ober- oder Untergrenze dar. Horowitz und McConnell (2003) haben zudem festgestellt, dass die VB fast immer größer als die ZB ist.

2.2.3 Zahlungsabsicht

In der akademischen Diskussion wird die *Zahlungsabsicht (ZA)* als eine eigenständige Kenngröße herausgestellt, die sich von der ZB abgrenzt. Im Unterschied zur ZB, die den maximalen Betrag abbildet, den eine Person bereit ist zu zahlen, wird die Zahlungsabsicht (ZA) mittels einer dichotomen Skala gemessen. Diese dichotome Messung erfasst lediglich, ob eine Person grundsätzlich geneigt ist, finanzielle Mittel für (digitale) Produkte aufzuwenden. Dieser Unterschied in der Operationalisierung der beiden Konstrukte wird in der Forschungsliteratur, unter anderem von Chyi und Yang (2009), deutlich hervorgehoben.

2.2.4 Vorausgegangene Zahlungen

Eine weitere wichtige Kenngröße im Rahmen dieser Arbeit ist das Konzept der *Vorausgegangenen Zahlungen (VZ)*. Die VZ bezieht sich auf die Frage, ob oder wie viel in der Vergangenheit für ein bestimmtes Gut gezahlt wurde. Dieses Konstrukt kann auf zwei Weisen erfasst werden: entweder dichotom, indem ermittelt wird, ob eine Person in der Vergangenheit für ein bestimmtes Gut gezahlt hat oder nicht oder auf einer kontinuierlichen Skala, indem konkret

erfragt wird, wie viel eine Person in der Vergangenheit für das Gut gezahlt hat (O'Brien et al., 2020). In der wissenschaftlichen Literatur wurde häufig darauf hingewiesen, dass eine starke Korrelation zwischen den VZ und der ZB besteht (O'Brien et al., 2020). Dies deutet darauf hin, dass frühere Zahlungen als Prädiktor für zukünftige ZB dienen können. Aufgrund dessen werden die VZ als Kontrollvariable in das Forschungsmodell aufgenommen.

2.3 Zahlungsbereitschaft für Texte

Im Journalismus wurde bereits ausgiebig untersucht, wie die ZB für verschiedene Formen journalistischer Inhalte aufgrund verschiedener Einflussfaktoren ausfällt. Dabei spielt vor allem die ZB für digitalen Journalismus eine übergeordnete Rolle und wird von mehreren Studien durchleuchtet. In einer systematischen Literaturrecherche stellten O'Brien et al. (2020) drei Kategorien für diverse Einflussfaktoren auf die ZB für digitale journalistische Inhalte heraus:

1. **Konsumentenbasierte Faktoren:** Alter, Geschlecht, Bildungsgrad, Ausmaß der Mediennutzung und Nachrichteninteresse
2. **Produktbasierte Faktoren:** Format/Medium, Personalisierung, technische Benutzerfreundlichkeit, Exklusivität, Qualität und Spezialisierung/Nische
3. **Ökonomische Faktoren:** Einkommen, Preis und Substitute/kostenlose Alternativen

Dabei fanden Sie heraus, dass vor allem die Faktoren Nachrichteninteresse und Format/Medium und der Preis einen signifikanten Einfluss auf die ZB haben. Da das Format Medium konstant gehalten wird, muss es nicht in das Modell aufgenommen werden. Das Nachrichten- bzw. Themeninteresse sowie die demografischen Faktoren Alter, Geschlecht, Bildungsgrad und Einkommen werden als Kontrollvariablen in das Forschungsmodell aufgenommen.

Abseits des Journalismus ist die Literatur zur ZB für literarische Texte begrenzt. Stejskal et al. (2021) untersuchten für Bibliotheksnutzende der Tschechischen Republik, welche Faktoren einen Einfluss auf die ZB für E-Books haben. Auch hier stellt sich erneut heraus, dass der Preis ein entscheidendes Kriterium ist.

Wie die ZB allerdings für KI-generierte Texte ausfällt und welche Faktoren diese beeinflussen, wurde in der Literatur noch nicht untersucht und soll in dieser Arbeit evaluiert werden. Wir gehen davon aus, dass Menschen nicht bereit sind, für KI-generierte Texte genau so viel oder mehr zu bezahlen wie für von Menschen verfasste Texte, da:

1. KI-generierte Texte in kürzester Zeit erzeugt werden können. Menschen sind vermutlich eher bereit, Geld für eine Leistung zu bezahlen, die sich aus der eingeflossenen Arbeit und der dafür benötigten Zeit ergibt. Dieses Verhältnis ist bei KI-generierten Texten nicht stimmig, was durch eine geringere ZB ausgedrückt werden könnte.

2. Texte von menschlichen Verfassern eine persönliche Note und Authentizität haben können, die KI-generierte Texte möglicherweise nicht bieten. Es fehlt KI-generierten Inhalten an eigenen Nuancen, die es (aktuell) in dieser Form nur bei menschlichen Verfassern gibt.
3. KI-generierte Texte auf Basis verschiedenster Texte menschlichen Ursprungs erzeugt werden, da diese die Grundlage für das Training der Modelle schaffen. Somit fehlt es KI-generierten Texten an Originalität und das zuvor angesprochene Problem der Urheberschaft wirft neben juristischen Problematiken weitere moralische Fragen auf, die möglicherweise die ZB drücken können.

Aus den angesprochenen Gründen folgern wir somit:

H₁: Ist der angegebene Verfasser eine künstliche Intelligenz, so sinkt die Zahlungsbereitschaft für den entsprechenden Text.

2.4 Auswirkungen der Erwartungshaltung auf die Zahlungsbereitschaft für algorithmisch erzeugte Inhalte

Abseits des Journalismus haben Rix et al. (2022) für algorithmisch erzeugte Inhalte untersucht, ob die Offenlegung des algorithmischen Ursprungs der Inhalte eine Auswirkung auf die ZA der Teilnehmenden hat. Unter der Annahme, dass eine Abneigung gegenüber algorithmisch erzeugten Inhalten besteht, stellten sie fest, dass die Teilnehmenden im Gegenteil eine Wertschätzung gegenüber den algorithmisch erzeugten Inhalten äußerten. Einen Teil des Effekts erklärten Sie sich durch die von den Teilnehmenden beobachtete Performanz der Algorithmen. Sie sprachen davon, dass die Algorithmen möglicherweise die Erwartungen der Teilnehmenden übertroffen haben, was zu diesem Effekt führen könnte.

Dieses Konstrukt lässt sich als *Widerlegung von Annahmen (WvA)*² operationalisieren und stammt aus der *Expectation Confirmation Theorie*, die versucht, die Zufriedenheit einer Person mit einem Produkt, einer Dienstleistung oder einem Technologieartefakt nach einem Kauf oder einer Adaption durch die Erwartungen an das Gut, die beobachtete Performanz und durch die WvA zu erklären (Oliver, 1980). Die WvA ist dabei das Urteil, das eine Person nach der Beobachtung der Performanz eines Gutes im Vergleich zu seinen vorherigen Erwartungen fällt und berechnet sich als Differenz aus der tatsächlich beobachteten Performanz und den ursprünglichen Erwartungen an die Performanz des Gutes. Ist die WvA positiv, so wurden die Erwartungen übertroffen. Die Expectation Confirmation Theorie wurde bereits in manchen Bereichen der IS-Forschung eingesetzt (Hossain und Quaddus, 2012), ist im Kontext der generativen KI in der Literatur aber noch unterentwickelt. Wir adaptieren das Konstrukt in

²englisch: Disconfirmation of Beliefs

dieser Arbeit, um zu messen, ob die Erwartungen der Teilnehmenden an die verschiedenen Verfasser über- oder untertroffen wurden. Da die Arbeit von Rix et al. (2022) noch vor dem globalen Start von ChatGPT veröffentlicht wurde und der Großteil der Bevölkerung zu diesem Zeitpunkt noch nicht in einer solchen Form mit generativer KI interagiert hat, vermuten wir, dass die Erwartungen an die Performanz von generativer KI seitdem stark gestiegen sind. Menschen, die mit generativer KI arbeiten, sind sich über die Form und Qualität der Ausgaben bereits bewusst und werden beispielsweise nicht von einem KI-generierten Klappentext für einen nicht existierenden Kriminalroman überrascht sein. Daraus folgern wir:

H_{2a}: Ist der angegebene Verfasser eines Textes eine künstliche Intelligenz, ist die WvA geringer, als bei einem menschlichen Verfasser.

Da nach Rix et al. (2022) ein Faktor für die ZA der Teilnehmenden für algorithmisch erzeugte Inhalte ein Übertreffen der ursprünglichen Erwartungen, also eine positive WvA, war, stellt sich die Frage, wie sich die ZA auf die ZB der Teilnehmenden auswirken kann. In einer Studie zur ZA und ZB für umweltfreundlichen Güterverkehr konnten Schniederjans et al. (2020) herausfinden, dass ein positiver Zusammenhang zwischen der ZA und der ZB besteht. So waren Teilnehmende mit einer stärker ausgeprägten ZA eher bereit, mehr Geld für ein T-Shirt auszugeben, das den umweltfreundlichen Güterverkehr unterstützt, als Teilnehmende mit einer geringer ausgeprägten ZA. Da also ein Zusammenhang zwischen der ZA und der ZB besteht und die ZA von einer positiven WvA beeinflusst wird, argumentieren wir:

H_{2b}: Mit steigender WvA steigt ebenso die Zahlungsbereitschaft für den Text.

2.5 Vertrauen

Nachdem die ZB als Zielkonstrukt eingeführt wurde, gilt es nun, Vertrauen als ein weiteres wichtiges Konstrukt im Rahmen dieser Arbeit näher zu betrachten. *Vertrauen* ist ein in der wissenschaftlichen Literatur viel diskutierter, oft falsch verstandener Begriff (Mayer et al., 1995). Eine häufig zitierte Definition nach Mayer et al. (1995) beschreibt Vertrauen als „die Bereitschaft einer Partei, sich den Handlungen einer anderen Partei auszusetzen, basierend auf der Erwartung, dass die andere Partei eine Handlung, die für den Vertrauensgeber wichtig ist, vornimmt, unabhängig von der Möglichkeit des Vertrauensgebers, den Vertrauensnehmer dabei zu überwachen oder zu kontrollieren“ (eigene Übersetzung aus Mayer et al., 1995, S. 712). Die fehlende Kontrolle über den Vertrauensnehmer ist ein wichtiger Aspekt dieser Definition. Um zu vertrauen, muss der Vertrauensgeber ein Risiko eingehen, etwas zu verlieren. Das kann etwa ein monetärer Verlust sein, wenn ein erworbenes Buch sich als uninteressant und schlecht geschrieben herausstellt. Dies bezeichnet man als die *Vulnerabilität* des Vertrauensgebers

gegenüber dem Vertrauensnehmer. Die Bereitschaft, vulnerabel gegenüber der anderen Partei zu sein und somit ein Risiko einzugehen, grenzt das Vertrauen von anderen Konstrukten wie der Vorhersehbarkeit, Zuversicht und Kooperation ab (Mayer et al., 1995). Die Eigenschaft, vulnerabel gegenüber dem Vertrauensnehmer zu sein, ist eine gemeinsame Eigenschaft in vielen anderen Vertrauensdefinitionen (Rousseau et al., 1998). In der Literatur finden sich diverse Modelle, die erklären, wie Vertrauen formativ beschrieben werden kann (Mayer et al., 1995; McKnight et al., 1998). Das Vertrauensmodell nach Mayer et al. (1995) definiert Eigenschaften, die sich in der Forschung als *Vertrauensüberzeugungen* etabliert haben (McKnight et al., 1998). Diese Vertrauensüberzeugungen setzen sich aus den Eigenschaften des Vertrauensnehmers und den Eigenschaften des Vertrauensgebers zusammen. So evaluiert der Vertrauensgeber die Eigenschaften der Fähigkeit, des Wohlwollens und der Integrität des Vertrauensnehmers. Diese drei Eigenschaften sind von Mayer et al. (1995) definiert:

- *Fähigkeit* beschreibt eine Menge an Kompetenzen, Fertigkeiten und Charakteristika, die es einer Partei ermöglichen, Einfluss auf eine spezielle Domäne zu haben.
- *Wohlwollen* definiert sich als der Grad, in dem ein Vertrauensnehmer dem Vertrauensgeber Gutes zukommen lassen möchte, ohne dabei von egozentrischen Motiven geleitet zu werden.
- *Integrität* wird als das Befolgen gewisser Prinzipien definiert, die der Vertrauensgeber akzeptabel findet.

Auf die Vertrauensbeziehung, sowohl auf das Gesamtvertrauen als auch medierend auf die Beziehung zwischen Fähigkeit, Wohlwollen und Integrität auf Vertrauen, wirkt sich dann zusätzlich die *Vertrauensneigung* des Vertrauensgebers aus. Die Vertrauensneigung ist dabei personenabhängig. Sie beschreibt unabhängig von den Eigenschaften des Vertrauensnehmers, wie sehr eine Person inhärent vertraut. Abbildung 2 bildet das Vertrauensmodell nach Mayer et al. (1995) ab.

Einige Forschende fragen Vertrauen oftmals lediglich reflektiv mit einer einzelnen Frage ab (Rahimi et al., 2022; Meng et al., 2020). Wird Vertrauen lediglich reflektiv abgefragt, berücksichtigt dies nicht den formativen Aspekt des Vertrauens, dass sich das Vertrauen durch andere Eigenschaften zusammensetzt.

2.5.1 Vertrauen in Informationssysteme

Die Arbeit von Mayer et al. (1995), wie der Großteil der Vertrauensforschung der Zeit, beschäftigt sich primär mit dem zwischenmenschlichen Vertrauen. Grundsätzlich schließt die Vertrauensdefinition nach Mayer et al. (1995) KI bzw. Technologie als Ganzes nicht aus. So muss man sich heutzutage auf die Handlungen von Technologie verlassen, ohne die Möglichkeit zu haben, diese überwachen oder kontrollieren zu können (McKnight, 2005).

Vertrauensüberzeugungen

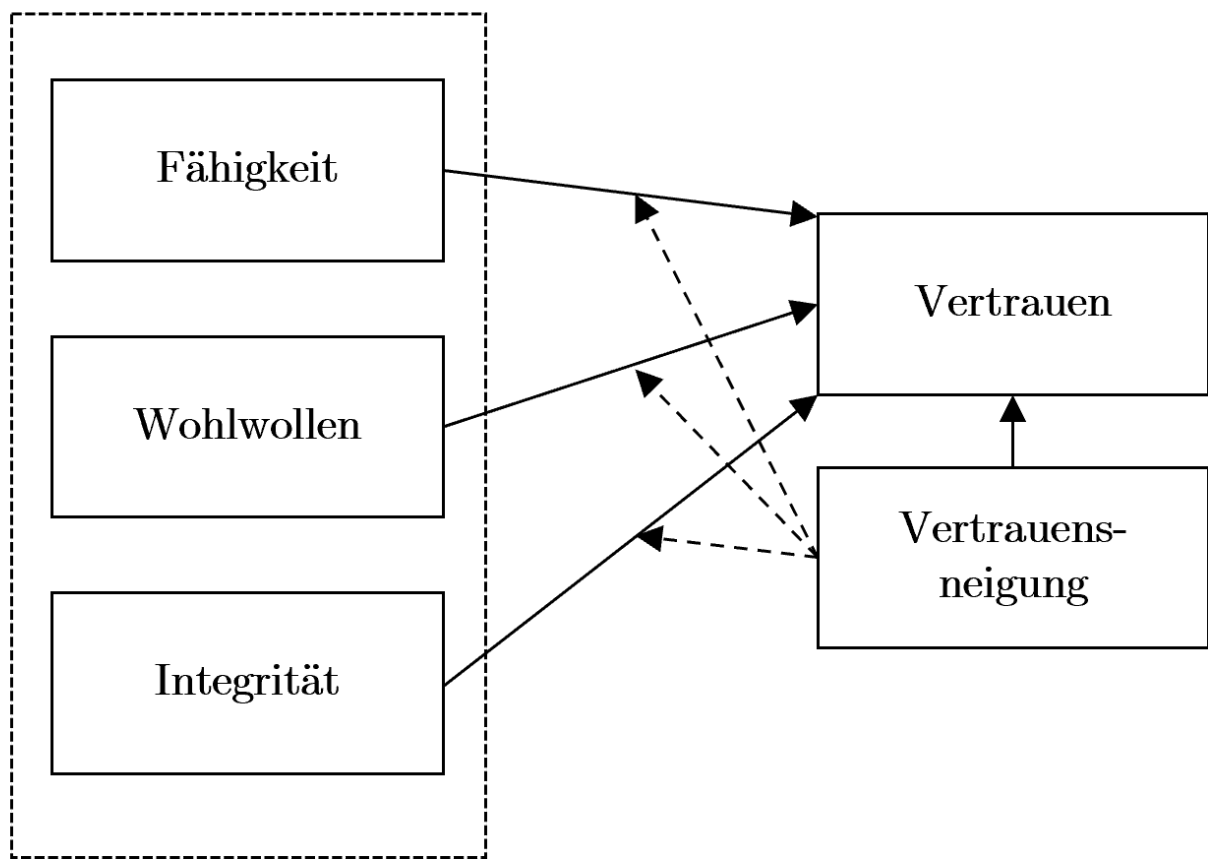


Abbildung 2: Formatives Vertrauen nach Mayer et al. (1995)

Mit dem anhaltenden technologischen Fortschritt sowie dem Aufkommen der Disziplinen Informationssysteme (IS) und Wirtschaftsinformatik, haben sich Forschende der letzten Jahre mit der Frage auseinandergesetzt, wie Vertrauen in IS modelliert werden kann. Die Frage, ob man Technologie überhaupt vertrauen kann, ist in der Literatur umstritten. Für einige Forschende kann Vertrauen in Technologie, sowohl herkömmliche Systeme sowie KI, nicht existieren. Das Abfragen von Vertrauen in Technologie stelle einen grundlegenden Fehler dar, denn Vertrauen existiere ausschließlich zwischenmenschlich (Pitt, 2010; Nickel et al., 2010; Ryan, 2020). Die richtige Frage sei es, das Vertrauen in die Menschen, die diese Systeme erschaffen haben, abzufragen. Andere Forschende halten dagegen. Fragen zu Vertrauen in IT-Systeme sind in der Literatur häufig anzutreffen (Söllner et al., 2012; Li et al., 2008; Rahimi et al., 2022). Ostrom et al. (2019) zeigen sogar die Wichtigkeit des Vertrauens in KI-Systeme für erhöhte Akzeptanz bei Verbrauchern auf. Untersuchungen aus der Soziologie fanden zudem heraus, dass Menschen Computer als soziale Wesen behandelten und Ihnen soziale Eigenschaften, die normalerweise ausschließlich Menschen zugewiesen wurden, zuweisen (Reeves und Nass, 1996; Nass und Moon, 2000). Erweiterungen der Vertrauensdefinitionen des Zwischenmenschlichen auf IS ist also ein naheliegender Schritt. In der Literatur finden

sich einige Arbeiten, welche die Vertrauensüberzeugungen nach Mayer et al. (1995) auf IS-Kontexte (Li et al., 2008; McKnight et al., 2002) sowie auf KI-Kontexte (Bayer et al., 2022) anpassen.

2.5.2 Vertrauen im literarischen und journalistischen Kontext

Vertrauen spielt also für die Akzeptanz von KI eine wichtige Rolle. Es gilt im Folgenden nun zu untersuchen, welche Rolle Vertrauen in KI-Systeme im literarischen und journalistischen Kontext spielt. Im journalistischen Bereich haben Lucassen und Schraagen (2012) ein Vertrauensmodell aufgestellt, welches erklärt, wie Vertrauen in Informationen, beispielsweise einen journalistischen Artikel, zustande kommt. Das Modell ist ein Schichtenmodell, bei welchem das Vertrauen in jede Schicht das Vertrauen in die nächste Schicht beeinflusst. Das Modell besteht aus vier Schichten:

1. *Vertrauensneigung*: Die Vertrauensneigung spiegelt, ähnlich zu dem Modell nach Mayer et al. (1995), eine Persönlichkeitsausprägung wider und beschreibt, wie wahrscheinlich es ist, dass eine Person vertraut. Jede Vertrauensentscheidung wird von dieser persönlichen Vertrauensausprägung beeinflusst.
2. *Vertrauen in das Medium*: Ein Medium ist recht breit als die Generalisierung bzw. Sammlung mehrerer Quellen beschrieben. Die Generalisierung kann eng oder breit gefasst werden. So ist sowohl das gesamte Internet ein Medium, als Generalisierung aller Internetseiten, als auch soziale Netzwerke, als Generalisierung aller sozialen Netzwerke wie beispielsweise X (ehemals Twitter), Facebook und Reddit. Sogar einzelne Websites können nach dieser Definition als Medium betrachtet werden. So ist das soziale Netzwerk Reddit auch nur eine Sammlung unterschiedlicher kleiner Foren. Viele Menschen hätten wohl ein geringeres Vertrauen in ein ihnen unbekanntes digitales Aktienforum, als in eine ihnen unbekannte Börsenzeitschrift, da viele Menschen dem Medium Zeitung wohl ein höheres Vertrauen zuordnen würden, wie dem Medium Internet (Newman et al., 2018). Vertrauen in ein Medium ist also wichtig, da oftmals von dem Vertrauen in ein Medium auf das Vertrauen in eine einzelne Quelle geschlossen wird.
3. *Vertrauen in die Quelle*: Das Vertrauen in die Quelle wird als eine Spezifizierung des Vertrauens in ein Medium beschrieben. Die Entscheidung, ob man einer Quelle vertraut, trägt maßgeblich dazu bei, ob man den enthaltenen Informationen vertraut.
4. *Vertrauen in die Information*: Die Entscheidung zu dem Vertrauen in die Information wird maßgeblich durch die Qualität des Textes bestimmt und zudem durch die äußeren Schichten beeinflusst.

Lucassen und Schraagen (2012) arbeiten zwar mit der Vertrauensdefinition nach Mayer et al. (1995), vernachlässigen jedoch die formative Natur des nach Mayer et al. (1995) definier-

ten Vertrauens über die Vertrauensüberzeugungen Wohlwollen, Integrität und Fähigkeit. Das Vertrauen in die einzelnen Schichten wird reflektiv abgefragt. Da es möglich ist, einer Information zu vertrauen obwohl man der Quelle und dem Medium nicht vertraut, beispielsweise durch sorgfältiges Prüfen durch eine andere Quelle, gestaltet es sich schwierig, Vertrauen als reflektives in Schichten getrenntes Konstrukt abzufragen und auszuwerten. Dieser Sachverhalt kann, im Gegensatz zu dem formativen Ansatz nach Mayer et al. (1995), nicht durch statistische Verfahren beschrieben werden. Es wird deswegen vorwiegend in dieser Arbeit primär mit der Definition und den Vertrauensüberzeugungen nach Mayer et al. (1995) gearbeitet. Dennoch liefert das Schichtenmodell interessante Ansätze bezüglich der Erklärbarkeit des Vertrauens: Vertrauen Menschen einem Text nicht, da er inhaltliche Schwächen hat? Oder liegt es daran, dass sie der Quelle nicht vertrauen?

Die Begriffe der Glaubwürdigkeit und Vertrauen werden in der Literatur teils als Synonyme verwendet (Wölker und Powell, 2018; Lucassen und Schraagen, 2012), sind aber unterschiedlich definiert. Eine häufig zitierte Definition der *Glaubwürdigkeit* definiert diese so: „Glaubwürdigkeit ist sinnvollerweise als ein Teilphänomen von Vertrauen rekonstruierbar und ist als eine Eigenschaft bestimmbar, die Menschen, Institutionen oder deren kommunikativen Produkten (mündliche oder schriftliche Texte [...]) von jemandem (Rezipienten) in Bezug auf etwas [...] zugeschrieben wird“ (Bentele, 1998, S. 305). Parteien seien dann glaubwürdig, wenn erwartet werde, dass alle Aussagen dieser Partei für die andere Partei als korrekt angesehen werden können. Parteien sehen die Aussagen von anderen Parteien eher als korrekt an, wenn Sie glauben, dass die andere Partei den eigenen Prinzipien folgt. Aufgrund dessen ordnen wir Glaubwürdigkeit als Aspekt der Integrität ein und folgen ähnlichen Einordnungen aus der Literatur. So ist auch McKnight et al. (1998) der Ansicht, dass die „Integrität [...] sich konzeptuell mit der [...] Glaubwürdigkeit“ überschneidet (eigene Übersetzung aus McKnight et al., 1998, S. 6). Es existiert nur begrenzte Forschung zur Wirkung der algorithmischen Herkunft von Texten auf das Vertrauen in den Erzeuger. Es konnten keine Arbeiten über den Einfluss algorithmischer Herkunft auf Vertrauen oder vertrauensähnliche Konstrukte für andere Einsatzgebiete generativer KI außerhalb des Journalismus gefunden werden. Im journalistischen Bereich existieren einige Arbeiten, die sich mit dem Einfluss der algorithmischen Herkunft auf die Glaubwürdigkeit eines Erzeugers oder seiner Erzeugnisse befassen (Graefe und Bohlken, 2020). Diese sollten somit aufgrund der erläuterten Überschneidung von Glaubwürdigkeit und Integrität Anwendung auf Diskussionen über die Integrität von Verfassern finden.

Erste Arbeiten in diesem Bereich befassen sich mit der Wahrnehmung von automatisiert erzeugten Texten, die wenig komplex und hochstrukturiert sind. Dazu zählen beispielsweise Zusammenfassungen von Quartalsberichten börsennotierter Unternehmen oder Sportnachrichten (Graefe, 2016). Die Glaubwürdigkeit dieser automatisiert erzeugten deterministischen

Texte sowie den dahinterliegenden Algorithmen sei wohl gleichauf, wenn nicht teilweise höher, als Menschengeschriebenes (Wölker und Powell, 2018). Dies widerspricht der grundsätzlichen Aversion von Menschen gegenüber Algorithmen und Ihren Erzeugnissen (Dietvorst et al., 2015). Eine mögliche Begründung von Wölker und Powell (2018) ist die versprochene algorithmische Objektivität: Deterministische Algorithmen haben keine Meinung und garantieren eine objektive Berichtserstattung. In der Literatur wird dieser Effekt als Maschinenheuristik beschrieben (Sundar, 2008). Andere Studien (Graefe et al., 2018) fanden währenddessen eine geringere Glaubwürdigkeit für KI als Tatsächlicher Verfasser (TV). In einer systematischen Literaturübersicht betrachteten Graefe und Bohlken (2020) die Wahrnehmung von journalistischen Artikeln menschlicher und algorithmischer Herkunft in Bezug auf die Faktoren Glaubwürdigkeit, Qualität und Lesbarkeit. Die Ergebnisse der Metaanalyse unterstützen grundsätzlich die Ergebnisse von Wölker und Powell (2018) insofern, dass kein Unterschied in der wahrgenommenen Glaubwürdigkeit zwischen Texten menschlicher und algorithmischer Herkunft festgestellt wurde. Ein interessanter Effekt, der in der Metaanalyse beobachtet wurde, ist eine erhöhte Glaubwürdigkeit für Texte, bei denen Teilnehmende angegeben wurde, dass der vorliegende Text von einem Menschen entstammt, unabhängig von dem TV. Die Ergebnisse widersprechen also anscheinend der Maschinenheuristik nach Sundar (2008) in Bezug auf Glaubwürdigkeit. Um den von Graefe und Bohlken (2020) beschriebenen Effekt auch für generative KI und das Zielkonstrukt ZB zu prüfen, wird der TV als Kontrollvariable in das Forschungsmodell mit aufgenommen.

Die Erzeugung der Texte aus der genannten Literatur erfolgt primär deterministisch und ist somit von den Erzeugnissen generativer KI abzugrenzen, welche in der Lage sind, nichtdeterministische, hochkomplexe Texte mithilfe diverser Deep Learning Methoden zu generieren (Lim et al., 2023). Aufgrund der Neuheit dieser generativen Modelle existiert noch wenig Literatur über die wahrgenommene Glaubwürdigkeit von nichtdeterministischen, komplexen Texten. Longoni et al. (2022) beschreibt eine geringere Glaubwürdigkeit von Nachrichten, die von generativer KI erzeugt wurden, während Henestrosa et al. (2023) keinen Unterschied in der Glaubwürdigkeit feststellen konnten. Im Bereich der generativen KI existiert also kein Konsens über die Auswirkung von generativer KI als Erzeuger auf die Glaubwürdigkeit. Wir folgen deswegen den Ergebnissen von Graefe und Bohlken (2020) und erwarten eine geringere Glaubwürdigkeit für Verfasser, wenn der Angegebener Verfasser (AV) eine generative KI ist. Durch die bereits erläuterte Überschneidung der Integrität und Glaubwürdigkeit ist es zu erwarten, dass der Integritätsaspekt der Vertrauensüberzeugungen von generativer KI einer ähnlich geringen Bewertung im Vergleich zu menschlichen Verfassern zugewiesen wird.

Die Angst, durch KI seine Anstellung zu verlieren, steigt seit Jahren (Fast und Horvitz, 2017) und hat sich durch die erhöhte mediale Aufmerksamkeit durch die Veröffentlichung von

ChatGPT erneut verschlimmert (EY, 2023). Diese Angst könnte das Ansehen generativer KI in Hinblick auf das Wohlfühlen gegenüber dem Menschen statistisch signifikant negativ beeinflussen. Durch die noch umstrittene Erkennbarkeit der Erzeugnisse generativer KI, ist nicht davon auszugehen, dass die Bewertung der Fähigkeit einer generativen KI, gute Texte zu schreiben, statistisch signifikant höher im Vergleich zu menschlichen Verfassern ausfallen wird. Aufbauend auf den Erwartungen der geringeren Vertrauensüberzeugungen Integrität und Wohlfühlen für algorithmische Erzeuger formulieren wir die Hypothese H_{3a} :

H_{3a} : Ist der angegebene Verfasser eines Textes eine künstliche Intelligenz, sinkt das Vertrauen in den Verfasser des Textes.

2.5.3 Vertrauen und Zahlungsbereitschaft

In einer systematischen Literaturübersicht schreiben O'Brien et al. (2020), dass „[...] keine quantitative Studie in [ihrem] Datensatz explizit die Beziehung zwischen Vertrauen und VZ, ZA oder ZB für digitalen Journalismus adressiert“ (eigene Übersetzung aus O'Brien et al., 2020, S.24). Der Einfluss psychologischer Faktoren, wie dem Vertrauen, auf die ZB ist in der wissenschaftlichen Literatur des digitalen Journalismus also unterentwickelt. In vielen anderen Forschungsgebieten lassen sich allerdings Belege für einen positiven Zusammenhang zwischen Vertrauen und ZB finden. Im Bereich des autonomen Fahrens (Rahimi et al., 2022), der Nanotechnologie (Roosen et al., 2015), Ernährungswissenschaften (Nocella et al., 2010; Aksoy und Özsönmez, 2019) sowie der Medizin (Meng et al., 2020) existieren quantitative Studien, die einen statistisch signifikanten positiven Zusammenhang zwischen Vertrauen und ZB belegen. Aufbauend darauf formulieren wir die Hypothese H_{3b} :

H_{3b} : Ist das Vertrauen in den Verfasser eines Textes hoch, so steigt die Zahlungsbereitschaft für den Text.

3 Experimentdesign

Um die Hypothesen empirisch zu testen, wurde ein online Experiment durchgeführt. In dem Experiment wurden die Teilnehmenden Texten von diverser Herkunft ausgesetzt. Die Texte sind dabei sowohl von generativer KI erzeugt worden, wie von menschlichen Verfassern verfasst worden. Bei dem Experiment handelte es sich um ein *Within-Subject* Design. Das bedeutet, dass Teilnehmende des Experimentes mehr als ein Setup durchliefen (Charness et al., 2012). Zwar bringen Within-Subject Designs diverse Probleme mit sich (Charness et al., 2012), dennoch erachten wir den Vorteil der erhöhten Teststärke von Within-Subject-Designs als bedeutsam. Kahnemann und Ritov (1994) untersuchten die ZB für öffentliche Projekte in einem Within-Subject Design und führten eine Korrelationsanalyse zwischen

der ZB und Fragennummer durch. Sie kamen zu dem Schluss, dass Teilnehmende in der Lage sind, die zu bewertenden Sachverhalte unabhängig zu betrachten ($r = -0.001$). Da dieses Experiment mit weniger Setups als Kahnemann und Ritov (1994) operierte, gingen auch wir davon aus, dass keine statistisch signifikante Korrelation zwischen der Fragennummer und der ZB existierte. Dennoch randomisierten wir die Reihenfolge der jeweiligen Setups, um einer möglichen Korrelation entgegenzuwirken. Das Within-Subject Design ermöglichte zusätzlich eine Komplexitätsreduzierung des Forschungsmodells, indem die Vertrauensneigung als mediiierendes Konstrukt auf das Gesamtvertrauen vernachlässigt werden konnte. Auswirkungen von überdurchschnittlich hoher bzw. niedriger Vertrauensneigung spiegelten sich durch das Within-Subject Design in allen Datensätzen eines Teilnehmenden wider. Da für die Auswertung primär der Unterschied zwischen dem Vertrauen in zwei AV (Mensch und generative KI) untersucht wurde, wirkte sich die Vertrauensneigung auf alle Datensätze gleichmäßig aus. Unterschiede in das Vertrauen in menschliche Verfasser und KI Erzeuger sollten somit primär mit den wahrgenommenen Vertrauensüberzeugungen des Vertrauensnehmers begründet werden können.

Der Ablauf des Experimentes ist in Abbildung 3 abgebildet.

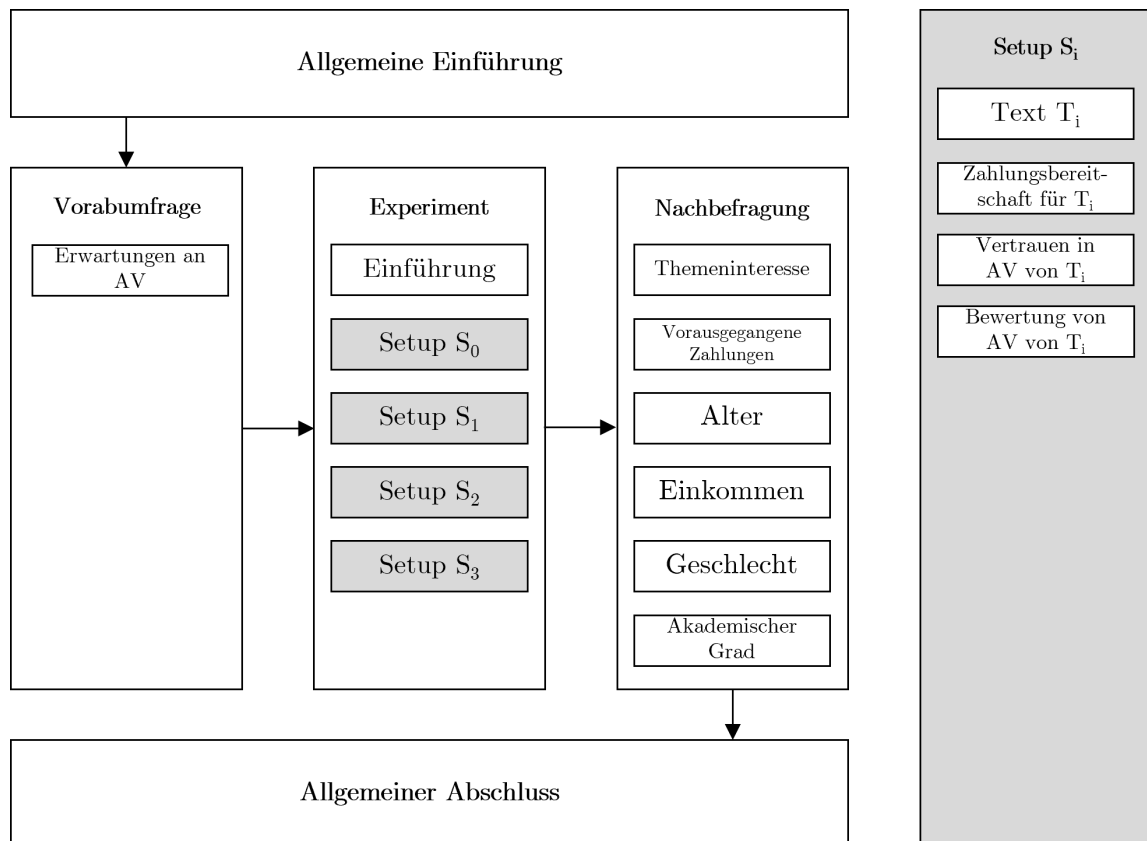


Abbildung 3: Experimentablauf

3.1 Auswahl der Teilnehmenden

Um die benötigte Anzahl an Teilnehmenden für das Experiment zu berechnen, verwendeten wir die Software G*Power (Faul et al., 2009). Aus den Angaben in Tabelle 1 lässt sich ableiten, dass wir für unsere Untersuchung im Rahmen der „Linearen multiplen Regression: Festes Modell, R²-Anstieg“ eine Mindestanzahl von 77 Teilnehmenden benötigten. Da aber alle Teilnehmenden exakt vier Setups durchliefen und somit vier Datenpunkte lieferten, ließ sich die Anzahl an Teilnehmenden unter der Annahme, dass die Beobachtungen unabhängig voneinander sind, auf 20 Teilnehmende³ reduzieren (Ratray et al., 2011). Allerdings sind die Daten einzelner Teilnehmenden in der Realität nicht vollständig unabhängig voneinander. Daher berücksichtigten wir auch die Ähnlichkeit der Daten innerhalb einzelner Teilnehmenden, indem wir einen durchschnittlichen Korrelationskoeffizienten von 0,3 ansetzten (Rutterford et al., 2015). Diese Annahme führte zu einer adjustierten erforderlichen Stichprobengröße von 37 Teilnehmenden⁴.

Variablen	Werte
Alter	KV
Einkommen	KV
Themeninteresse	KV
Vorausgegangene Zahlungen	KV
Tatsächlicher Verfasser	KV
Zahlungsbereitschaft	AV
Vertrauen	UV
Widerlegung von Vermutungen	UV
KI ist angegebener Verfasser	UV
Effektgröße f	.15
Signifikanzniveau α	.05
Teststärke $1 - \beta$.8
Anzahl getesteter Prädiktoren	3
Anzahl Prädiktoren gesamt	8
Stichprobengröße	77
Datenpunkte pro Teilnehmer	4
Intra-cluster Korrelation	.3
Teilnehmeranzahl	37

Tabelle 1: Ermittlung der benötigten Teilnehmeranzahl

Die Auswahl der Teilnehmenden erfolgte hauptsächlich mittels direkter Ansprache über diverse Kommunikationskanäle, unter anderem WhatsApp, LinkedIn und persönliche Kommunikation. Um den Teilnehmendenkreis möglichst divers zu gestalten, wurden vor allem gezielt

³Stichprobengröße geteilt durch die Anzahl der Datenpunkte pro Teilnehmender und aufgerundet:
 $77/4 = 19,25 \rightarrow 20$

⁴Stichprobengröße $\times (1 + (\text{Datenpunkte pro Teilnehmer/in} - 1) \times \text{intra-cluster Korrelation})$ aufgerundet:
 $19,25 \times (1 + (4 - 1) \times 0.03) = 36,575 \rightarrow 37$

Personen aus dem direkten Bekanntenkreis mit unterschiedlichen demografischen Kriterien angesprochen. Da eine Teilnahme an dem Experiment gute Kenntnisse in der deutschen Sprache voraussetzte, mussten wir bei der Auswahl sicherstellen, dass die Personen die deutsche Sprache beherrschen. Dementsprechend wurden von uns hauptsächlich Kommiliton/innen und Familienmitglieder angesprochen, was die Diversität der Teilnehmenden beeinflussen könnte. Teilweise trugen die Teilnehmenden selbst zur Erweiterung der Reichweite bei, indem sie das Experiment in ihren eigenen Netzwerken an geeignete Personen weiterleiteten. Dies wird als *Snowball-Sampling* bezeichnet und ist eine häufig in der wissenschaftlichen Literatur anzutreffende Rekrutierungsmethode (Vedapradha et al., 2019, Habib und Zaidi, 2021) und war ein unerwarteter und wertvoller Aspekt der Rekrutierungsstrategie, da es zur Diversifizierung der Stichprobe beitrug. Vorherige Erfahrungen mit Systemen der generativen KI waren dabei nicht vorausgesetzt, da die Fragen zum Vertrauen in solche Systeme auch ohne einen initialen Umgang mit diesen beantwortet werden konnten. Da wir die Texte teilweise bewusst falsch gekennzeichnet haben (Kapitel 3.2), mussten wir sicherstellen, dass die Teilnehmenden keine Kenntnis über die verwendeten menschengeschriebenen Werke („Die Blutschule“ von Rhode (2016) und „Der Hammer der Götter“ von Hohlbein (2010)) hatten. Es wäre beispielsweise möglich, dass die Teilnehmenden das von Wolfgang Hohlbein geschriebene Werk „Der Hammer der Götter“ vorgelegt bekamen, aber als AV die fiktive KI ChatRev angegeben war. Wenn die Teilnehmenden das Werk von Hohlbein bereits kannten, könnte dies die Ergebnisse verzerren, da sie der KI potenziell Einfallslosigkeit oder Plagiarismus vorwerfen würden und dies in einer geringeren Ausprägung der Vertrauensneigungen oder der ZB äußern. Deshalb wurde am Ende des Experiments eine spezifische Abfrage durchgeführt, um zu ermitteln, ob die Teilnehmenden bereits mit diesen Werken vertraut waren. Damit wir sicherstellen konnten, dass die Teilnehmenden die verschiedenen Texte aufmerksam lasen und die Fragen zu den Texten nicht zufällig beantwortet wurden, haben wir in dem Experiment an bestimmten Stellen Fragen zur Überprüfung der Aufmerksamkeit eingebaut (Cabrera et al., 2023). So mussten die Teilnehmenden regelmäßig die erste Antwortoption auf einer Likert-Skala auswählen, um ihre Aufmerksamkeit zu validieren.

3.2 Gezieltes Falschinformieren

Die Teilnehmenden durchliefen während des Experiments vier Setups. In jedem der Setups bekamen die Teilnehmenden jeweils einen Text vorgelegt, sodass insgesamt zwei Texte KI-generiert und zwei Texte menschengeschrieben sind. Dabei blieb der TV während eines Kontextes (Belletristik oder Journalismus) konsistent. Hierbei wurden die Teilnehmenden jedoch bei jeweils einem der beiden Texte aus einem Kontext gezielt über den Verfasser des Textes falsch informiert. So wurde in dem KI-generierten Kontext einer der beiden Texte einem menschlichen Verfasser zugeordnet und in dem von Menschen verfassten Kontext einer der beiden Texte als KI-generiert gekennzeichnet. Für den falsch gekennzeichneten Text, der

menschlicher Hand entstammt, ist also der TV Mensch (TVM und das Gegenstück TVK) und der AV eine generative KI (AVK und das Gegenstück AVM)⁵. Da die teilnehmende Person über das Falschinformieren erst nach Vervollständigung des Experimentes informiert wurde, mussten zwangsläufig für beide Texte, über deren Herkunft die Teilnehmenden falsch informiert wurden, gelten, dass $AV \neq TV$. Die genaue Heuristik, wie ausgewählt wurde, in welchem Kontext der Benutzer falsch informiert wurde sowie in welcher Reihenfolge diese Texte erschienen, kann der Abbildung 4 entnommen werden. Um die ethische Unbedenklichkeit sicherzustellen, wurde ein Ethikantrag an der Universität Hohenheim gestellt. Teilnehmende wurden aus diversen Gründen falsch informiert. Wenn der technische Fortschritt seine jetzige Geschwindigkeit beibehält, wird die Qualität der erzeugten Texte von generativer KI sich in den nächsten Jahren wohl weiterhin drastisch verbessern. Indem wir die Teilnehmenden falsch informierten, konnten wir für diesen (möglichen) Qualitätsunterschied, der derzeit eventuell noch besteht, kompensieren und uns auf die Forschungsfrage und Titel, der sich mit dem Wert der menschlichen Arbeit als Ganzes befasst, fokussieren. Zudem erlaubte uns das falsche Informieren die Hintergründe des Vertrauens besser zu untersuchen. Das Schichtenmodell nach Lucassen und Schraagen (2012) definiert, wie bereits erläutert, die vier Schichten Vertrauensneigung, Vertrauen in Medium, Vertrauen in Quelle und Vertrauen in Information. Die ersten zwei Schichten hielten wir durch das Within-Subject-Design und ein gleichbleibendes Medium konstant. Das Vertrauen in die Information und die Quelle ließ sich nun über das falsch Informieren herausfinden. Ausgehend von einer ähnlichen Qualität ließ sich auf ein ähnliches Vertrauen in die Information schließen. So haben auch Lucassen und Schraagen (2012) das Vertrauen in die Information mithilfe von Artikeln, die einen Qualitätsunterschied aufwiesen, gemessen. War also kein signifikanter Effekt zwischen dem TV und der ZB vorhanden, ließ sich, indem wir Teilnehmende über den TV falsch informieren, auf eine gleichbleibende wahrgenommene Qualität der Texte schließen.

3.3 Auswahl der Texte

Wir verwendeten unterschiedliche Textarten, um die Robustheit des Experimentes sicherzustellen. Um sinnvolle Ergebnisse zu erhalten, haben wir Texte gewählt, die die folgenden Eigenschaften erfüllen:

1. Breites, allgemeines Interesse am Textinhalt. Wie bereits erläutert ist das Themeninteresse ein Einflussfaktor auf die ZB (O'Brien et al., 2020). Um möglichst aussagekräftige Ergebnisse zu erhalten, sollte der Inhalt des Textes möglichst viele Teilnehmende ansprechen.

⁵Aus Gründen der Verständlichkeit wird auch für Texte der generativen KI in diesem Kapitel der Verfasserbegriff bzw. die dementsprechende Abkürzung genutzt. Wie bereits erläutert, soll sich das nicht auf Verfasser bzw. Urheber nach UrhG beziehen.

2. Begrenzte Länge. Da es sich um ein Online-Experiment handelt, musste die Zeit der Teilnehmenden respektiert werden.
3. Keine Texte, die allgemein kostenlos sind. Die ZB für einen Wikipedia Artikel liegt bei den meisten Teilnehmenden wohl nahe 0 €, da diese Informationen frei an anderer Stelle verfügbar sind.

Mit Hinblick auf diese Kriterien haben wir uns entschieden, zwei Kontexte aufzunehmen: Journalismus (Auszüge aus Produktrezensionen von Smartphones) und Belletristik (Klappentext eines belletristischen Werkes). Bei der Produktrezension wurde auf eine künstliche Bezahlschranke gesetzt: Teilnehmende erhielten die ersten zwei Absätze einer Produktrezension gegeben, bevor ihre ZB und andere Einstellungen abgefragt wurden. Diese Art der Bezahlschranke, die einen Einblick in den Artikel gewährt, bevor eine Kaufentscheidung getroffen werden muss, ist im Journalismus häufig anzutreffen (Franklin, 2014) und ein Geschäftsmodell von Plattformen wie Stiftung Warentest, die einzelne Produkttests oder Testsammlungen zum digitalen Erwerb anbieten. Smartphones wurden als Rezensionsinhalt gewählt, da ein Großteil der Menschen in Deutschland eines besitzen (Bitkom e.V., 2023) und somit ein hohes thematisches Interesse zu erwarten war. Belletristische Werke werden von allen Büchern mit am häufigsten verkauft (Crosby, 2022), was bedeutet, dass auch hier ein hohes thematisches Interesse vorgelegen haben sollte. Die Kaufentscheidung für ein Buch wird dabei maßgeblich durch den Inhalt des Klappentextes bestimmt (Leitão et al., 2018), weswegen die Teilnehmenden die Klappentexte erhielten. GPT-4 ist bereits jetzt in der Lage, Bücher ohne menschlichen Inhaltsanweisungen zu schreiben (Coetzee, 2023). In einem HITL-Kontext sowie durch weiteren Fortschritt bezüglich der Tokenlimits diverser GPT-Modelle ist davon auszugehen, dass sich die Fähigkeit dieser Modelle, ganze Bücher zu schreiben, erneut verbessern wird. Deshalb betrachteten wir die Aufnahme dieses Kontextes als problemlos. Die Klappentexte von menschlichen Verfassern im Bereich der Belletristik wurden von bekannten Autoren übernommen: Es wurden die Klappentexte der Werke „Die Blutschule“ von Sebastian Fitzek, veröffentlicht unter dem Pseudonym Max Rhode, und „Der Hammer der Götter“ von Wolfgang Hohlbein übernommen. Für die journalistischen Texte wurden Produktrezensionen von den Smartphone-Modellen „Pixel 8 Pro“ von Kremp (2023) und „iPhone 14“ von Mansholt (2022) aus den Magazinen „Der Spiegel“ und „Stern“ übernommen. Die Texte der generativen KI wurden von ChatGPT mit dem unterliegenden Modell GPT-4 erzeugt. Die Texte der menschlichen Verfasser und die Texte der generativen KIs sowie die Anfragen an ChatGPT, die zur Erzeugung der Texte führten, können dem Anhang entnommen werden. Die Texte der generativen KI wurden nicht ChatGPT zugeschrieben, sondern zwei fiktiven generativen KIs „ChatRev“, eine generative KI, die Produktrezensionen erzeugen kann und „ChatLit“ eine generative KI, die belletristische Literatur erzeugen kann. Die Entscheidung, die Texte fiktiven KIs zuzuschreiben, betraf die Generalisierbarkeit des Experimentes. ChatGPT ist eine inzwischen bekannte Plattform, welche aber nicht repräsentativ für generative KI oder

LLMs als Ganzes betrachtet werden sollte. Das Experiment sollte nicht die ZB für Texte von ChatGPT untersuchen, sondern die ZB für generative KIs im Allgemeinen. Indem wir eine fiktive generative KI als Erzeuger angaben, konnte somit die ZB für generative KI als Ganzes anstelle der ZB für ChatGPT im Einzelnen untersucht werden.

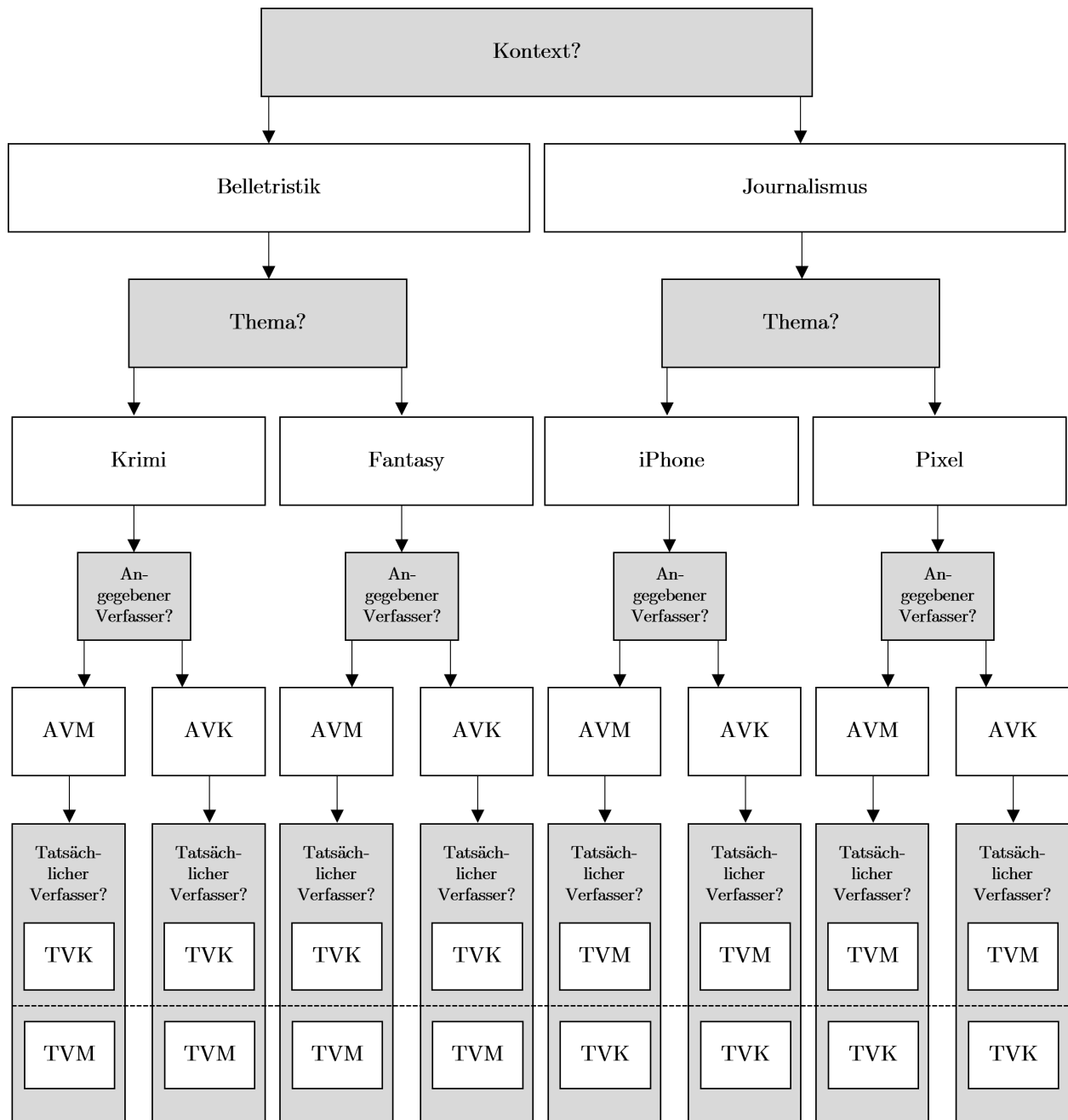


Abbildung 4: Entscheidungsbaum für das Experiment

Die Texte T_0 bis T_3 wurden für eine teilnehmende Person vor Eintritt in das Experiment in einem Schritt bestimmt. Die Auswahl der Texte erfolgte dabei anhand des in Abbildung 4 zu erkennenden Entscheidungsbaums. Dieser wurde viermal von Wurzel- bis zu einem der Blattknoten durchlaufen, wobei es für jeden ausgewählten Pfad einige Regeln zu beachten galt:

- **Regel 1:** Wurde ein bestimmter Pfad bereits für die Auswahl eines Textes abgelaufen, kann dieser nicht erneut ausgewählt werden. Dies stellt sicher, dass keine Texte doppelt gewählt werden.

- **Regel 2:** Der Kontext wird initial zufällig ausgewählt. Ist diese Auswahl getroffen, so wird der nächste Pfad zur Entscheidung über den folgenden Text ebenfalls innerhalb des ausgewählten Kontextes sein.
- **Regel 3:** Das Thema wird initial zufällig gewählt. Ist diese Auswahl getroffen, so wird der nächste Pfad zur Entscheidung über den folgenden Text über das andere Thema laufen.
- **Regel 4:** Der AV wird initial zufällig ausgewählt. Pro Kontext darf dabei jeder AV nur exakt einmal vorkommen. Dementsprechend wird für den Teilbaum unterhalb des Themas jeweils exakt ein Pfad gewählt.
- **Regel 5:** Nach der Entscheidung über den ersten TV werden die TV für alle anderen Setups fixiert. Es wird entsprechend der Grafik entweder das Szenario oberhalb der gestrichelten Linie oder das darunterliegende ausgewählt. Dies stellt sicher, dass eine teilnehmende Person in beiden Kategorien über den Ursprung von jeweils einem der Texte falsch informiert wird und wurde, wie in Abschnitt 3.3 angesprochen, gezielt in dieser Form gewählt.

Mit diesen Regeln ergeben sich insgesamt 64 verschiedene Kombinationen für die Texte T_0 bis T_3 :

- **Erster Schritt:** Wählen zwischen Belletristik und Journalismus. Da beide Pfade durchlaufen werden müssen und die Reihenfolge eine Rolle spielt, gibt es 2 Möglichkeiten (erst Belletristik dann Journalismus oder umgekehrt).
- **Zweiter Schritt:** Für jeden der beiden Pfade (Belletristik und Journalismus) muss die Entscheidung zwischen den beiden Themen gefällt werden. Diese Entscheidung muss für beide Pfade getroffen werden, was zu $2 \times 2 = 4$ Kombinationen führt.
- **Dritter Schritt:** Nach Regel 3 wird für jedes Thema nur ein Pfad gewählt. Die Entscheidung, welcher Pfad gewählt wird, muss für beide Themen getroffen werden, was zu $2 \times 2 = 4$ Kombinationen führt.
- **Vierter Schritt:** Nach Regel 5 wird nach der Wahl des ersten TV der TV für alle anderen Setups fixiert, was bedeutet, dass für den ersten Durchlauf die Entscheidung zwischen TVM oder TVK besteht. Für die nachfolgenden Durchläufe gibt es keine weiteren Optionen, da der TV bereits festgelegt ist. Dementsprechend berücksichtigen wir nur die Entscheidungen für den ersten Durchlauf.

Die Gesamtanzahl der Szenarien ergibt sich dann wie folgt:

$$\text{GesamtanzahlKombinationen} = \text{OptionenSchritt1} \times \text{OptionenSchritt2} \times \text{OptionenSchritt3} \times \text{OptionenSchritt4}$$

$$\text{GesamtanzahlKombinationen} = 2 \times (2 \times 2) \times (2 \times 2) \times 2$$

$$\text{GesamtanzahlKombinationen} = 2 \times 4 \times 4 \times 2$$

$$\text{GesamtanzahlKombinationen} = 64$$

3.4 Datenerhebung

Die Datenerhebung war dreigeteilt: Zuerst erfolgte eine Vorabumfrage, daraufhin wird das Experiment durchgeführt und zuletzt erfolgte eine Nachbefragung. Um mögliche beeinflussende Effekte zu vermeiden, wurden demografische und sonstige für die Setups nicht benötigte Fragen erst in der Nachbefragung gestellt. Für jedes Konstrukt wurden validierte Skalen aus vorheriger Forschung verwendet. Alle Konstrukte bis auf die ZB wurden als 7-Punkte Likert Skala abgefragt (1: Stimme stark zu, 7: Stimme gar nicht zu) und wenn möglich durch mehrere Fragen reflektiv gebildet. Der gesamte Fragenkatalog kann der Tabelle 4 des Anhangs entnommen werden. Für jeden Teilnehmenden war jede Frage, die erhalten wurde, verpflichtend, wodurch ein vollständiger Datensatz garantiert wurde.

3.4.1 Vorabumfrage

In der Vorabumfrage wurden Teilnehmende zuerst nach ihren Erwartungen an die generierenden Systeme ChatLit und ChatRev sowie an die menschlichen Verfasser befragt. Das Abfragen der Erwartung war dabei der erste Teil des Konstruktes WvA. Für beide Teile des Konstruktes wurden die Fragen von Kocielnik et al. (2019) übernommen und angepasst.

3.4.2 Experiment

Im eigentlichen Experiment durchliefen Teilnehmende, wie in Abbildung 3 beschrieben, vier Setups. Zu jedem Text erhielten die Teilnehmenden Fragen in absteigender Wichtigkeit in Bezug auf das Forschungsmodell. Dies diente der Minimierung von möglichen beeinflussenden Effekten auf die wichtigsten Konstrukte. Zuerst wurden Teilnehmende nach ihrer ZB für den Ihnen vorliegenden Text befragt. Die ZB wurde angelehnt an Marbeau (1987) direkt abgefragt. Dabei wurde versucht, die preisliche Obergrenze des Textes zu ermitteln, um die Definition nach Chyi und Yang (2009) zu erfüllen. Die ZB wurde als eine einzige reelle Zahl $0 < ZB < 999,99$ ermittelt. Die Teilnehmenden wurden zudem über die Durchschnittspreise für den Text informiert, da sich die tatsächliche ZB stark verändern kann, wenn die Teilnehmenden den durchschnittlichen Marktwert eines Produktes erfahren (Breidert et al., 2015). Indem Teilnehmende bereits vor Angabe ihrer ZB informiert wurden, versuchten wir diesem Effekt entgegenzusteuern. Die Durchschnittspreise für Produktrezensionen entstammten den Kosten einer Testsammlung auf der Seite test.de von Stiftung Warentest. Für die belletristischen Texte wurde der Durchschnittspreis des Jahres 2020 (Börsenverein des Deutschen Buchhandels, 2021) anhand der durchschnittlichen Inflationsrate bis November

2023 transponiert (Statistisches Bundesamt, 2024), um einen Schätzwert für den Buchpreis im Jahr 2023 zu erhalten. Das Vertrauen in den AV wurde ebenso mit angepassten Fragen von Li et al. (2008) ermittelt. Für den Fall, dass sich das formative Vertrauensmodell nach Mayer et al. (1995) mit den angepassten Fragen von Li et al. (2008) durch statistischen Auswertung nicht als angemessenes Higher-order Konstrukt herausstellen sollte, wurden zusätzlich die reflektiven Fragen zu Vertrauen von Söllner et al. (2012) übernommen, die im Falle eines invaliden Higher-order Konstrukts für die weitere statistische Auswertung der Hypothesen anstelle des Vertrauensmodells nach Mayer et al. (1995) genutzt werden sollten. Zuletzt wurde die Bewertung der Performanz des AV zur Berechnung der WvA abgefragt.

3.4.3 Nachbefragung

In der Nachbefragung mussten Teilnehmende einige demografische Informationen angeben sowie einige Fragen zu den Kontrollvariablen beantworten. Um beeinflussende Effekte der demografischen Variablen zu vermeiden, wurden diese hinter den Fragen zu den Kontrollvariablen angestellt. Zuerst wurde das Themeninteresse an beiden Kontexten abgefragt, wozu die Fragen von Schiefele (1990) übernommen wurden. Zudem wurden für beide Kontexte die VZ der vergangenen 12 Monate abgefragt. Dafür wurden die Fragen von O'Brien et al. (2020) übernommen. Um eine angemessene Stichprobenverteilung zu gewährleisten und mögliche demografische Einflussfaktoren auf die ZB nach O'Brien et al. (2020) abzubilden, wurden Fragen zum Alter, Geschlecht und dem monatlichen Haushaltseinkommen (klassiert in 1000€-Schritten) gestellt.

Eine Gesamtübersicht über die erhobenen Daten kann Tabelle 2 entnommen werden.

Variable	Messung	Wertebereich
Unabhängige Variablen: 4 Datenpunkte, 1 Punkt pro Setup		
Tatsächlicher Verfasser	Manipulation	{Mensch, KI}
Angegebener Verfasser	Manipulation	{Mensch, KI}
Setups: 4 Datenpunkte, 1 Punkt pro Setup		
Erwartungen an Autor/Erzeuger	1 Datenpunkt	$[1; 7] \subset \mathbb{N}$
Zahlungsbereitschaft	1 Datenpunkt	$[0; 999, 99]^1 \subset \mathbb{Q}$
Fähigkeit des Autors/Erzeuger	4 Datenpunkte, CFA	$[1; 7] \subset \mathbb{N}$
Wohllollen des Autors/Erzeuger	3 Datenpunkte, CFA	$[1; 7] \subset \mathbb{N}$
Integrität des Autors/Erzeuger	4 Datenpunkte, CFA	$[1; 7] \subset \mathbb{N}$
Vertrauen in Autor/Erzeuger	3 Datenpunkte, CFA	$[1; 7] \subset \mathbb{N}$
Bewertung des Autor/Erzeuger	1 Datenpunkt	$[1; 7] \subset \mathbb{N}$
Vor- und Nachbefragung: 1 Datenpunkt pro Variable und Teilnehmer		
Fähigkeit generativer KI	4 Datenpunkte, CFA	$[1; 7] \subset \mathbb{N}$
Wohllollen generativer KI	3 Datenpunkte, CFA	$[1; 7] \subset \mathbb{N}$
Integrität generativer KI	4 Datenpunkte, CFA	$[1; 7] \subset \mathbb{N}$
Vertrauen in generativer KI	3 Datenpunkte, CFA	$[1; 7] \subset \mathbb{N}$
Themeninteresse Produktrezension	7 Datenpunkte CFA	$[1; 7] \subset \mathbb{N}$
Themeninteresse Belletristik	7 Datenpunkte CFA	$[1; 7] \subset \mathbb{N}$
VZ Produktrezension	1 Datenpunkt	$[0; 1] \subset \mathbb{R}$
VZ Belletristik	1 Datenpunkt	$[0; 1] \subset \mathbb{N}$
Alter	1 Datenpunkt	$[18; 102] \subset \mathbb{N}$
monatliches Nettoeinkommen	1 Datenpunkt	$\{0 - 999, 1000 - 1999, 2000 - 2999, 3000 - 3999, 4000 - 4999, 5000 - 5900, > 6000\}^1$
Geschlecht	1 Datenpunkt	{Männlich, Weiblich, Andere, Keine Angabe}
Akademischer Grad	1 Datenpunkt	{Abitur, Ausbildung, Bachelor, Master, Doktor, Sonstige}

1: Alle Geldbeträge in Euro

Tabelle 2: Übersicht über die erhobenen Daten

4 Ergebnisse

4.1 Grundlagen der multiplen linearen Regression

4.1.1 Einführung in die multiple lineare Regression

4.1.2 Annahmen der multiplen linearen Regression

Die Resultate einer OLS-Regression können in ihrer Aussagekraft vermindert sein, wenn nicht gewisse Annahmen über die zugrundeliegenden stetigen Variablen erfüllt sind (Osbourne und Waters, 2002). Osbourne und Waters (2002) beschreiben vier Annahmen, die bei einer OLS-Regression getestet werden sollten:

1. Normalverteilung der Variablen. Sollte bei Variablen keine Normalverteilung vorliegen, kann dies Beziehungen und Signifikanztests beeinflussen. Die Normalverteilung kann entweder durch statistische Tests wie dem Kolmogorov-Smirnov Test sichergestellt werden oder durch Inspektion von Histogrammen sowie Schiefe und Kurtosis der Verteilung evaluiert werden.
2. Lineare Beziehung zwischen den unabhängigen und abhängigen Variable(n). Sollte zwischen den unabhängigen Variablen und den abhängigen Variablen keine lineare Abhängigkeit vorliegen (z.B: exponentiell), unterschätzt die lineare Regression den Effekt der Beziehung. Nichtlineare Beziehungen sind unter Umständen bereits in der Literatur bekannt oder können über statistische Analyseverfahren herausgefunden werden.
3. Verlässliche, fehlerfreie Messungen. Viele Variablen sind nur schwer operationalisierbar bzw. schwer messbar. Unzuverlässige Messungen können dazu führen, dass Beziehungen unterschätzt werden. Die interne Zuverlässigkeit einer Confirmatory Factor Analysis (CFA) kann über ein valides (Heck und Thomas, 2015; Hu und Bentler, 1999) Modell angenommen werden.
4. Homoskedastizität der Variable. Die Homoskedastizität beschreibt, dass die Varianz der Residuen über alle Werte der (unabhängigen) Variablen konstant sein sollte. Ist das nicht gegeben, spricht man von Heteroskedastizität. So ist leichte Heteroskedastizität zwar noch unproblematisch, höhere Werte können jedoch zu ungenaueren statistischen Tests und somit geringerer Aussagekraft dieser führen. Über die Prüfung der Residuenverteilung oder durch statistische Tests wie den Goldfeld-Quandt Test kann eine mögliche Heteroskedastizität geprüft werden.

Darüber hinaus existieren zudem noch Annahmen, die entweder bei Verstoß keinen signifikanten Einfluss auf die Aussagekraft der Regression haben (z. B. Normalverteilung

der Fehlerwerte) oder durch gutes Studiendesign nicht anfallen (z. B. Unabhängigkeit der Beobachtungen) (Osbourne und Waters, 2002). Zwar sind die Annahmen bei Nichterfüllung korrigierbar, dennoch ist dies oft mit hohem Aufwand und ungenaueren Ergebnissen verbunden (Osbourne und Waters, 2002), weswegen sich einige statistische Ansätze entwickelt haben, um diese Einschränkungen der OLS-Regression zu korrigieren. Der Schätzer von Huber (1964) sowie weitere Anpassungen des Maximum-Likelihood-Schätzers z. B. nach Yuan und Bentler (1998) erlaubt es, lineare Regressionen auf Variablen anzuwenden, die u.U nicht normalverteilt sowie nicht homoskedastisch sind. Dieser Schätzer wird in der Literatur als Maximum Likelihood Robust (MLR) Schätzer beschrieben (Li, 2016). Dieser ist in diversen Statistikprogrammen wie Mplus eingebaut (Muthén und Muthén, 1998-2010) und wird in den nachfolgenden Analysen genutzt. Da der MLR-Schätzer genutzt wird, muss keine Prüfung der Variablen auf Normalverteilung oder Homoskedastizität vorgenommen werden. Die Linearitätsannahme setzt stetige Variablen voraus. Für die ZB ist dies gegeben. Für die Items, die mithilfe der 7-Punkte Likert Skala abgefragt wurden, folgen wir den Empfehlungen von Harpe (2015) und behandelten diese als stetig, da die Likert-Skala 7 Punkte enthält und dies somit den von Harpe (2015) definierten Grenzwert von 5 überschreitet. Die eigentliche Linearitätsannahme konnte vor der Regression nicht geprüft werden, da diese Frage im Rahmen dieser Arbeit beantwortet werden soll. Wir prüften die interne Zuverlässigkeit der Multi-Item-Variablen, indem wir das entstehende Modell auf Validität prüften (Kapitel 4.2). Wir nahmen eine externe Validität an, da alle relevanten Items der Literatur entstammen. Der gesamte Fragenkatalog kann Tabelle 4 des Anhangs entnommen werden.

4.1.3 Interpretation der Regressionskoeffizienten

Bei der multiplen linearen Regression muss nicht jeder Regressionskoeffizient statistisch signifikant sein. Nach Keith (2006) wird gewöhnlich ein Einstichproben-t-Test durchgeführt, um die Nullhypothese „Der Koeffizient des Regressors ist Null“ zu verwerfen. Um den t-Wert zu erhalten, teilt man den Regressionskoeffizienten des zu prüfenden Regressors durch seinen Standardfehler (SE), also $t = \frac{b}{SE_b}$. Danach vergleicht man den berechneten t-Wert mit dem kritischen Wert aus der t-Verteilung. Der kritische Wert hängt dabei vom gewählten Signifikanzniveau (üblicherweise 0,01, 0,05 oder 0,10) und den Freiheitsgraden der Verteilung ab, die sich aus der Anzahl der Beobachtungen minus der Anzahl der unabhängigen Variablen minus 1 berechnen. Liegt der berechnete t-Wert über dem kritischen Wert, kann die Nullhypothese verworfen werden und der Regressionskoeffizient wird als statistisch signifikant angesehen.

Wichtig zu betonen ist hierbei allerdings, dass die statistische Signifikanz nicht gleich bedeutet, dass der Regressor auch in der Praxis relevant sein muss. Der Regressionskoeffizient kann so klein sein, dass die praktische Anwendung davon nichtig ist (Hanson et al., 1986, Erb, 1990).

4.1.4 Vergleich mit SEM und Begründung der Methodenwahl

Eine andere häufig anzutreffende Methode zur Validierung von mehreren Hypothesen sind Strukturgleichungsmodelle. Die Strukturgleichungsmodellierung (Structural Equation Modeling, kurz SEM) ist „ein multivariater statistischer Analyserahmen, der zur Modellierung komplexer Beziehungen zwischen direkt und indirekt beobachteten (latenten) Variablen verwendet wird“ (eigene Übersetzung aus Stein et al., 2012, S.1). Latente Variablen sind dabei Konstrukte, wie in dieser Arbeit das Vertrauen (Söllner et al., 2010), die nicht mittels einer direkten Messung bewertet werden können. Sie setzen sich meist aus mehreren beobachtbaren Variablen zusammen. In unserem Fall wurde versucht, das Vertrauen über die Indikatoren Fähigkeit, Wohlwollen und Integrität zur erklären, welche durch das Stellen bestimmter Fragen messbar sind (Li et al., 2008). SEM wird in der Literatur häufig als leistungsfähiger gegenüber der multiplen linearen Regression betrachtet (Nusair und Hua, 2010, Stein et al., 2012), da sie die Modellierung von Messfehlern sowohl für den Prädiktor als auch für das Ergebnis ermöglicht und ein hohes Maß an Flexibilität bei der Modellierung der Korrelation zwischen den verschiedenen Fehlertermen bietet (Stein et al., 2012).

Allerdings ist SEM statistisch deutlich komplexer und bedarf höheren Anforderungen an die erforderliche Stichprobengröße (Wolf et al., 2013) als die multiple lineare Regression, um signifikante Ergebnisse liefern zu können. Zusätzlich steigt bei SEM die Komplexität bei wiederholten Messungen (Deng und Yuan, 2015). Da unser Datensatz nur über eine Stichprobengröße von 37 Teilnehmenden verfügte und wiederholte Messungen enthielt, ließ sich dieser in SEM nur mit erheblichem zusätzlichem Aufwand korrekt modellieren. In Anbetracht dieser Einschränkungen lieferte die multiple lineare Regression eine robuste und ressourcenschonendere Alternative. Zusätzlich erzeugte die multiple lineare Regression Resultate, die einfacher zu interpretieren sind.

4.2 Deskriptive Analysen und Datenbeschreibung

An dem Experiment haben insgesamt 89 Personen teilgenommen. Aufgrund unvollständiger Antworten durch vorzeitiges Abbrechen des Experimentes sowie Falschbeantwortung von diversen Kontrollfragen in dem Experiment erhielten wir $38 \times 4 = 152$ Experimentreihen von 38 Teilnehmenden. Eine Analyse der Antworten fand für eine teilnehmende Person starke Ausreißerwerte für die ZB für Produktrezensionen. In den Produktrezensionen wurden Smartphones thematisiert. Die angegebene ZB der teilnehmenden Person von 500 € würde sich mit dem Verkaufspreis eines neuen Smartphones decken. Die ZB dieser Person für Belletristik war wieder in einem angemessenen Bereich, weswegen wir davon ausgingen, dass die Person die Frage zu den Produktrezensionen falsch verstanden hat und wir die Experimentreihen dieser Person herausfilterten. Keine teilnehmende Person kannte die Werke von Max Rhode und Wolfgang Hohlbein bereits vor Teilnahme an dem Experiment, weswegen dadurch keine

weiteren Personen herausgefiltert wurden. Damit ergab sich eine valide Stichprobe von 37 sowie einem Datensatz von $37 \times 4 = 148$ Einträgen. Die genaue Filterung der validen Stichprobe kann der Abbildung 5 entnommen werden.

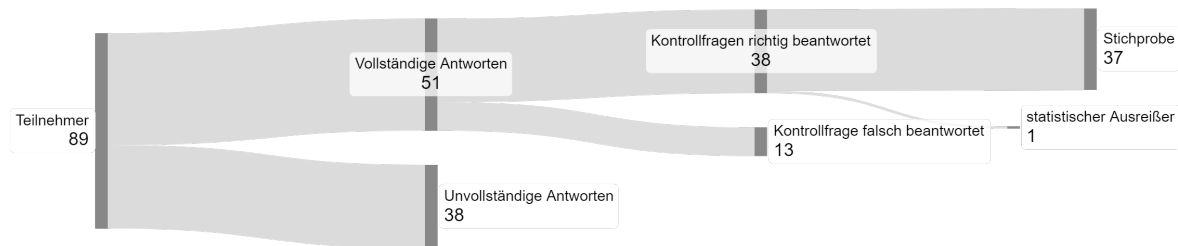


Abbildung 5: Auswahl der Teilnehmenden

Die erhaltene Stichprobe von 37 erfüllte somit die errechnete Mindeststichprobengröße von 37, welche in Kapitel 3.1 errechnet wurde. Um einen möglichen Effekt der Fragenreihenfolge auf die ZB zu prüfen, wurde die Korrelation sowie eine mögliche statistische Signifikanz dieser Korrelation, zwischen der Fragennummer und ZB geprüft. Sofern nicht anders gekennzeichnet, wird in dieser Arbeit grundsätzlich das Fisher Kriterium (Fisher, 1925) von $p < 0.05$ zur Ablehnung der Nullhypothese verwendet, welches sich in der Literatur als Standard etabliert hat (Cowles und Davis, 1982). Aufgrund der großen Anzahl von Personen, die das Experiment abgebrochen haben, sowie dem Falschbeantworten von Kontrollfragen, konnten keine ausgeglichenen Randomisierungsgruppen mit Hinblick auf die kontextliche Reihenfolge der Texte garantiert werden⁶. Das war für diese Untersuchung problematisch, da sich die ZB für den journalistischen Text und Belletristik, wie zu erwarten, statistisch signifikant unterscheidet hat (Zweistichproben gepaarter t-test: $p < 2.155 \times 10^{-15}$). Eine Korrelationsanalyse hätte demnach diesen statistisch signifikanten Unterschied gefunden, anstatt die Korrelation zwischen der Fragennummer und der ZB zu betrachten. Aufgrund dessen wurde die Korrelation zwischen der Fragennummer und ZB für die beiden Kontexte Belletristik und Journalismus separat durchgeführt. Sowohl für die Belletristik ($r = -.181$, $p = 0.1214$) als auch den Journalismus ($r = 0.018$, $p = 0.8786$) konnte keine statistisch signifikante Korrelation durch den Pearson-Korrelationstests gefunden werden, was die Ergebnisse von Kahnemann und Ritov (1994) stützt. Um einen möglichen Effekt der Robustheitskriterien auszuschließen,

⁶25 Teilnehmende erhielten zuerst Belletristik, während nur 12 zuerst den journalistischen Text erhielten.

wurde geprüft, ob die Robustheitskriterien einen statistisch signifikanten Einfluss auf die Zielgröße ZB hatten. Hierfür wurden zwei Zweistichproben gepaarte t-Tests durchgeführt, um einen statistisch signifikanten Unterschied zwischen der ZB für die jeweiligen Textquellen auszuschließen. Weder bei den belletristischen Klappentexten ($p = 0.4057$) noch bei den Produktrezensionen ($p = 0.8754$) wurde ein statistisch signifikanter Unterschied in der ZB gefunden. Der p-Wert der belletristischen Klappentexte war vermutlich geringer, da auch hier eine ungleiche Verteilung der Randomisierungsgruppen mit Bezug auf den AV vorliegt⁷, wodurch der Effekt des AV einen Einfluss auf die ZB hatte.

x	Messung	\bar{x}			σ_x			γ
		Gesamt	Belletristik	Journalismus	Gesamt	Belletristik	Journalismus	
Between								
Akademiker	<i>direkt</i>	0.32			0.47			
Geschlecht	<i>direkt</i>	0.35			0.48			
Alter	<i>direkt</i>	27.30			10.61			
Einkommensgruppe	<i>direkt</i>	1.49			1.59			
Vertrauen in generative KI	<i>multi-item</i>	4.09			1.23			
	GT1	4.19			1.21			0.829
	GT2	3.97			1.48			1
	GT3	4.11			1.41			0.945
Within								
ZB	<i>direkt</i>	9.35	14.61	4.09	8.67	8.73	4.36	
WvA	<i>direkt</i>	0.13	0.08	0.18	1.28	1.20	1.37	
VZ	<i>direkt</i>	0.36	0.57	0.16	0.48	0.50	0.37	
Vertrauen in den Autor	<i>multi-item</i>	4.22	4.17	4.28	1.40	1.26	1.52	
	T1	4.16	4.18	4.14	1.47	1.34	1.60	0.922
	T2	4.13	4.05	4.20	1.55	1.43	1.67	1
	T3	4.38	4.27	4.49	1.41	1.35	1.47	0.848
Interest	<i>multi-item</i>	4.37	4.56	4.18	1.17	1.19	1.12	
	I1	4.59	4.76	4.43	1.67	1.77	1.56	0.728
	I2	4.16	5.03	3.30	1.65	1.39	1.42	0.787
	I3	4.84	4.89	4.78	1.43	1.44	1.43	1
	I4	3.65	3.89	3.41	1.39	1.38	1.37	0.750
	I5	4.61	4.22	5.00	1.50	1.52	1.38	0.806

Tabelle 3: Deskriptive Statistiken und CFA der Daten

Die Werte für γ weisen alle ein $p < 0.01$ auf.

Tabelle 3 zeigt die deskriptiven Statistiken und die Ergebnisse der CFA. Die Between-Variablen waren dabei nur zwischen den einzelnen Teilnehmenden verschieden und innerhalb eines Teilnehmenden konsistent. Analog dazu waren die Within-Variablen innerhalb eines Teilnehmenden und zwischen den Teilnehmenden verschieden. Die Variable „Akademiker“ war dabei binär und beschriebte, ob eine teilnehmende Person einen akademischen Grad erlangt

⁷25 Teilnehmende erhielten die Pixel 8 Pro Produktrezension, bei welcher die generative KI der AV ist, während nur 12 die iPhone 14 Produktrezension erhielten, bei welcher die generative KI der AV ist.

hat (1) oder nicht (0). Als akademischer Grad waren die Abschlüsse Bachelor, Master und Diplom kodiert. Da es keine Teilnehmenden gab, die als Geschlecht „divers“ angegeben haben, wurde Geschlecht als binäre Variable mit den Werten 1 für „weiblich“ und 0 für „männlich“ repräsentiert. Die Einkommensgruppen aus Tabelle 2 wurden jeweils zu den Zahlen 0 bis 6 kodiert. Hierbei repräsentierte 0 die unterste Einkommensgruppe (0 € - 999 €) und 6 die höchste Einkommensgruppe (> 6000 €). Die ZB wurde von den Teilnehmenden in Euro angegeben. WvA und VZ wurden bereits in Kapitel 2.4 sowie 2.2.4 diskutiert. Die multi-item Variablen „Vertrauen in generative KI“, „Vertrauen in den Verfasser“ und „Themeninteresse“ wurden mithilfe einer CFA validiert, die eine in der Literatur häufig anzutreffende Methode zur Validierung latenter Variablen ist (Bayer et al., 2022; Credé und Harms, 2015). Diese prüft mit einem Signifikanztest und Modellgüteindizes, ob die beobachteten Daten ein vorher spezifiziertes Modell stützen oder nicht und mit welcher Güte sie dieses Modell stützen (Bühner, 2006, S. 263). In einem ersten Schritt wurden die Items aus der weiteren Analyse entfernt, die keinen signifikanten Einfluss auf das jeweilige Konstrukt aufwiesen (Hair et al., 2019)⁸. Da Vertrauen über die reflektiv gemessenen Konstrukte Fähigkeit, Wohlwollen und Integrität formativ nach Mayer et al. (1995) gebildet werden sollte, galt es diese nun in der nächsten CFA-Iteration zu prüfen. Trotz des Entfernens der insignifikanten Variablen wurden die Grenzwerte für ein valides Modell nach Heck und Thomas (2015) sowie Hu und Bentler (1999) ($CFI < 0.9$; $TLI > 0.9$; $SRMR < 0.08$; $RMSEA < 0.08$) nicht erfüllt, da sowohl $CFI = 0.762$, $TLI = 0.729$ und $SRMR = 0.091$ die aufgestellten Gütekriterien verletzen. Um die geringen Werte zu kompensieren, wurde das Vertrauensmodell nach Mayer et al. (1995) zum Messen des Vertrauens in den AV sowie in generative KI als Ganzes durch die reflektiven Items zu Vertrauen nach Söllner et al. (2012) ersetzt. Die bereits ermittelten invaliden Items Interest6 und Interest7 wurden in dem neuen Modell nicht weiter berücksichtigt. Das neue Modell erfüllte die Gütekriterien ($CFI = 0.922$; $TLI = 0.908$; $SRMR = 0.074$; $RMSEA = 0.063$).

4.3 Ergebnisse der multiplen linearen Regression

Nachdem die Validität der reflektiven Messungen nun gegeben war, konnte die multiple lineare Regression berechnet werden, um das aufgestellte Forschungsmodell (Abbildung 1) quantitativ zu prüfen. Auch hierfür wurde die Statistiksoftware Mplus 8.0 (Muthén und Muthén, 1998-2010) genutzt. Um möglichen Problemen bezüglich der Homoskedastizität und Verteilung der unabhängigen Variablen entgegenzuwirken, wurde der MLR Schätzer verwendet (Li, 2016). Die Ergebnisse der Regression können Abbildung 6 entnommen werden. Die Hypothese H_1 , für welche wir einen negativen Einfluss von generativer KI als AV auf die ZB erwarteten, wurde statistisch signifikant bestätigt. Ebenso bestätigte sich die Hypothese H_{3a} statistisch sehr

⁸Konkret betraf dies die Items Interest6 und Interest7 für das Themeninteresse sowie GenAIBenevolence2, GenAIBenevolence3 und GenAIIntegrity3 für Vertrauen in generative KI

signifikant ($p < 0.01$), die ein geringeres Vertrauen bei generativer KI als Erzeuger von Texten postulierte. Die Hypothese H_{3b} , für die wir einen positiven Effekt zwischen Vertrauen und ZB erwarteten, bestätigte sich knapp nicht statistisch signifikant ($p = 0.073, \gamma = 0.902$). Ebenso abgelehnt wurden die Hypothesen H_{2a} und H_{2b} , welche mit $p = 0.447$ und $p = 0.754$ weit von der Bestätigung entfernt waren. Die Kontrollvariable VZ hatte einen statistisch signifikanten Einfluss auf die ZB ($p = 0.033, \gamma = 6.565$). Alle anderen Kontrollvariablen hatten keinen statistisch signifikanten Einfluss auf die ZB (Alter: $p = 0.257, \gamma = -0.102$; Themeninteresse: $p = 0.994, \gamma = 0.011$; TV ist KI: $p = 0.033, \gamma = 0.474$; Geschlecht: $p = 0.080, \gamma = 3.069$; Ist Akademiker: $p = 0.724, \gamma = 0.616$). Das Einkommen wurde klassiert in 1000€ Schritten abgefragt, wobei die letzte Klasse jegliches Monatseinkommen größer 6000€ enthält. Da die oberste Klasse theoretisch unendlich nach oben skaliert, kann hier nicht die gleiche Annahme wie für die Likert-Skala Items nach Harpe (2015) getroffen werden. Um einen möglichen Einkommenseffekt auf die ZB zu prüfen, wurde stattdessen ein Kruskal-Wallis Test durchgeführt ($\chi^2 = 10.695, p = 0.0302$). Die Residuenvarianz der Zielvariable ZB betrug $\sigma^2 = 48.735$, während die gesamte Stichprobe eine Varianz von $\sigma^2 = 74.636$ für die ZB aufwies. Daraus ergab sich $R^2 = 0.347$, was bedeutet, dass das Modell 34.7 Prozent der Gesamtvarianz der ZB erklärte und somit den von Ozili (2023) aufgestellten Grenzwert von $R^2 = 0.1$ überschritt. Da mindestens ein Regressor statistisch signifikant und $R^2 > 0.1$ war, wurde das Regressionsmodell akzeptiert (Ozili, 2023).

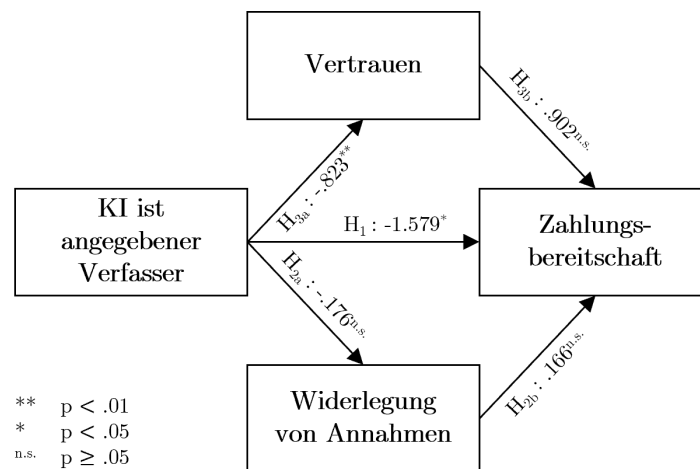


Abbildung 6: Geschätztes Modell

5 Diskussion

In dieser Arbeit sollte untersucht werden, wie die ZB für Texte durch eine generative KI als AV beeinflusst wird. Hier versuchten wir, Teile dieses Einflusses durch das Vertrauen in den Verfasser und über die WvA zu erklären. Dazu wurde ein online Experiment mit 37 Teilnehmenden durchgeführt, um die notwendigen Daten zur Validierung der Hypothesen zu erheben. Die Daten wurden anschließend mittels einer multiplen linearen Regression ausgewertet. Durch die Ergebnisse der Regression hat sich bestätigt, dass Menschen eine geringere ZB für Texte haben, bei denen eine generative KI als Erzeuger angegeben ist. Zwar vertrauten die Teilnehmenden einer generativer KI als Erzeuger weniger als einem menschlichen Verfasser, jedoch konnte kein signifikanter Einfluss von Vertrauen in den Verfasser auf die ZB festgestellt werden. Wenn eine KI als Erzeuger angegeben wurde, hatte dies keinen signifikanten Effekt auf die WvA, die wiederum auch die ZB nicht signifikant beeinflusst.

5.1 Eine KI als angegebener Verfasser senkt die Zahlungsbereitschaft für den Text

Die multiple lineare Regression hat bestätigt, dass eine KI als AV die ZB für den jeweiligen Text senkt (H_1). Die von uns vermuteten Faktoren für die geringere ZB zielen eher auf Aspekte der menschlichen Schreibweise ab. Da der tatsächliche Verfasser aber keinen signifikanten Einfluss auf die ZB hatte, verdeutlicht dies das Problem der Erkennbarkeit von KI-generierten Inhalten: Die Teilnehmenden konnten nicht zuverlässig unterscheiden, ob ein Text von einer KI oder einem Menschen verfasst wurde, aber ihre ZB sank dennoch bei der bloßen Annahme, dass der Text KI-generiert sei. Somit liegt die Vermutung nahe, dass die Teilnehmenden ihre geringere ZB für als KI-generiert gekennzeichnete Texte nicht aufgrund der fehlenden persönlichen Note und Originalität der Texte äußerten. Da die Teilnehmenden allerdings über die Kenntnis des AV der Texte verfügten und sich über die unterschiedlichen Produktionsgeschwindigkeiten von menschlichen und KI-generierten Werken bewusst waren, beeinflusste dies potenziell ihre Wertschätzung der Texte, die wiederum die ZB senken könnte. Diese allgemeine Präferenz für menschliche Werke würde somit auch die Ergebnisse von Zhang und Gosline (2023) bestätigen.

Die Ergebnisse dieser Arbeit liefern einige interessante Implikationen, die sich sowohl auf die Praxis, als auch auf die weiterführende Forschung im Bereich der KI und ihren Einsatzgebieten beziehen. Bereits heute stellen die Ergebnisse dieser Arbeit Unternehmen vor ein Dilemma: Wenn Konsumenten nicht über den TV informiert werden oder, wie in dieser Studie dargestellt, absichtlich falsche Angaben gemacht werden, zeigen sie keine statistisch signifikant niedrigere ZB für KI-generierte Inhalte. Aus Sicht der Unternehmen könnte demnach der erhöhte

Produktivitätsgewinn, den generative KI verspricht, ohne Einbußen der ZB genutzt werden, solange dem Konsumierenden angegeben wird, dass der vorliegende Text von einem Menschen verfasst wurde oder keine exakte Angabe getätigt wird⁹. Sollten Unternehmen diese neue Technologie nicht nutzen, droht es ihnen von Unternehmen abgehängt zu werden, die von den Vorteilen generativer KI Gebrauch machen. Die juristischen Implikationen dessen sind derzeit noch nicht klar geregelt, wie bereits ausführlich in Kapitel 2.1.1 diskutiert. Ein klarer gesetzlicher Handlungsrahmen seitens des Gesetzgebers, beispielsweise durch Ausweitung der diskutierten KI-Verordnung seitens der EU auf Transparenzpflicht für Texte, (Füllsack, 2023) werden notwendig sein, um Transparenz in den Einsatz von generativer KI zu bringen. Curry und Stroud (2021) zeigen einen positiven Zusammenhang zwischen erhöhter Transparenz und erhöhte Glaubwürdigkeit.

Da der TV keinen signifikanten Einfluss auf die ZB hatte, kann noch nicht von einem Qualitätsunterschied zwischen menschenverfassten und KI-generierten Texten gesprochen werden. Es ist jedoch nicht unrealistisch zu erwarten, dass sich ein möglicher Qualitätsunterschied zwischen menschenverfassten und KI-generierten Inhalten durch technologischen Fortschritt ergeben wird. Sollte in Zukunft ein signifikanter Unterschied zwischen der Qualität von menschlichen und von generativer KI erzeugten Texten existieren, wäre es interessant, die ZB erneut zu untersuchen. Somit kann der Zielkonflikt zwischen der Wertschätzung, der investierten Zeit und Arbeit des Menschen und der höheren Qualität der KI-generierten Texte in Bezug auf die ZB untersucht werden. Schließlich scheinen Menschen nach unseren Ergebnissen die investierte Zeit der menschlichen Verfasser bei gleichbleibender Qualität zu schätzen. Es bleibt zu klären, ob diese Wertschätzung weiterhin besteht, wenn die Qualität der von generativer KI erstellten Texte, die laut O'Brien et al. (2020) einen signifikanten Einfluss auf die ZB hat, höher ist als die Qualität menschlicher Texte.

5.2 Das Vertrauen in generative KI als angegebener Verfasser ist geringer als bei menschlichen Verfassern

Ebenfalls wurde bestätigt, dass KI als AV das Vertrauen in den Verfasser negativ beeinflusst (H_{3a}). Wurde KI als Erzeuger angegeben, vertrauten die Teilnehmenden dieser im Schnitt um 0.832 Punkte auf der Likert-Skala weniger, als einem Menschen als AV. Das geringer ausfallende Vertrauen bestätigt im Vergleich zu Rix et al. (2022) eine Aversion gegenüber generativer KI als Erzeuger von journalistischen und belletristischen Texten. Dies könnte an der höheren medialen Aufmerksamkeit bezüglich (generativer) KI der letzten Jahre liegen. Durch die technologischen Fortschritte in diesem Gebiet, die den Meisten erst durch die Veröffentlichung von ChatGPT Ende 2022 bewusst geworden sind, kann es sein, dass Menschen die Befürchtung haben, dass KI die menschliche Kreativität und Authentizität in

⁹Denkbar wäre beispielsweise der Angegebene Verfasser „Die Redaktion“ oder ähnliche

künstlerischen und schriftlichen Werken ersetzen könnte. Dadurch könnte das wahrgenommene Wohlwollen der KI gegenüber dem Menschen geschädigt sein, was das geringere Vertrauen erklären würde¹⁰.

Eine Einordnung der Ergebnisse in das Vertrauensmodell nach Lucassen und Schraagen (2012) kann über den Unterschied des Einflusses von TV ist KI auf die ZB sowie AV ist KI auf die ZB vorgenommen werden. Da die ZB für Texte, bei denen der TV eine generative KI ist, nicht signifikant niedriger war, als bei menschenverfassten Texten, scheint es keine qualitativen Unterschiede bei den Inhalten der Texte zu geben. Unsere Ergebnisse bestätigen somit die Ergebnisse von Graefe und Bohlken (2020), die einen ähnlichen Effekt auf die Glaubwürdigkeit beobachteten. Somit scheint es Vertrauensdefizite in die Quelle anstatt ihrer Erzeugnisse zu geben. Wie Vertrauen in (generative) KI verbessert werden kann, ist Gegenstand vieler Forschung, beispielsweise durch erklärbare KI (Bayer et al., 2022). Es ist denkbar, dass die in Kapitel 2.1.1 beschriebenen Transparenzvorschriften die von uns herausgefundenen Vertrauensproblematiken in Bezug auf generative KI als Quelle verbessern könnte, indem eine klare Abgrenzung zwischen den Erzeugnissen generativer KI und menschlichen Verfassern erzwungen wird. Wir schließen uns deswegen vorherigen Aufrufen für erhöhte Transparenz und Ethikvorschriften im journalistischen Bereich an (Graefe et al., 2018; Diakopoulos, 2015).

5.3 Das Vertrauen in den Verfasser beeinflusst die Zahlungsbereitschaft nicht signifikant

Wir konnten nicht nachweisen, dass das Vertrauen in den Verfasser die ZB positiv beeinflusst (H_{3b}). Allerdings könnte das Vertrauen auch einen anderen, nicht berücksichtigten Faktor beeinflussen, der sich wiederum auf die ZB auswirkt. Wir konnten somit auch die Überlegungen von O'Brien et al. (2020) in Anlehnung an Newman et al. (2018) über einen positiven Einfluss des Vertrauens auf die ZB nicht bestätigen. Die Insignifikanz widerspricht dabei etlicher Literatur aus diversen Bereichen (Rahimi et al., 2022; Roosen et al., 2015; Nocella et al., 2010; Aksoy und Özsönmez, 2019; Meng et al., 2020). Aufgrund der vielen Belege aus der Literatur sowie des lediglich knappen Überschreitens des Grenzwertes ($p = 0.073$), stellen wir als mögliche Begründung die ungenaue Operationalisierung der Variable Vertrauen auf. Die geplante Operationalisierung mithilfe des Vertrauensmodells nach Mayer et al. (1995) stellte sich als ungeeignet heraus, wodurch die Items nach Söllner et al. (2012) genutzt werden mussten. Die reflektiven Items nach Söllner et al. (2012) sind in der Forschung ungeprüfter als das formative Modell nach Mayer et al. (1995) und bilden somit eventuell das Vertrauen nicht adäquat ab.

¹⁰Die Analyse verwendete reflektive Items gemäß Söllner et al. (2012); dennoch ist das wahrgenommene Wohlwollen bei der Vertrauensevaluation einer anderen Partei wichtig (Mayer et al., 1995).

5.4 Widerlegung von Annahmen als mediierendes Konstrukt

Dass die WvA bei einer KI als AV signifikant geringer ist als bei einem Menschen als AV konnte nicht bestätigt werden (H_{2a}). Dennoch blieb eine positiv signifikante WvA im Vergleich zum Mensch als AV, die von Rix et al. (2022) als mögliche Begründung für eine höhere ZA von Inhalten generativer KI aufgestellt wurde, aus. Darüber hinaus konnte ebenso nicht bestätigt werden, dass sich die WvA positiv auf die ZB auswirkt (H_{2b}). Unser Versuch, das Konstrukt WvA aus der Expectation Confirmation Theorie auf die ZB für Texte generativer KI anzuwenden, erwies sich als inadäquat. Als möglicher Grund könnte die von uns vorgenommene Vereinfachung des Modells genannt werden. Laut der Expectation Confirmation Theorie beeinflusst die WvA die Zufriedenheit (Oliver, 1980). Zwischen Zufriedenheit und ZB existiert nach Homburg et al. (2005) ein statistisch signifikanter Einfluss. Somit hätte Zufriedenheit als zusätzliches Konstrukt in unserem Forschungsmodell den eigentlichen Nutzen der WvA aus der Expectation Confirmation Theorie exakter abgebildet. Die Untersuchung des Effekts der WvA auf die Zufriedenheit, die sich wiederum auf die ZB auswirken könnte, wäre somit eine angebrachte Anpassung des Modells.

5.5 Kontrollvariablen

Zusätzlich zu den Hypothesen des Forschungsmodells wurden einige Kontrollvariablen aufgenommen. Diesen wurde in der Literatur ein möglicher signifikanter Einfluss auf die ZB zugeschrieben (O'Brien et al., 2020). Wie erwartet hatten VZ einen signifikanten Einfluss auf die Zahlungsbereitschaft eines Textes. Wer bereits in der Vergangenheit eine Produktrezension erworben hat, wird eher bereit sein, dies erneut zu tun, als jemand, der noch nie für eine Produktrezension gezahlt hat. Dies unterstreicht die Wichtigkeit von Kundenbindung sowie der damit einhergehenden Loyalität (Griffin, 1995). Interessanterweise hatte das Themeninteresse keinen signifikanten Einfluss auf die ZB. Eine mögliche Erklärung dafür könnten kostenfreie Alternativen sein, wie sie insbesondere für den journalistischen Kontext oft existieren. Teilnehmende mit hohem thematischen Interesse waren sich vermutlich eher über diese kostenlosen Alternativen bewusst als Teilnehmende, die ein geringes Interesse an den Produkten aufwiesen. Die Signifikanz des Einkommens wurde nicht, wie in Kapitel 4.2 erläutert, durch das Regressionsmodell berechnet, sondern mithilfe des Kruskal-Wallis Test evaluiert. Der in der Literatur vielfach beschriebene Einkommenseffekt (Horowitz und McConnell, 2003; Jacobsen und Hanley, 2009) hat sich durch den Kruskal-Wallis Test auch für dieses Experiment signifikant bestätigt. Alle anderen Kontrollvariablen waren demografischer Natur und stellten sich als insignifikant heraus. Wir konnten in der Literatur beschriebene demografische Effekte auf die ZB (O'Brien et al., 2020; Goyanes, 2014) demnach nicht bestätigen.

6 Beitrag zur Forschung

Wir tragen mit dieser Arbeit zu der Literatur über ZB für generativen KI bei, indem wir Theorie über Zahlungsbereitschaft in herkömmliche digitalen Medien auf generative KI anwenden und diese quantitativ testen. Wir betrachten zusätzlich die Problematik der Erkennbarkeit von Erzeugnissen von generativer KI und diskutieren Auswirkungen auf die Praxis. Dabei unterstreichen wir die Relevanz von Regulierungen für generativer KI, insbesondere im Hinblick auf Transparenzrichtlinien. Zudem folgen wir dem Forschungsauftrag von O'Brien et al. (2020), indem wir den Einfluss des Vertrauens auf die ZB untersuchen. Wir erweitern die Literatur zu Vertrauen in KI-Systeme, indem wir die Validität des viel zitierten Vertrauensmodells nach Mayer et al. (1995) auf generative KI als Erzeuger prüfen. Zuletzt erweitern wir die Literatur über Expectation-Confirmation Theorie im Technologiebereich (Oliver, 1980) und untersuchen den von Rix et al. (2022) vermuteten Überraschungseffekt über die Erzeugnisse generativer KI.

7 Limitationen und offene Forschungsimpulse

Das Experimentdesign beinhaltet mehrere Problematiken, die unter Umständen die Aussagekraft der Ergebnisse beeinflussen könnte. Ein erster Diskussionspunkt ist die Erhebung der ZB. In dieser Arbeit wurde die ZB direkt abgefragt. Die direkte Abfrage der ZB bringt dabei einige verzerrende Effekte mit sich. Breidert et al. (2015) nennen folgende Probleme mit der direkten Abfrage:

1. Ein unnatürlicher Fokus auf den Preis. Indem direkt nach dem Preis gefragt wird, werden andere, ggf. wichtigere Attribute des Produktes überschattet. Da die ZB eines der zentralen Konstrukte in dieser Arbeit ist, halten wir diesen Punkt für unbedeutend. In anderen Experimenten, bei denen beispielsweise das Vertrauen im Zentrum der Arbeit steht, wäre der dadurch entstehende größere Fokus auf den Preis problematischer.
2. Teilnehmende haben nicht unbedingt einen Anreiz, ihre tatsächliche ZB offenzulegen. Breidert et al. (2015) beschreiben als Begründung vor allem sog. Prestige- und Zusammenhaltseffekte. Der Prestigeeffekt beschreibt, dass man nicht als geizig gegenüber den Forschenden auftreten möchte, während der Zusammenhaltseffekt dafür sorgt, dass Teilnehmende gemeinsam versuchen könnten, den Preis niedrig zu halten, damit sie das Produkt billiger erwerben können. Da das Experiment anonym durchgeführt wurde, vernachlässigen wir den Prestigeeffekt. Der Zusammenhaltseffekt stellt jedoch eine Limitation der Arbeit dar, da dieser außerhalb unseres Einflussbereichs liegt.
3. Selbst wenn Teilnehmende ihre tatsächliche ZB offenlegen, impliziert das nicht, dass Teilnehmende das Produkt tatsächlich kaufen werden. Dieses Problem konnte durch

das Experimentdesign nicht gelöst werden. Innerhalb einer Umfrage kann nicht geprüft werden, ob Teilnehmende beispielsweise ein belletristisches Werk tatsächlich zu dem genannten Preis erwerben werden.

4. Die direkte Abfrage der ZB ist eine kognitiv belastende Aufgabe. Ob der Effekt dazu führt, dass Produkte über- oder unterbewertet werden, sei noch unklar. Ein gewisser Bias sei laut Breidert et al. (2015) jedoch wahrscheinlich. Wir versuchen diesem Effekt durch die Randomisierung der Textreihenfolge und somit der Abfrage der ZB entgegenzuwirken. Zusätzlich wurden Kontrollfragen eingebaut, um die Aufmerksamkeit der Teilnehmenden zu validieren. Wir erhofften uns, dass Teilnehmende, die diese Tests bestehen, ebenso bedacht ihre ZB angeben. Ganz herausrechnen ließ sich der Effekt nicht, weshalb dies eine Limitation der Arbeit darstellt und die Ergebnisse beeinflusst.
5. Teilnehmende sind sich oft des Marktwertes eines Produktes nicht bewusst, was die ZB bei Erkenntnis über den tatsächlichen Marktwert unter Umständen sogar stark beeinflusst. Indem ein Durchschnittspreis angegeben wird, kann für diesen beeinflussenden Effekt kontrolliert werden.

Breidert et al. (2015) sowie Balderjahn (2003) schlagen diverse andere Methoden zur Erfassung der ZB vor, die aber im Rahmen einer (Online) Befragung nur schwierig bis gar nicht umsetzbar sind. Online Experimente können generell die Generalisierbarkeit von Experimenten negativ beeinflussen (Chyi und Lee, 2013). Durch die digitale Durchführung sind alle präsentierten Texte ebenso vollkommen digital. Grundsätzlich ist die ZB für digitale Texte geringer (Franklin, 2014). Aus zeitlichen Gründen war es nicht möglich, alle Teilnehmenden persönlich zu befragen. Zudem beeinflussen die oben beschriebenen Prestigeeffekte bei direkter persönlicher Befragung die Ergebnisse vermutlich stärker als bei einer anonymisierten Befragung. In nachfolgenden Studien wäre daher ein Experiment interessant, in welchem die ZB der Teilnehmenden durch die von Breidert et al. (2015) und Balderjahn (2003) vorgeschlagenen indirekten Methoden ermittelt wird.

Eine weitere Limitation der Arbeit ist der Anteil der menschlichen Teilnahme an der Erzeugung des Textes. Die Forschungsfrage zielt auf den Unterschied der ZB für menschliche und KI-generierte Texte ab. In dem Experiment wurden jedoch die KI-generierten Texte nicht unüberwacht generiert. Stattdessen wurden Texte teilweise sowohl mit gleicher Aufforderung als auch mit angepasster Aufforderung erneut generiert, um möglichst qualitativ hochwertige inhaltlich passende Texte zu erhalten. Dementsprechend liegt in dieser Arbeit kein HOOTL-System vor.

Zudem stellt die Stichprobengröße eine weitere Limitation der Arbeit dar. Zwar erfüllt diese, wie in Kapitel 3.1 beschrieben, gerade so die Mindeststichprobengröße, ist aber noch zu klein, um ein SEM anzuwenden. Außerdem ist denkbar, dass sich manche Hypothesen (z. B. H_{3b}) durch eine größere Stichprobe als signifikant herausstellen könnten. Ursprünglich

war eine deutlich höhere Stichprobengröße angepeilt, diese wurde jedoch aufgrund der hohen Abbruchrate und der Rekrutierungsstrategie der Teilnehmenden nicht erreicht.

Eine weitere Einschränkung der Arbeit lag in dem Kontext des Konsums vor. Rix et al. (2022) konnten zeigen, dass ein utilitaristischer Kontext die ZA eines Konsumenten positiv beeinflusst. Zwar sei es möglich, den Kontext künstlich über das jeweilige Thema herzustellen (Produktrezensionen als utilitaristischer, Belletristik als hedonischer Konsumkontext), jedoch fehlt in unserer Arbeit die Validierung, ob die jeweiligen Texte wirklich in dem angezielten Kontext konsumiert werden. Da wir in unserem Experiment keine Rücksicht auf die Unterscheidung zwischen den Kontexten des Konsums der Texte genommen haben, wäre dies ein weiterer Untersuchungsgegenstand, der in zukünftigen Forschungsarbeiten betrachtet werden könnte.

Des Weiteren gelang es uns nicht, das Vertrauen nach Mayer et al. (1995) formativ zu bilden. Durch den Umstieg auf die reflektive Messung ist die Aussagekraft des Konstruktes somit möglicherweise ungenauer, was zu einer eingeschränkten Aussagekraft der Ergebnisse bezüglich des Vertrauens führen könnte. Zukünftige Studien in diesem Kontext könnten Vertrauen durch ein anderes formatives Modell messen. Wie von O'Brien et al. (2020) gefordert, können auch die Auswirkung anderer psychologischer Faktoren, wie beispielsweise der Glaubwürdigkeit, auf die ZB untersucht werden. Außerdem haben wir in dieser Arbeit keine Effekte von initialem Vertrauen in generative KI auf andere Konstrukte untersucht. Dieser Aspekt könnte in zukünftigen Studien betrachtet werden.

Zur weiteren Vertiefung der Thematik wäre es interessant zu untersuchen, wie sich die KI als Erzeuger von Texten auf andere Konstrukte neben der ZB auswirkt. Hierbei ist vor allem die Akzeptanz der algorithmisch erzeugten Texte von Interesse. Erste Untersuchungen zu dieser Thematik im Bereich des Journalismus führten Kim und Kim (2021) durch. Die Ausweitung ihrer Untersuchungen auf andere Bereiche wäre dabei eine spannende Möglichkeit für zukünftige Forschungsarbeiten.

Abschließend wurden in dieser Arbeit lediglich Texte untersucht. Generative KI kann, wie bereits diskutiert, deutlich mehr als nur Texte generieren. Wie sich die ZB für andere Erzeugnisse generativer KI verhält, stellt eine weitere interessante Forschungslücke dar.

8 Fazit

Das Ziel dieser Arbeit war es, die ZB für KI-generierte Texte im Vergleich zu menschenverfassten Texten zu untersuchen. Da keine signifikanten Qualitätsunterschiede in den Texten von generativer KI und menschlichen Autoren vermutet wurden, sollte evaluiert werden, inwiefern Konsumierende die geleistete Arbeit von menschlichen Verfassern noch wertschätzen oder ob sich die gleichbleibende Qualität durch eine gleichbleibende ZB ausdrückt. Die Ergebnisse der Arbeit zeigen auf, dass die Wertschätzung der Arbeit menschlicher Verfasser weiterhin exis-

tiert. Darüber hinaus konnten wir signifikante Vertrauensdefizite in generative KI feststellen. Gleichzeitig konnte das Erkennbarkeitsproblem bestätigt werden, welches Unternehmen vor komplexe Entscheidungen bezüglich der Vorteile generativer KI und Transparenz gegenüber Konsumierenden stellt. Unsere Ergebnisse unterstreichen die Signifikanz von integrierter, transparenter Nutzung generativer KI zum Wohle von Konsumierenden als auch Arbeitnehmenden. Ob diese integrale Nutzung durch Abmachungen zwischen Arbeitnehmenden und Arbeitgebern, wie im Falle der WGA, erreicht werden kann, wird sich herausstellen. Gesetzliche Regelungen, wie die geplante KI-Verordnung der EU, können eine andere Lösung sein. Dass langfristig eine nachhaltige Lösung über den Einsatz generativer KI gefunden werden muss, verdeutlichen unsere Ergebnisse.

Literaturverzeichnis

- Ahmed, S., Nielsen, I. E., Tripathi, A., Siddiqui, S., Ramachandran, R. P., & Rasool, G. (2023). Transformers in Time-Series Analysis: A Tutorial. <https://doi.org/10.1007/s00034-023-02454-8>
- Aksoy, H., & Özsönmez, C. (2019). How Millennials' Knowledge, Trust, and Product Involvement Affect the Willingness to Pay a Premium Price for Fairtrade Products? *Asian Journal of Business Research*. <https://doi.org/10.14707/ajbr.190062>
- Balderjahn, I. (2003). Erfassung der Preisbereitschaft. In H. Diller & A. Herrmann (Hrsg.), *Handbuch Preispolitik: Strategien - Planung - Organisation - Umsetzung* (S. 387–404). Gabler Verlag. https://doi.org/10.1007/978-3-322-90512-3_18
- Bayer, S., Gimpel, H., & Markgraf, M. (2022). The role of domain expertise in trusting and following explainable AI decision support systems. *Journal of Decision Systems*, 32(1), 110–138. <https://doi.org/10.1080/12460125.2021.1958505>
- Bentele, G. (1998). Vertrauen / Glaubwürdigkeit. In *Politische Kommunikation in der demokratischen Gesellschaft: Ein Handbuch mit Lexikonteil* (S. 305–311). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-80348-1_19
- Bitkom e.V. (2023). Vom Knochen zum Must-have: 40 Jahre Handy. *Bitkom e.V.* Verfügbar 4. Januar 2024 unter <https://www.bitkom.org/Presse/Presseinformation/Vom-Knochen-zum-Must-Have-40-Jahre-Handy>
- Bitkom e.V. (n.d.). Deutsche Wirtschaft drückt bei Künstlicher Intelligenz aufs Tempo. *Bitkom e.V., 2023*. Verfügbar 24. Januar 2024 unter <https://www.bitkom.org/Presse/Presseinformation/Deutsche-Wirtschaft-drueckt-bei-Kuenstlicher-Intelligenz-aufs-Tempo>
- Bolter, J. D. (1993). *Turing's man: Western culture in the computer age* (4. Aufl.). Univ. of North Carolina Pr. http://www.jstor.org/stable/10.5149/9781469616308_bolter

- Börsenverein des Deutschen Buchhandels. (2021). *Tabellenkompendium*. https://www.boersenverein.de/tx_file_download?tx_theme_pi1%5BfileUid%5D=8430&tx_theme_pi1%5BpageUid%5D=128&tx_theme_pi1%5Breferer%5D=https%3A%2F%2Fwww.boersenverein.de%2Fmarkt-daten%2Fmarktforschung%2Fwirtschaftszahlen%2F&cHash=29d0bf8fc4a6e1a694fc0d040ba08736
- Breidert, C., Hahsler, M., & Reutterer, T. (2015). A Review of Methods for Measuring Willingness-to-Pay. *Innovative Marketing*, 1(4). https://www.researchgate.net/publication/242382759_A_Review_of_Methods_for_Measuring_Willingness-to-Pay
- Briganti, G. (2023). How ChatGPT works: a mini review. *European Archives of Oto-Rhino-Laryngology*, 1–5. <https://doi.org/10.1007/s00405-023-08337-7>
- Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., Boselie, P., Lee Cooke, F., Decker, S., DeNisi, A., Dey, P. K., Guest, D., Knoblich, A. J., Malik, A., Paauwe, J., Papagiannidis, S., Patel, C., Pereira, V., Ren, S., ... Varma, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, 33(3), 606–659. <https://doi.org/10.1111/1748-8583.12524>
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2., aktualisierte Aufl.). Pearson Studium. <https://elibrary.pearson.de/book/99.150005/9783863268138>
- Cabrera, Á. A., Perer, A., & Hong, J. I. (2023). Improving Human-AI Collaboration With Descriptions of AI Behavior. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 1–21. <https://doi.org/10.1145/3579612>
- Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. <http://arxiv.org/pdf/2303.04226.pdf>
- Catak, M., AlRasheedi, S., AlAli, N., AlQallaf, G., AlMeri, M., & Ali, B. (2021). Artificial Intelligence Composer. *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 608–613. <https://doi.org/10.1109/3ICT53449.2021.9581896>
- Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1), 1–8. <https://doi.org/10.1016/j.jebo.2011.08.009>
- Chyi, H. I., & Lee, A. M. (2013). ONLINE NEWS CONSUMPTION. *Digital Journalism*, 1(2), 194–211. <https://doi.org/10.1080/21670811.2012.753299>
- Chyi, H. I., & Yang, M. J. (2009). Is Online News an Inferior Good? Examining the Economic Nature of Online News among Users. *Journalism & Mass Communication Quarterly*, 86(3), 594–612. <https://doi.org/10.1177/107769900908600309>

- Coetzee, C. (2023). *Generating a full-length work of fiction with GPT-4*. <https://medium.com/@chiaracoetzee/generating-a-full-length-work-of-fiction-with-gpt-4-4052cfeddef3>
- Cowles, M., & Davis, C. (1982). On the origins of the .05 level of statistical significance. *American Psychologist*, 37(5), 553–558. <https://doi.org/10.1037/0003-066X.37.5.553>
- Coyle, J. (2023a). Could AI pen 'Casablanca'? Screenwriters take aim at ChatGPT. *AP News*. Verfügbar 16. Januar 2024 unter <https://apnews.com/article/ai-hollywood-writers-strike-artificial-intelligence-dc71c4cabcca0ee1b1afe0050d392a36>
- Coyle, J. (2023b). In Hollywood writers' battle against AI, humans win (for now). *AP News*. Verfügbar 16. Januar 2024 unter <https://apnews.com/article/hollywood-ai-strike-wga-artificial-intelligence-39ab72582c3a15f77510c9c30a45ffc8>
- Credé, M., & Harms, P. D. (2015). 25 years of higher-order confirmatory factor analysis in the organizational sciences: A critical review and development of reporting recommendations. *Journal of Organizational Behavior*, 36(6), 845–872. <https://doi.org/10.1002/job.2008>
- Crosby, P. (2022). Don't Judge a Book by Its Cover: Examining Digital Disruption in the Book Industry Using a Stated Preference Approach. In S. Cameron (Hrsg.), *The Economics of Books and Reading* (S. 91–121). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-18199-3_5
- Curry, A. L., & Stroud, N. J. (2021). The effects of journalistic transparency on credibility assessments and engagement intentions. *Journalism*, 22(4), 901–918. <https://doi.org/10.1177/1464884919850387>
- Deng, L., & Yuan, K.-H. (2015). Multiple-Group Analysis for Structural Equation Modeling With Dependent Samples. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(4), 552–567. <https://doi.org/10.1080/10705511.2014.950534>
- Diakopoulos, N. (2015). Algorithmic Accountability. *Digital Journalism*, 3(3), 398–415. <https://doi.org/10.1080/21670811.2014.976411>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of experimental psychology. General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Erb, H. N. (1990). A Statistical Approach for Calculating the Minimum Number of Animals Needed in Research. *ILAR Journal*, 32(1), 11–16. <https://doi.org/10.1093/ilar.32.1.11>
- EY. (2023). Arbeitsplatzsorgen wegen Künstlicher Intelligenz bei Top-Verdienern am größten. *EY*. Verfügbar 21. Januar 2024 unter https://www.ey.com/de_de/news/2023/09/ey-jobstudie-digitalisierung-2023
- Fast, E., & Horvitz, E. (2017). Long-Term Trends in the Public Perception of Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.10635>

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fisher, R. (1925). *Statistical methods for research workers* ((11th ed. rev)). Oliver and Boyd. https://link.springer.com/chapter/10.1007/978-1-4612-4380-9_6
- Franklin, B. (2014). The Future of Journalism. *Journalism Studies*, 15(5), 481–499. <https://doi.org/10.1080/1461670X.2014.930254>
- Füllsack, A. L. (2023). Künstliche Intelligenz und der Journalismus der Zukunft. *CMS Hasche Sigle*. Verfügbar 20. Januar 2024 unter <https://www.cmshs-bloggt.de/rechtsthemen/kuenstliche-intelligenz/kunstliche-intelligenz-ki-und-der-journalismus-der-zukunft/>
- Gangadharbatla, H. (2022). The Role of AI Attribution Knowledge in the Evaluation of Artwork. *Empirical Studies of the Arts*, 40(2), 125–142. <https://doi.org/10.1177/0276237421994697>
- Gehrmann, S., Strobelt, H., & Rush, A. (2019). GLTR: Statistical Detection and Visualization of Generated Text. In M. R. Costa-jussà & E. Alfonseca (Hrsg.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (S. 111–116). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-3019>
- Ghosh, A., & Fossas, G. (2022). Can There be Art Without an Artist? <http://arxiv.org/pdf/2209.07667.pdf>
- Gmyrek, P., Berg, J., & Bescond, D. (2023). *Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality*. <https://doi.org/10.2139/ssrn.4584219>
- Google Trends. (21.01.2024). *Google Trends*. Verfügbar 21. Januar 2024 unter <https://trends.google.de/trends/explore?date=today%205-y&geo=DE&q=%2Fm%2F0mkz&hl=de>
- Goyanes, M. (2014). An Empirical Study of Factors that Influence the Willingness to Pay for Online News. *Journalism Practice*, 8(6), 742–757. <https://doi.org/10.1080/17512786.2014.882056>
- Graefe, A. (2016). Guide to Automated Journalism. https://www.researchgate.net/publication/289529155_Guide_to_Automated_Journalism/link/568edc4508aead3f42f07586/download
- Graefe, A., & Bohlken, N. (2020). Automated Journalism: A Meta-Analysis of Readers' Perceptions of Human-Written in Comparison to Automated News. *Media and Communication*, 8(3), 50–59. <https://doi.org/10.17645/mac.v8i3.3019>
- 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>
- Griffin, J. (1995). *Customer loyalty*. <http://altfeldinc.com/pdfs/customer%20loyalty.pdf>

- Habib, S. S., & Zaidi, S. (2021). Exploring willingness to pay for health insurance and preferences for a benefits package from the perspective of women from low-income households of Karachi, Pakistan. *BMC Health Services Research*, 21(1), 380. <https://doi.org/10.1186/s12913-021-06403-6>
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Hanson, M., Abramson, M., & McNamara, J. F. (1986). Practical Significance in Special Education Research. *The Journal of Special Education*, 20(4), 401–408. <https://doi.org/10.1177/002246698602000403>
- Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, 7(6), 836–850. <https://doi.org/10.1016/j.cptl.2015.08.001>
- Heck, R., & Thomas, S. L. (2015). *An Introduction to Multilevel Modeling Techniques*. Routledge. <https://doi.org/10.4324/9781315746494>
- Henestrosa, A. L., Greving, H., & Kimmerle, J. (2023). Automated journalism: The effects of AI authorship and evaluative information on the perception of a science journalism article. *Computers in Human Behavior*, 138, 107445. <https://doi.org/10.1016/j.chb.2022.107445>
- Heston, T., & Khun, C. (2023). Prompt Engineering in Medical Education [PII: ime2030019]. *International Medical Education*, 2(3), 198–205. <https://doi.org/10.3390/ime2030019>
- Hitsuwari, J., Ueda, Y., Yun, W., & Nomura, M. (2023). Does human–AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. *Computers in Human Behavior*, 139, 107502. <https://doi.org/10.1016/j.chb.2022.107502>
- Hohlbein, W. (2010). *Der Hammer der Götter: Die Asgard-Saga*. http://www.hohlbein.de/asgard_saga.html
- Homburg, C., Koschate, N., & Hoyer, W. D. (2005). Do Satisfied Customers Really Pay More? A Study of the Relationship between Customer Satisfaction and Willingness to Pay. *Journal of Marketing*, 69(2), 84–96. <https://doi.org/10.1509/jmkg.69.2.84.60760>
- Horowitz, J. K., & McConnell, K. E. (2003). Willingness to accept, willingness to pay and the income effect. *Journal of Economic Behavior & Organization*, 51(4), 537–545. [https://doi.org/10.1016/S0167-2681\(02\)00216-0](https://doi.org/10.1016/S0167-2681(02)00216-0)
- Hossain, M. A., & Quaddus, M. (2012). Expectation–Confirmation Theory in Information System Research: A Review and Analysis. In Y. K. Dwivedi, M. R. Wade & S. L. Schneberger (Hrsg.), *Information Systems Theory: Explaining and Predicting Our Digital Society, Vol. 1* (S. 441–469). Springer Science+Business Media, LLC. https://doi.org/10.1007/978-1-4419-6108-2_21

- Hu, L.-t., & Bentler, P. M. (1999). *Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives* (Bd. 6). <https://doi.org/10.1080/10705519909540118>
- Huang, H., Fu, R., & Ghose, A. (2023). *Generative AI and Content-Creator Economy: Evidence from Online Content Creation Platforms*. <https://doi.org/10.2139/ssrn.4670714>
- Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1), 73–101. <https://doi.org/10.1214/aoms/1177703732>
- Hui, X., Reshef, O., & Zhou, L. (2023). *The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market*. <https://doi.org/10.2139/ssrn.4527336>
- Jacobsen, J. B., & Hanley, N. (2009). Are There Income Effects on Global Willingness to Pay for Biodiversity Conservation? *Environmental and Resource Economics*, 43(2), 137–160. <https://doi.org/10.1007/s10640-008-9226-8>
- Kahnemann, D., & Ritov, I. (1994). Determinants of Stated Willingness to Pay for Public Goods: A Study in the Headline Method. *Journal of Risk and Uncertainty*, 9(1), 5–37. <http://www.jstor.org/stable/41760735>
- Keith, T. Z. (2006). *Multiple regression and beyond*. Pearson/Allyn and Bacon.
- Kim, D., & Kim, S. (2021). A model for user acceptance of robot journalism: Influence of positive disconfirmation and uncertainty avoidance. *Technological Forecasting and Social Change*, 163, 120448. <https://doi.org/10.1016/j.techfore.2020.120448>
- Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, 114, 106553. <https://doi.org/10.1016/j.chb.2020.106553>
- Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will You Accept an Imperfect AI? In S. Brewster, G. Fitzpatrick, A. Cox & V. Kostakos (Hrsg.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (S. 1–14). ACM. <https://doi.org/10.1145/3290605.3300641>
- Kremp, M. (2023). Pixel 8 und Pixel 8 Pro im Test: Das können Googles neue KI-Smartphones. *DER SPIEGEL*. Verfügbar 19. Januar 2024 unter <https://www.spiegel.de/netzwelt/gadgets/pixel-8-und-pixel-8-pro-im-test-was-googles-neue-ki-smartphones-koennen-a-d5591f35-d631-4b66-9fbe-ab3668a341f1>
- Kreps, S., McCain, R. M., & Brundage, M. (2022). All the News That’s Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *Journal of Experimental Political Science*, 9(1), 104–117. <https://doi.org/10.1017/XPS.2020.37>
- Kucharavy, A., Schillaci, Z., Maréchal, L., Würsch, M., Dolamic, L., Sabonnadiere, R., David, D. P., Mermoud, A., & Lenders, V. (2023). Fundamentals of Generative Large Language Models and Perspectives in Cyber-Defense. <http://arxiv.org/pdf/2303.12132.pdf>

- Kulkarni, P., Mahabaleshwarkar, A., Kulkarni, M., Sirsikar, N., & Gadgil, K. (2019). Conversational AI: An Overview of Methodologies, Applications & Future Scope. *2019 Fifth International Conference on Computing, Communication Control and Automation (ICCUBEA)*, 1–7. <https://doi.org/10.1109/ICCUBEA47591.2019.9129347>
- Legner, S. (2020). *Are Works of Artificial Intelligence in Need for Further Protection?* <https://doi.org/10.2139/ssrn.3735360>
- Leitão, L., Amaro, S., Henriques, C., & Fonseca, P. (2018). Do consumers judge a book by its cover? A study of the factors that influence the purchasing of books. *Journal of Retailing and Consumer Services*, 42, 88–97. <https://doi.org/10.1016/j.jretconser.2018.01.015>
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems*, 17(1), 39–71. <https://doi.org/10.1016/j.jsis.2008.01.001>
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), 100790. <https://doi.org/10.1016/j.ijme.2023.100790>
- Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022). News from Generative Artificial Intelligence Is Believed Less. In T. Speith (Hrsg.), *A review of taxonomies of explainable artificial intelligence (XAI) methods* (S. 97–106). ACM; Saarländische Universitäts- und Landesbibliothek. <https://doi.org/10.1145/3531146.3533077>
- Lucassen, T., & Schraagen, J. M. (2012). Propensity to trust and the influence of source and medium cues in credibility evaluation. *Journal of Information Science*, 38(6), 566–577. <https://doi.org/10.1177/0165551512459921>
- Mansholt, M. (2022). iPhone 14 im Test: Ein tolles iPhone – das Sie sich lieber sparen sollten. *STERN.de*. Verfügbar 19. Januar 2024 unter <https://www.stern.de/digital/tests/iphone-14-im-test--ein-tolles-iphone---das-sie-sich-lieber-sparen-sollten-32729470.html>
- Marbeau, Y. (1987). What value pricing research today. *Journal of the Marketing Research Society*, (Vol. 29), 153–182. <http://web.lib.aalto.fi/en/oa/db/SCIMA/?cmd=listget&id=53586&q=%40indexterm%20RETAIL%20PRICING&p=63&cnt=107>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3), 709. <https://doi.org/10.2307/258792>
- McCarthy, J. (2004). What is Artificial Intelligence?

- McKnight, D. H. (2005). Trust in Information Technology. *The Blackwell Encyclopedia of Management. Vol. 7 Management Information Systems*, 329–331. https://www.researchgate.net/publication/297731446_Trust_in_Information_Technology
- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research*, 13(3), 334–359. <https://doi.org/10.1287/isre.13.3.334.81>
- McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial Trust Formation in New Organizational Relationships. *The Academy of Management Review*, 23(3), 473. <https://doi.org/10.2307/259290>
- Meng, F., Ji, Z., Song, F., Bai, T., Fan, X., & Wang, D. (2020). Patients’ familiarity with, trust in and willingness to pay for traditional Chinese medicine in Chinese community health care centres. *European Journal of Integrative Medicine*, 36, 101118. <https://doi.org/10.1016/j.eujim.2020.101118>
- Metz, C. (2024). OpenAI Says New York Times Lawsuit Against It Is ‘Without Merit’. *The New York Times*. Verfügbar 23. Januar 2024 unter <https://www.nytimes.com/2024/01/08/technology/openai-new-york-times-lawsuit.html>
- Midjourney. (2024). *Are there group plan discounts or enterprise billing options?* Verfügbar 24. Januar 2024 unter <https://help.midjourney.com/en/articles/8150091-are-there-group-plan-discounts-or-enterprise-billing-options>
- Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (2023). DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature. <http://arxiv.org/pdf/2301.11305.pdf>
- MMC. (2019). *The MMC State of AI 2019 Report*. <https://mmc.vc/resources/fund-brochures/The-MMC-State-of-AI-2019-Report.pdf>
- Muthén, L. K., & Muthén, B. O. (1998-2010). Mplus user’s guide. (6th ed.)
- Nass, C., & Moon, Y. (2000). *Machines and Mindlessness: Social Responses to Computers* (Bd. 56). <https://doi.org/10.1111/0022-4537.00153>
- Nature. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, 613(7945), 612. <https://doi.org/10.1038/d41586-023-00191-1>
- Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2018). Digital News Report 2018. <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/digital-news-report-2018.pdf>
- Nickel, P. J., Franssen, M., & Kroes, P. (2010). Can We Make Sense of the Notion of Trustworthy Technology? *Knowledge, Technology & Policy*, 23(3-4), 429–444. <https://doi.org/10.1007/s12130-010-9124-6>
- Nocella, G., Hubbard, L., & Scarpa, R. (2010). Farm Animal Welfare, Consumer Willingness to Pay, and Trust: Results of a Cross-National Survey. *Applied Economic Perspectives and Policy*, 32(2), 275–297. <https://doi.org/10.1093/aepp/ppp009>

- Nusair, K., & Hua, N. (2010). Comparative assessment of structural equation modeling and multiple regression research methodologies: E-commerce context. *Tourism Management*, 31(3), 314–324. <https://doi.org/10.1016/j.tourman.2009.03.010>
- O’Brien, D., Wellbrock, C.-M., & Kleer, N. (2020). Content for Free? Drivers of Past Payment, Paying Intent and Willingness to Pay for Digital Journalism – A Systematic Literature Review. *Digital Journalism*, 8(5), 643–672. <https://doi.org/10.1080/21670811.2020.1770112>
- Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research*, 17(4), 460–469. <https://doi.org/10.1177/002224378001700405>
- Ooi, K.-B., Tan, G. W.-H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., Lee, V.-H., Loh, X.-M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P., Sarker, P., Sharma, A., ... Wong, L.-W. (2023). The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions. *Journal of Computer Information Systems*, 1–32. <https://doi.org/10.1080/08874417.2023.2261010>
- OpenAI. (2022). *Introducing ChatGPT*. Verfügbar 21. Januar 2024 unter <https://openai.com/blog/chatgpt>
- OpenAI. (2023). *Introducing ChatGPT Enterprise*. Verfügbar 24. Januar 2024 unter <https://openai.com/blog/introducing-chatgpt-enterprise>
- Osbourne, J. W., & Waters, E. (2002). Four Assumptions of Multiple Regression That Researchers Should Always Test. *Practical Assessment*, 8(2). https://www.researchgate.net/publication/234616195_Four_Assumptions_of_Multiple_Regression_That_Researchers_Should_Always_Test
- Ostrom, A. L., Fotheringham, D., & Bitner, M. J. (2019). Customer Acceptance of AI in Service Encounters: Understanding Antecedents and Consequences. *Handbook of Service Science, Volume II*, 77–103. https://doi.org/10.1007/978-3-319-98512-1_5
- Ozili, P. K. (2023). *The acceptable R-square in empirical modelling for social science research*. <https://mpra.ub.uni-muenchen.de/116496/>
- Pitt, J. C. (2010). It’s Not About Technology. *Knowledge, Technology & Policy*, 23(3-4), 445–454. <https://doi.org/10.1007/s12130-010-9125-5>
- Rahimi, A., Azimi, G., & Jin, X. (2022). What Drives Commuters to Pay for Autonomous Vehicles? *Transportation Research Record: Journal of the Transportation Research Board*, 2676(4), 267–280. <https://doi.org/10.1177/03611981211058095>
- Raj, R., Singh, A., Kumar, V., & Verma, P. (2023). Analyzing the potential benefits and use cases of ChatGPT as a tool for improving the efficiency and effectiveness of business operations. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 3(3), 100140. <https://doi.org/10.1016/j.tbench.2023.100140>

- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. <http://arxiv.org/pdf/2102.12092.pdf>
- Rana, S. (2023). AI and GPT for Management Scholars and Practitioners: Guidelines and Implications. *FIIB Business Review*, 12(1), 7–9. <https://doi.org/10.1177/23197145231161408>
- Rattray, J. E., Lauder, W., Ludwick, R., Johnstone, C., Zeller, R., Winchell, J., Myers, E., & Smith, A. (2011). Indicators of acute deterioration in adult patients nursed in acute wards: a factorial survey. *Journal of clinical nursing*, 20(5-6), 723–732. <https://doi.org/10.1111/j.1365-2702.2010.03567.x>
- Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Pla.* https://www.researchgate.net/publication/37705092_The_Media_Equation_How_People_Treat_Computers_Television_and_New_Media_Like_Real_People_and_Pla
- Rhode, M. (2016). *Die Blutschule: Thriller* (Vollständige Taschenbuchausgabe, Bd. Band 17502). Bastei Lübbe AG. <https://sebastianfitzek.de/buch/die-blutschule/>
- Rix, J., Rußell, R., Rühr, A., & Hess, T. (2022). Human vs. AI: Investigating Consumers' Context-Dependent Purchase Intentions for Algorithm-Created Content. In T. Bui (Hrsg.), *Proceedings of the 55th Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences. <https://doi.org/10.24251/HICSS.2022.554>
- Roosen, J., Bieberstein, A., Blanchemanche, S., Goddard, E., Marette, S., & Vandermoere, F. (2015). Trust and willingness to pay for nanotechnology food. *Food Policy*, 52, 75–83. <https://doi.org/10.1016/j.foodpol.2014.12.004>
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not So Different After All: A Cross-Discipline View Of Trust. *The Academy of Management Review*, 23(3), 393–404. <https://doi.org/10.5465/AMR.1998.926617>
- Russell, S. J., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Prentice Hall.
- Rutterford, C., Copas, A., & Eldridge, S. (2015). Methods for sample size determination in cluster randomized trials. *International Journal of Epidemiology*, 44(3), 1051–1067. <https://doi.org/10.1093/ije/dyv113>
- Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics*, 26(5), 2749–2767. <https://doi.org/10.1007/s11948-020-00228-y>
- Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-Generated Text be Reliably Detected? <https://arxiv.org/pdf/2303.11156.pdf>
- Samoili, S., Cobo, M. L., Gomez, E., de Prato, G., Martinez-Plumed, F., & Delipetrev, B. (2020). *AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*. Joint Research Centre (Seville site). <https://eprints.ugd.edu.mk/28047/>

- Schiefele, U. (1990). Thematisches Interesse, Variablen des Leseprozesses und Textverstehen. *Zeitschrift für experimentelle und angewandte Psychologie*, (37), 304–332. <https://dnb.info/121838204X/34>
- Schirner, G., Erdogmus, D., Chowdhury, K., & Padir, T. (2013). The Future of Human-in-the-Loop Cyber-Physical Systems. *Computer*, 46(1), 36–45. <https://doi.org/10.1109/MC.2013.31>
- Schniederjans, D. G., Curado, C., & Khalajhedayati, M. (2020). Supply chain digitisation trends: An integration of knowledge management. *International Journal of Production Economics*, 220, 107439. <https://doi.org/10.1016/j.ijpe.2019.07.012>
- Schwenk, H. (2007). Continuous space language models. *Computer Speech & Language*, 21(3), 492–518. <https://doi.org/10.1016/j.csl.2006.09.003>
- Söllner, M., Hoffmann, A., Hirdes, E. M., Rudakova, L., Leimeister, S., & Leimeister, J. (2010). Towards a Formative Measurement Model for Trust. *BLED 2010 Proceedings*. <https://aisel.aisnet.org/bled2010/37>
- Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., & Leimeister, J. M. (2012). Understanding the Formation of Trust in IT Artifacts. http://dx.doi.org/10.1007/978-3-319-05044-7__3
- Statistisches Bundesamt. (2024). *Verbraucherpreisindex und Inflationsrate*. Verfügbar 21. Januar 2024 unter https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Verbraucherpreisindex/_inhalt.html
- Stein, C. M., Morris, N. J., & Nock, N. L. (2012). Structural equation modeling [Journal Article]. *Statistical Human Genetics*, 850, 495–512. https://doi.org/10.1007/978-1-61779-555-8_27
- Stejskal, J., Hajek, P., & Prokop, V. (2021). The role of library user preferences in the willingness to read and pay for e-books: case of the Czech Republic. *The Electronic Library*, 39(4), 639–660. <https://doi.org/10.1108/EL-01-2021-0001>
- Sundar, S. S. (2008). *The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility* / Request PDF. Verfügbar 13. Januar 2024 unter https://www.researchgate.net/publication/323990996_The_MAIN_Model_A_Heuristic_Approach_to_Understanding_Technology_Effects_on_Credibility
- Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, 59(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- UK Public General Acts. (1988). Copyright, Designs and Patents Act. <https://www.legislation.gov.uk/ukpga/1988/48/contents>
- Upwork Inc. (n. D.). *Investor Relations* / Upwork Inc. Verfügbar 21. Januar 2024 unter <https://investors.upwork.com/>
- Urheberrechtsgesetz [UrhG]. (2021). Verfügbar 21. Januar 2024 unter <https://www.gesetze-im-internet.de/urhg/>

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. <http://arxiv.org/pdf/1706.03762.pdf>
- Vedapradha, R., Hariharan, R., & Shivakami, R. (2019). Artificial Intelligence: A Technological Prototype in Recruitment. *Journal of Service Science and Management*, 12(03), 382–390. <https://doi.org/10.4236/jssm.2019.123026>
- ver.di. (2023). *Generische KI im Journalismus – Sorgfalt, Transparenz und Qualität gewährleisten*. Verfügbar 21. Januar 2024 unter <https://dju.verdi.de/ueber-uns/nachrichten/++co++7b088aa6-5623-11ee-880a-001a4a160111>
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and psychological measurement*, 76(6), 913–934. <https://doi.org/10.1177/0013164413495237>
- Wölker, A., & Powell, T. E. (2018). Algorithms in the newsroom? News readers’ perceived credibility and selection of automated journalism. *Journalism*, 2021.
- Writers Guild of America. (2023). *Summary of the 2023 WGA MBA*. Verfügbar 21. Januar 2024 unter <https://www.wga.org/contracts/contracts/mba/summary-of-the-2023-wga-mba>
- Writers Guild of America East. (2024). *What is the Guild? / Writers Guild of America, East*. Verfügbar 21. Januar 2024 unter <https://www.wgaeast.org/what-is-the-guild/>
- Yuan, K.-H., & Bentler, P. M. (1998). 9. Structural Equation Modeling with Robust Covariances. *Sociological Methodology*, 28(1), 363–396. <https://doi.org/10.1111/0081-1750.00052>
- Zhang, Y., & Gosline, R. (2023). Human favoritism, not AI aversion: People’s perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. *Judgment and Decision Making*, 18. <https://doi.org/10.1017/jdm.2023.37>

Kapitelangaben und Erklärungen

Kapitelangaben

Folgende Kapitel wurden von Sidney Strasser verfasst:

- 2.1
- 2.2
- 2.3
- 2.4
- 3.1
- 4.1.1
- 4.1.3
- 4.1.4

Folgende Kapitel wurden von Thomas Schaffrath verfasst:

- 2.5
- 3
- 3.2
- 4.1.2
- 4.3

Alle anderen Kapitel wurden gemeinsam verfasst.



Anlage „Erklärung zur Verwendung generativer KI-Systeme“

Bei der Erstellung der Arbeit habe ich die folgenden auf künstlicher Intelligenz (KI) basierten Systeme¹ benutzt:

1. ChatGPT
2. Consensus

Ich erkläre weiterhin, dass ich

- ☒ mich aktiv über die Leistungsfähigkeit und Beschränkungen der oben genannten KI-Systeme informiert habe,
- ☒ die aus den oben angegebenen KI-Systemen übernommenen Passagen gekennzeichnet habe,
- ☒ überprüft habe, dass die mithilfe der oben genannten KI-Systeme generierten und von mir übernommenen Inhalte faktisch richtig sind,
- ☒ mir bewusst bin, dass ich als Autor:in dieser Arbeit die Verantwortung für die in ihr gemachten Angaben und Aussagen trage.

Die oben genannten KI-Systeme habe ich wie im Folgenden dargestellt eingesetzt.

Arbeitsschritt	Eingesetzte(s) KI-System(e)	Beschreibung der Verwendungsweise
Generierung von Ideen und Konzeption der Arbeit	/	
Literatursuche	Consensus	Wurde teilweise für die Literaturrecherche verwendet.
Literaturanalyse	/	
Literaturverwaltung und Zitationsmanagement	/	
Auswahl von Methoden und Modellen	/	
Datensammlung und -analyse	/	
Generierung von Programmcodes	/	

¹ Wenn Sie unsicher sind, ob Sie eine verwendetes KI-System angeben müssen, wenden Sie sich an Ihre:n Betreuer:in.

Erstellung von Visualisierungen	/	
Interpretation und Validierung	/	
Strukturierung des Texts der Arbeit	/	
Formulierung des Texts der Arbeit	/	
Übersetzung des Texts der Arbeit	/	
Redigieren des Texts	ChatGPT	Generierung von alternativen Formulierungen für bestimmte Sätze. Dabei wurden keine Sätze exakt übernommen. Sie dienten lediglich der Inspiration für eigene, sinnvolle Formulierungen.
Vorbereitung der Präsentation des Texts	/	
Sonstiges	ChatGPT	ChatGPT wurde für die Generierung von Texten in unserem Experiment genutzt. Die Verwendungsweise wird detailliert in der Arbeit beschrieben.

Stuttgart, 24.01.2023,

Unterschrift Strasser



Stuttgart, 24.01.2023,

Unterschrift Schaffrath



Eigenständigkeitserklärung

Hiermit versichere ich, Thomas Schaffrath,

- dass die Arbeit, bzw. bei einer Gruppenarbeit mein entsprechend gekennzeichnete Teil, selbstständig verfasst wurde,
- dass keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet wurden,
- dass keine anderen als die angegebenen Hilfsmittel verwendet wurden,
- dass die eingereichte Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens war,
- dass die Arbeit weder vollständig noch in Teilen bereits veröffentlicht wurde und
- dass das elektronische Exemplar mit den anderen Exemplaren übereinstimmt.

Stuttgart, den 23. Januar 2024

Unterschrift



Hiermit versichere ich, Sidney Strasser,

- dass die Arbeit, bzw. bei einer Gruppenarbeit mein entsprechend gekennzeichnete Teil, selbstständig verfasst wurde,
- dass keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet wurden,
- dass keine anderen als die angegebenen Hilfsmittel verwendet wurden,
- dass die eingereichte Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens war,
- dass die Arbeit weder vollständig noch in Teilen bereits veröffentlicht wurde und
- dass das elektronische Exemplar mit den anderen Exemplaren übereinstimmt.

Stuttgart, den 23. Januar 2024

Unterschrift



Anhang

Produktrezension iPhone 14 - Malte Mansholt:

Das iPhone 14 ist schneller als der Vorgänger und knipst bessere Fotos. Trotzdem kann man den Kauf leider kaum empfehlen. Warum, erklärt der Test

Es ist ein kleines Paradoxon: Technisch ist das neue iPhone 14 das beste Basis-iPhone, dass Apple je auf den Markt gebracht hat. Doch einige Entscheidungen des Konzerns lassen es erheblich weniger attraktiv wirken – vor allem die Konkurrenz aus dem eigenen Haus. [...] Dabei ist das iPhone 14 im Vergleich zum Basismodell des iPhone 13 ein Fortschritt. Beim Design hat sich zunächst wenig getan. Das iPhone 14 setzt wieder auf die durchaus schicke Kombination aus Aluminium-Rahmen und Glasrückseite, fühlt sich auch weitgehend genauso an. Der einzige Unterschied: Die Kamera-Ausbuchtung ist minimal größer geworden. Der Unterschied ist kaum sichtbar, sorgt aber dafür, dass die Hüllen des Vorgängers haarscharf nicht mehr passen. Wer ein iPhone 13 hat, kann die Hüllen des 14 nutzen – andersherum klappt es nicht. Auch das Display ist dasselbe. Der OLED-Bildschirm löst weiter in 2532 x 1170 Bildpunkten auf, ist mit maximal 1200 Nits gleich hell. Auch die Ausbuchtung für die Gesichtserkennung FaceID und die Frontkamera, die sogenannte Notch, ist weiter in vollem Umfang vorhanden. Die innovative „Dynamic Island“, bei der sich die Aussparung je nach Kontext verändert, hat Apple nur den beiden iPhone 14 Pro spendiert. Das iPhone 14 ist dadurch beim Look und dem Nutzungsgefühl zunächst kaum vom Vorgänger zu unterscheiden. Auf der technischen Seite gibt es da mehr Neuerungen. Mit dem A15 Bionic setzt Apple zwar wieder auf denselben Prozessor wie beim iPhone 13, es gibt aber einen Grafikkern mehr. In der Rechenleistung ist das aber kaum messbar. Im Benchmark-Programm Geekbench schneidet das iPhone 14 sogar etwas schwächer aus als das iPhone 13 Pro mit demselben Chip. Das ist etwas irritierend. Fairerweise muss man aber sagen, dass [...]

Produktrezension iPhone 14 - ChatRev:

Das iPhone 14, Apples jüngster Zugang in der Welt der Smartphones, steht im Rampenlicht und wird, wie alle Produkte des Technologiegiganten, sowohl hochgelobt als auch kritisch betrachtet. In unserem Test nehmen wir das iPhone 14 unter die Lupe und beleuchten seine vielfältigen Aspekte, von der Performance bis hin zur Nutzererfahrung.

Beim ersten in-die-Hand-nehmen des iPhone 14 fällt die hochwertige Verarbeitung auf. Apple hat es erneut geschafft, ein Gerät zu schaffen, das sich nicht nur hochwertig anfühlt, sondern auch ästhetisch ansprechend ist. Das Zusammenspiel von Glas und Metall sorgt für eine angenehme Haptik und gibt dem Nutzer das Gefühl von Stabilität und Qualität. Dies spiegelt sich auch in den technischen Daten wider: Mit einem A15 Bionic Chip, der bereits im Vorgängermodell beeindruckte, liefert das iPhone 14 eine solide Performance. Das Display

des iPhone 14 ist mit seiner Super Retina XDR-Technologie ein wahrer Hingucker. Mit einer Auflösung von 2532 x 1170 Pixeln bei 460 ppi bietet es eine kristallklare Darstellung, und die Farben wirken lebendig und naturgetreu. Die Helligkeit ist beeindruckend, auch unter direktem Sonnenlicht, was das iPhone 14 zu einem idealen Begleiter für den Außenbereich macht. Jedoch ist das Fehlen einer höheren Bildwiederholrate bemerkenswert, insbesondere wenn man bedenkt, dass dies bei anderen High-End-Smartphones mittlerweile Standard ist. Interessant ist die Abwesenheit des Dynamic Island, eines innovativen Features, das in anderen Modellen der Serie zu finden ist. Dieses Merkmal, das die Bildschirmnutzung optimiert und eine neue Art der Interaktion ermöglicht, hätte dem iPhone 14 eine zusätzliche Dimension verleihen können. In einem Markt, der sich immer weiterentwickelt und in dem sich die Grenzen zwischen Mittelklasse und Premium-Smartphones zunehmend verwischen, könnte das Fehlen solcher Features [...]

Prompt

Schreibe einen Produkttest zum iPhone 14, in dem du dieses kritisierst aber ab und an einige positive Aspekte hervorhebst. Der Text soll sich auf alle Aspekte des iPhones konzentrieren, was du zu Beginn des Textes erwähnen sollst. Du schreibst aber nur über die Elemente Haptik, Display und das fehlen des Dynamic Island, was du nicht in der Einleitung oder irgendwo anders im Text erwähnen sollst. Gehe dabei auch auf die technischen Daten des iPhones ein. Der Text soll an einer spannenden Stelle abrupt aufhören. Außerdem soll es ein Fließtext ohne Unterabsätze sein, der ungefähr 250 Wörter umfasst. Schreibe den Text so, als wäre er in einem Technologie-Magazin veröffentlicht worden.

Produktrezension Pixel 8 Pro - Matthias Kremp

Das können Googles neue KI-Smartphones

Für seine neuen Pixel-Smartphones hat Google einen Chip entwickelt, neue Bildschirme und Kameras eingebaut. Die wichtigen Änderungen liegen aber in der Software, wie unser Test zeigt. Wären da nicht unterschiedliche Farben und Oberflächen, könnte man die neuen Pixel 8 und Pixel 8 Pro kaum von Googles Smartphones vom Vorjahr unterscheiden. Es gibt ein paar neue Farben, unter anderem ein wunderschönes Hellblau. Zudem ist die Rückseite der Pro-Modelle jetzt matt. Das sieht gut aus, weil sich Fingerabdrücke darauf nicht so stark abzeichnen wie bei den Vorgängern und beim Pixel 8, das weiterhin hochglänzend geliefert wird. Ein weiterer Unterschied zu den Vorjahresmodellen: Die Bildschirme sind an den Seiten nicht mehr so stark in den Rahmen hineingebogen. Stattdessen sind nun die Ecken kräftiger gerundet, das fühlt sich gut an. Auch technisch haben die OLED-Displays der Pixel-8-Serie zugelegt. Das Display des Pixel 8 ist schneller geworden, kann jetzt wie

das Pro-Modell bis zu 120 Bilder pro Sekunde liefern, was beim Scrollen und Spielen für einen ruhigeren Bildeindruck sorgt. Das Pixel 8 Pro hat einen Langsamgang spendiert bekommen und kann etwa beim Lesen von rasanten 120 Bildern pro Sekunde auf ein Bild pro Sekunde herunterschalten. Das spart Strom. Ohnehin zeigten beide Geräte im Test gute Ausdauer, vor allem das größere Pixel 8 Pro. Über einen normalen Arbeitstag komme ich damit meist problemlos. Daran dürfte auch der neue Tensor G3-Chip seinen Anteil haben, der leistungsfähiger und zugleich effizienter geworden sein soll als der G2. Ein konkreter Vergleich der Leistung des Tensor G3 mit der des Tensor G2 aus den Pixel-7-Modellen ist aber derzeit nicht möglich. Google hat die Installation von Testprogrammen wie etwa Geekbench und Antutu auf den Testgeräten für die Presse unterbunden. Die Kameras sind nah dran an denen der Vorgänger, aber doch verbessert worden. So sind beide Modelle mit [...]

Produktrezension Pixel 8 Pro - ChatRev

Pixel 8 vs. Pixel 8 Pro: Überraschende Wahrheiten Enthüllt - Erfahren Sie, Was Google Nicht Preisgibt!

In unserem heutigen Test werfen wir einen umfassenden Blick auf Googles neueste Smartphone-Erweiterungen, das Pixel 8 und das Pixel 8 Pro. Diese beiden Modelle sind nicht nur eine Fortsetzung der Pixel-Serie, sondern sie repräsentieren auch eine evolutionäre Entwicklung im Vergleich zu ihren Vorgängern. Beide Modelle, das Pixel 8 und das Pixel 8 Pro, bringen eine Vielzahl von Verbesserungen und neuen Features. Dabei fällt sofort die hochwertige Verarbeitung auf, die sich in einer angenehmen Haptik und Robustheit äußert. Google setzt bei beiden Geräten auf ein klares und minimalistisches Design, das sowohl elegant als auch funktional ist. Im Vergleich zu den Vorgängermodellen ist allerdings kein signifikanter Fortschritt in der Materialqualität oder im Design zu erkennen, was einige Nutzer möglicherweise enttäuschen könnte. Die Displays beider Modelle haben deutliche Fortschritte gemacht. Während das Pixel 8 mit einem Full-HD+-OLED-Display ausgestattet ist, bietet das Pixel 8 Pro ein QHD+-OLED-Display. Beide Bildschirme überzeugen mit ihrer scharfen Darstellung und den lebendigen Farben, wobei das Pro-Modell mit einer höheren Auflösung und Bildqualität punktet, was insbesondere bei der Wiedergabe von hochauflösenden Inhalten einen deutlichen Unterschied macht. Unter der Haube zeigen beide Geräte ihre Stärken. Ausgestattet mit Googles neuestem Tensor-Prozessor, bieten sie eine verbesserte Leistung gegenüber ihren Vorgängern. Diese Steigerung zeigt sich nicht nur in einer schnelleren Verarbeitungsgeschwindigkeit, sondern auch in einer effizienteren Energieverwaltung, was die Akkulaufzeit positiv beeinflusst. Insbesondere beim Pixel 8 Pro ist die Performance beeindruckend und macht es zu einer hervorragenden Wahl für Power-User, die hohe Ansprüche an ihre Geräte stellen. Während das Pixel 8 und das Pixel 8 Pro in vielen Aspekten überzeugen können, gibt es dennoch Bereiche, in denen...

Prompt

Schreibe einen Produkttest zum Google Pixel 8 und 8 Pro, in dem du diese kritisierst aber ab und an einige positive Aspekte hervorhebst. Der Text soll sich auf alle Aspekte der Pixel konzentrieren, was du zu Beginn des Textes erwähnen sollst. Du schreibst aber nur über die Elemente Haptik, Display und den Prozessor, was du auf keinen Fall schreiben sollst. Es soll der Eindruck entstehen, als würde sich der Test auf alle Facetten des Smartphones konzentrieren. Gehe dabei auch auf die technischen Daten der Pixel ein und vergleiche die beiden Modelle miteinander und zu den Vorgängermodellen. Der Text soll an einer spannenden Stelle abrupt aufhören. Außerdem soll es ein Fließtext ohne Unterabsätze sein, der ungefähr 250 Wörter umfasst. Schreibe den Text so, als wäre er in einem Technologie-Magazin veröffentlicht worden. Gebe dem Test noch eine Spannende, „Clickbait“-artige Überschrift.

Belletristik Kriminalroman - Max Rhode

Eine unbewohnte Insel im Storkower See. Eine Holzhütte, eingerichtet wie ein Klassenzimmer. Eine Schule mit den Fächern: Fallen stellen. Opfer jagen. Menschen töten. Die Teenager Simon und Mark können sich keinen größeren Horror vorstellen, als aus der Metropole Berlin in die Einöde Brandenburgs zu ziehen. Das Einzige, worauf sie sich freuen, sind sechs Wochen Sommerferien, doch auch hier macht ihnen ihr Vater einen Strich durch die Rechnung. Er nimmt sie mit auf einen Ausflug zu einer ganz besonderen Schule. Gelegen mitten im Wald auf einer einsamen Insel. Mit einem grausamen Lehrplan, nach dem sonst nur in der Hölle unterrichtet wird ...

Das Vergessene Erbe, Max Rhode

Belletristik Kriminalroman - ChatRev

In einem Berliner Krankenhaus erwacht Kommissar Jakob Stern ohne Erinnerung an die letzten 48 Stunden. Neben ihm: Ein rätselhafter Zettel mit einer Warnung und der Name einer Frau, die kurz darauf tot aufgefunden wird. Stern taucht in einen Fall ein, der ihn in die Abgründe des menschlichen Geistes führt. Mit jeder aufgedeckten Lüge verstrickt er sich tiefer in ein Netz aus Korruption und Wahnsinn. Stern muss die Wahrheit finden, bevor der Täter erneut zuschlägt – und bevor seine eigene Vergangenheit ihn einholt. Ein packender Thriller, der die Grenzen der Realität auf die Probe stellt.

Das Vergessene Erbe, ChatLit

Prompt

Schreibe einen Klappentext für einen Kriminalroman im Stil von Sebastian Fitzek. Der Text soll ca. 100 Wörter umfassen und den Namen von Fitzek sowie generische Aussagen wie „Ein atemberaubendes Rätsel, das bis zur letzten Seite fesselt“ nicht enthalten.

Belletristik Fantasyroman - Wolfgang Hohlbein

Eine Flotte auf dem Weg ins Nichts. Tausend zu allem entschlossene Krieger. Ein Sturm, der das Meer aufpeitscht. Ein tödliches Geheimnis. Und ein Mann, der angetreten ist, es zu lösen. In den Augen seiner Männer ist er Thor, der Gott. In seinen eigenen ist er ein verzweifelt Suchender. Der Bezwinger, der einst bezwungen wurde und nun alles zu verlieren droht: seine Schiffe, seine Krieger, seine Freunde, sein Leben. Nur was, wenn es nichts zu gewinnen gibt?

Der Hammer der Götter, Wolfgang Hohlbein

Belletristik Fantasyroman - ChatLit

In der verborgenen Stadt Eldoria, umhüllt von ewigem Nebel, entdeckt die junge Magierin Lysandra ein uraltes Artefakt, das die Grenzen zwischen den Welten zu zerreißen droht. Gejagt von einer dunklen Bruderschaft, muss sie die Geheimnisse des Artefakts lüften, bevor es in die falschen Hände gerät. Auf ihrer gefährlichen Reise durch verzauberte Wälder und vergessene Ruinen findet sie unerwartete Verbündete: einen entehrten Ritter und einen geheimnisvollen Wanderer mit einer verborgenen Agenda. Zusammen stellen sie sich einer Bedrohung, die nicht nur Eldoria, sondern die gesamte Welt zu verschlingen droht. Ein Kampf um Macht, Verrat und das Schicksal beginnt.

Die Schatten von Eldoria, ChatLit

Prompt

Schreibe einen Klappentext für einen Fantasyroman im Stil von Wolfgang Hohlbein. Der Text soll ca. 100 Wörter umfassen und den Namen von Hohlbein sowie generische Aussagen wie „Ein atemberaubendes Rätsel, das bis zur letzten Seite fesselt“ nicht enthalten.

Tabelle 4: Fragenkatalog

Konstrukt	Frage	Quelle
Zahlungsbereitschaft (Produktrezension)	Gehen Sie davon aus, dass <Angegebener Verfasser> noch mehr Tests zu aktuellen Smartphones geschrieben hat und diese gebündelt verfügbar sind. Ab welchem Preis würden Sie die Artikel von <Angegebener Verfasser> auf keinen Fall kaufen, weil Sie es sich nicht leisten können oder weil Sie der Meinung sind, dass es das Geld nicht wert ist? Eine solche Sammlung von Produkttests kostet normalerweise ca. 4,90 €	Marbeau (1987)
Zahlungsbereitschaft (Belletristik)	Gehen Sie davon aus, dass der vorherige Text der Klappentext eines Taschenbuches ist. Ab welchem Preis würden Sie das Buch von <Angegebener Verfasser> auf keinen Fall kaufen, weil Sie es sich nicht leisten können oder weil Sie der Meinung sind, dass es das Geld nicht wert ist? Ein Taschenbuch aus diesem Genre kostet im Schnitt ca. 14,88 €.	Marbeau (1987)
Wohlwollen der Verfasser (CFA)	1. Ich glaube, dass <Angegebener Verfasser> in meinem Interesse handelt.	Li et al. (2008)

Konstrukt	Frage	Quelle
	2. Wenn ich Hilfe benötige, gibt <Angegebener Verfasser> sein Bestes, um mir zu helfen	Li et al. (2008)
	3. <Angegebener Verfasser> kümmert sich um mein Wohlergehen, nicht nur um das eigene	Li et al. (2008)
Fähigkeit der Verfasser (CFA)	1. <Angegebener Verfasser> ist kompetent und effektiv im Schreiben von Texten	Li et al. (2008)
	2. <Angegebener Verfasser> erfüllt seine Aufgabe, Texte zu schreiben, sehr gut	Li et al. (2008)
	3. Insgesamt ist <Angegebener Verfasser> fähig und kompetent beim Schreiben von Texten	Li et al. (2008)
	4. Insgesamt hat <Angegebener Verfasser> ein großes Wissen über den Inhalt der Texte, die er schreibt	Li et al. (2008)
Integrität der Verfasser (CFA)	1. <Angegebener Verfasser> ist in seinen Interaktionen mit mir wahrheitsgemäß	Li et al. (2008)
	2. <Angegebener Verfasser> ist ehrlich	Li et al. (2008)
	3. <Angegebener Verfasser> hält seine Versprechen	Li et al. (2008)
	4. <Angegebener Verfasser> ist aufrichtig und authentisch	Li et al. (2008)
Vertrauen der Verfasser (reflektiv, CFA)	1. Ich vertraue <Angegebener Verfasser>	Söllner et al. (2012)
	2. Ich habe ein gutes Gefühl, wenn ich mich auf <Angegebener Verfasser> verlasse	Söllner et al. (2012)

Konstrukt	Frage	Quelle
	3. Ich kann den Informationen von <Angenehmer Verfasser> vertrauen	Söllner et al. (2012)
Widerlegung von Annahmen (Produktrezensionen)	1. Ich erwarte, dass <Angenehmer Verfasser> gute Produktrezensionen schreibt 2. <Angenehmer Verfasser> schreibt gute Produktreviews	Kocielnik et al. (2019) Kocielnik et al. (2019))
Widerlegung von Annahmen (Belletristik)	1. Ich erwarte, dass <Angenehmer Verfasser> gute Unterhaltungsliteratur schreibt 2. <Angenehmer Verfasser> schreibt gute Unterhaltungsliteratur	Kocielnik et al. (2019) Kocielnik et al. (2019)
Thematisches Interesse	1. <Kontext> ¹¹ langweilt mich nicht 2. <Kontext> heitert mich auf 3. <Kontext> interessiert mich 4. <Kontext> engagiert mich 5. <Kontext> ist für mich bedeutsam 6. <Kontext> ist für mich nicht unwichtig 7. <Kontext> ist für mich nützlich	Schiefele (1990) Schiefele (1990) Schiefele (1990) Schiefele (1990) Schiefele (1990) Schiefele (1990) Schiefele (1990)
Vorausgegangene Zahlungen (Produktrezensionen)	Haben Sie in den vergangenen 12 Monaten für Produkttests und/oder Magazine, in denen Produkttests erscheinen, bezahlt?	O'Brien et al. (2020)

¹¹Hier ist die Textart gemeint, also Belletristik oder Produktrezension

Konstrukt	Frage	Quelle
Vorausgegangene Zahlungen (Belletristik)	Haben Sie in den vergangenen 12 Monaten für Unterhaltungsliteratur bezahlt?	O'Brien et al. (2020)
Kontrolle	1. Kannten Sie das Werk „Der Hammer der Götter“ von Wolfgang Hohlbein bereits vor der Teilnahme an diesem Experiment? ¹² 2. Kannten Sie das Werk „Die Blutschule“ von Max Rhode bereits vor der Teilnahme an diesem Experiment?	
Alter	Bitte geben Sie ihr Alter in Jahren an	
Einkommen	Wie hoch ist Ihr Nettohaushaltseinkommen pro Monat?	
Geschlecht	Bitte geben Sie ihr Geschlecht an	
Abschluss	Bitte geben Sie Ihren höchsten erreichten Abschluss an	

¹²Teilnehmende, die das Werk bereits kannten, wurden herausgefiltert, wie in Kapitel 3.2 beschrieben

Pfad	Regressionskoeffizient	σ_{SE}	p
Hypothesen			
H_1	-1.579	0.670	0.018
H_{2a}	-0.176	0.231	0.447
H_{2b}	0.166	0.532	0.754
H_{3a}	-0.832	0.233	0.000
H_{3b}	0.902	0.502	0.073
Kontrollvariablen auf ZB			
Alter	-0.102	0.090	0.257
Themeninteresse	0.011	1.493	0.994
TV ist KI	1.397	1.952	0.474
Ist Weiblich	3.069	1.7554	0.080
Ist Akademiker	0.616	1.741	0.724
VZ	6.565	3.086	0.033

Tabelle 5: Alle Pfade des Regressionmodells