

KI-Verfahren im Recruiting: Analyse von Persönlichkeit, Konsistenz, Ernsthaftigkeit und **Urheberschaft** (KaRiEbo)

Finanziert durch:

Adecco Stiftung

Datum:

17.03.2025

Autoren in alphabetischer Reihenfolge (Projektrolle):

Christin Buley, M.Sc. (Wissenschaftliche Mitarbeiterin)

Hannah Poppe (Studentische Hilfskraft)

Marc Roedenbeck, Prof. Dr. (Projektleitung)

Rainer Stollhoff, Prof. Dr. (Projektleitung)

Inhaltsverzeichnis

1	Einleitung	4
2	Theorie	4
3	Forschungsfragen.....	6
4	Methoden	7
4.1	Versuchsdesign.....	7
4.1.1	Aufgabe 1: Unabhängige Variable	7
4.1.2	Aufgabe 2: Unabhängige Variable	8
4.1.3	Technische Umsetzung	8
4.2	Studienvariablen	8
4.2.1	Akzeptanz von / Einstellung gegenüber dem Nutzen von KI im Recruiting	9
4.2.2	Augenschein-Validität von Anschreiben unterschiedlicher Herkunft	9
4.2.3	Person-Job Fit	10
4.2.4	Aufgebrachte Mühe zur Erstellung einer Bewerbung.....	11
4.2.5	Intention eine Intervieweinladung auszusprechen	11
4.2.6	Soziodemografische Variablen	11
4.2.7	Kontrollvariablen	12
4.3	Versuchsmaterial	14
4.3.1	Material Aufgabe 1: Auswahl der Anschreiben nach Schwierigkeit.....	15
4.3.2	Material Aufgabe 2: Auswahl der Anschreiben nach Eignung zur Stelle	17
4.4	Versuchsablauf.....	20
4.5	Datenauswertung.....	22
5	Datenanalyse	22
5.1	Auswertung des Experimentes	22
5.1.1	Deskriptive Auswertung	23
5.1.1.1	Stichprobe	23
5.1.1.2	Akzeptanz	23
5.1.1.3	Augenschein-Validität	24

5.1.2	Teststatistik.....	24
5.1.2.1	Kontrollvariablen: Anzahl Codes, Textlänge & Schwierigkeitsindex.....	24
5.1.2.2	Hypothese 1: Urheberschaft unterscheiden.....	26
5.1.2.3	Hypothese 2: Sichtbarkeit Urheberschaft & Mühe.....	27
5.1.2.4	Hypothese 3: Sichtbarkeit Urheberschaft & Eignung	27
5.1.2.5	Hypothese 4: Sichtbarkeit Urheberschaft & Intention	28
5.1.3	Qualitative Inhaltsanalyse Freitextfeld.....	29
5.1.3.1	Methode.....	29
5.1.3.2	Datenanalyse.....	30
5.1.3.3	Limitation qualitative Auswertung.....	34
6	Diskussion	35
7	Literaturverzeichnis	38

1 Einleitung

Im Personalkontext wird Künstliche Intelligenz (KI) für verschiedene Anwendungen diskutiert, wobei die Nutzung von KI durch Unternehmen im Fokus steht (Lederer et al., 2021). Aufgrund neuerer Entwicklungen (z.B. Einführung ChatGPT, November 2022) scheint sich eine Erweiterung des KI-Anwenderkreises abzuzeichnen. Laut einer Umfrage des Unternehmens softgarden e-recruiting GmbH (2023) nutzen immer mehr Bewerbende KI zur Formulierung ihrer Anschreiben. Wie sich jedoch die Nutzung von KI durch Bewerbende auf ihre Chancen im Bewerbungsprozess auswirkt, ist bisher wenig erforscht. Die Studie von Weiss et al. (2022) gibt erste Hinweise und zeigt, dass Bewerbende als weniger kompetent beurteilt und seltener für den weiteren Bewerbungsprozess berücksichtigt werden, wenn bei der Sichtung der Bewerbung erkennbar ist, dass KI genutzt wurde. Ein möglicher Erklärungsansatz ist, dass Beurteilende annehmen, dass Bewerbende weniger zum Anschreiben beigetragen haben, wenn KI zur Texterstellung genutzt wurde (Celiktutan et al., 2024). Der Eigenbeitrag der Bewerbenden wird von den Beurteilenden möglicherweise als ein Hinweis auf die Qualität der Bewerbung angesehen (Effort-Heuristic, Kruger et al., 2004). Die Nutzung von KI seitens der Bewerbenden scheint Signalwirkung zu haben, die hier differenzierter untersucht werden soll.

2 Theorie

Mit der Rolle von Signalen und deren Interpretation beschäftigt sich die Signal-Theorie (Spence, 1973). Konkret wird auf Basis einer Sender-Empfänger-Kommunikation (Connelly et al., 2011) angenommen, dass zwischen mind. zwei Akteuren (Sender, Empfänger) ein Informationsungleichgewicht besteht, dass durch den Austausch von positiven Informationen (= Signale) verringert werden soll. Dabei geht es insbesondere um den Austausch der für die andere Partei nicht direkt zugänglichen Informationen (Connelly et al., 2011).

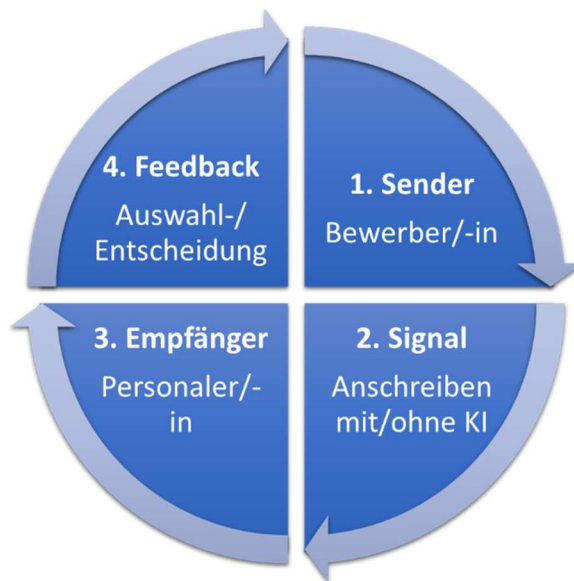


Abbildung 1. Elemente der Signaltheorie am Beispiel der Bewerbervorauswahl. Quelle: In Anlehnung an Connolly et al. (2011), S. 44.

Überträgt man die Elemente der Signaltheorie auf die Personalvorauswahl (Abbildung 1), zeigt sich, dass die Bewerbenden (= Sender) ihre Passung für eine vakante Stelle durch ihre Bewerbung (= Signal) kommunizieren. Die Personalverantwortlichen (= Empfänger) analysieren und interpretieren die Bewerbungsunterlagen dahingehend, ob die Person für die Stelle geeignet ist und teilen den Bewerbenden das Ergebnis anschließend mit (= Feedback).

Damit die Information, dass KI für die Anschreiben genutzt wurde, Signalwirkung entfalten kann, sind gemäß Signal-Theorie zwei Signalmerkmale (Kosten, Beobachtbarkeit) entscheidend (Connolly et al., 2011). Die Kosten beziehen sich hier auf die Transaktionskosten, die mit der Umsetzung des Signals einhergehen (Connolly et al., 2011). KI-Anwendungen, die frei zugänglich und kostenlos sind (z.B. ChatGPT), können von der breiten Bevölkerung genutzt werden. Daher werden die Kosten hier nicht weiter betrachtet. Die Beobachtbarkeit hingegen adressiert, inwieweit das Signal selbst zu erkennen ist (Connolly et al., 2011). Die Einschätzung, ob ein Text mit KI verfasst wurde, scheint Beurteilenden eher schwerzufallen (Clark et al., 2021). Daher wird die Beobachtbarkeit differenzierter in den folgenden Forschungsfragen betrachtet.

3 Forschungsfragen

Ausgehend von der Problemstellung zur Beobachtbarkeit, ob ein Text mit KI verfasst wurde oder nicht, lautet die erste Forschungsfrage, inwieweit es für Personalverantwortliche erkennbar ist, ob KI als Hilfsmittel für die Anschreiben genutzt wurde (1). Passend zur Forschungsfrage wurde folgende Hypothese formuliert:

Hypothese 1: Personalverantwortliche können nicht gut unterscheiden, ob ein Bewerbungsanschreiben von den Bewerbenden selbst (vs. mithilfe von KI) geschrieben wurde.

Wenn das Signal (Nutzung von KI) für Personalverantwortliche erkennbar ist, wird dieses gemäß der Signaltheorie (Connelly et al., 2011) interpretiert und beeinflusst die Auswahlentscheidung (Abbildung 1). Daher lautet die zweite Forschungsfrage, ob die erkennbare Nutzung von KI als Hilfsmittel für Anschreiben von den Personalverantwortlichen als geringere Eigenleistung der Bewerbenden (Celiktutan et al., 2024) interpretiert wird (2). Auch hier wurde passend zur Forschungsfrage folgende Hypothese formuliert:

Hypothese 2: Personalverantwortliche geben an, dass sich Bewerbende weniger Mühe gegeben haben, wenn auf den Anschreiben angezeigt wird, dass KI beim Verfassen genutzt (vs. KI nicht genutzt) wurde.

Um des Weiteren den Einfluss des Signals auf die Beurteilung im Bewerbungsprozess zu prüfen, lautet die dritte Forschungsfrage, wie sich die erkennbare Nutzung von KI als Hilfsmittel für Anschreiben auf den Bewerbungserfolg (Weiss et al., 2022) der Jobsuchenden auswirkt (3). Für diese Forschungsfrage wurden zwei Hypothesen aufgestellt:

Hypothese 3: Personalverantwortliche beurteilen Bewerbende als weniger geeignet, wenn auf den Anschreiben angezeigt wird, dass KI beim Verfassen genutzt (vs. KI nicht genutzt) wurde.

Hypothese 4: Personalverantwortliche laden Bewerbende mit einer geringeren Wahrscheinlichkeit zum Interview ein, wenn auf den Anschreiben angezeigt wird, dass KI beim Verfassen genutzt (vs. KI nicht genutzt) wurde.

4 Methoden

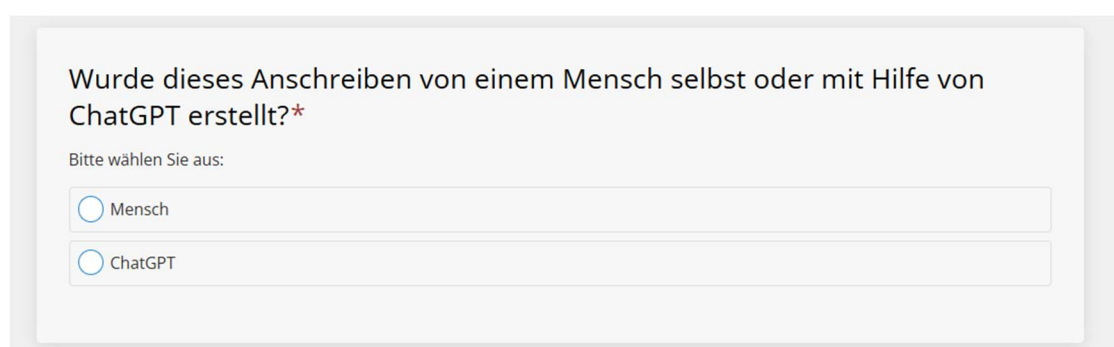
Im folgenden Abschnitt werden die verwendeten Methoden im Rahmen dieses Forschungsprojektes beschrieben.

4.1 Versuchsdesign

Zu Beginn der Studie wurden die Teilnehmenden per Zufall in zwei verschiedene Gruppen aufgeteilt. Um ein möglichst realitätsnahes Setting zu gestalten wurde im Experiment ein Bewerbungsprozess simuliert. Der Bewerbungsprozess fokussierte auf die Bewerbungsvorauswahl und umfasste zwei Aufgaben. Die Ausgangsbasis für die Bewerbungsvorauswahl bildete eine Stellenanzeige. In der Stellenanzeige wurde eine wissenschaftliche Hilfskraft für die Forschungsgruppe Management Analytics gesucht. Für die zwei Aufgaben bekamen die Teilnehmenden eine Reihe von Anschreiben zur Sichtung und Bewertung vorgelegt. Die Stellenanzeige blieb in allen Aufgaben gleich. Die Anschreiben hingegen änderten sich abhängig von der Aufgabe und der Gruppenzugehörigkeit. Das war aufgrund der experimentellen Manipulation notwendig.

4.1.1 Aufgabe 1: Unabhängige Variable

In der 1. Aufgabe wurde die *Urheberschaft* manipuliert, also ob ein Anschreiben von einem Menschen oder mithilfe von KI erstellt wurde. Das Datenset umfasst eine fixe Anzahl von Anschreiben ($n = 8$), die vorab randomisiert in eine Reihenfolge gebracht wurden. Diese Reihenfolge blieb in der Studie für alle Teilnehmenden gleich, da eine randomisierte Auswahl der Anschreiben auf Basis der Gruppenzugehörigkeit aus technischen Gründen nicht möglich war. Beide Gruppen bekamen somit die gleichen Anschreiben zur Bewertung vorgelegt.



Wurde dieses Anschreiben von einem Mensch selbst oder mit Hilfe von ChatGPT erstellt?*

Bitte wählen Sie aus:

☐ Mensch

☐ ChatGPT

Abbildung 2. Bewertung der Urheberschaft in Aufgabe 1

4.1.2 Aufgabe 2: Unabhängige Variable

In der 2. Aufgabe wurde die *Sichtbarkeit der Urheberschaft* manipuliert. Das Datenset umfasst zwei Anschreiben (A1 und A2), die beide von Menschen geschrieben wurden. Die Manipulation der Sichtbarkeit erfolgte indem auf den Anschreiben jeweils ein Piktogramm angebracht wurde (**Fehler! Verweisquelle konnte nicht gefunden werden.**). Die Piktogramme bezeichnen wir im weiteren Verlauf als „Label“. In jeder Gruppe wurde zuerst Anschreiben (A1) und danach Anschreiben (A2) beurteilt. Die Reihenfolge der Label auf den Anschreiben wurde zwischen den Gruppen ausbalanciert (**Fehler! Verweisquelle konnte nicht gefunden werden.**), um Sequenzeffekte zu kontrollieren.

Tabelle 1 Manipulation der Sichtbarkeit der Urheberschaft unterteilt nach Gruppen.

Gruppe	Anschreiben	Urheberschaft	Label
1	A1	Mensch	
	A2	Mensch	 Mensch
2	A1	Mensch	 Mensch
	A2	Mensch	

4.1.3 Technische Umsetzung

Für das Experiment wurde das Umfragetool Lamapoll verwendet. Für die Pilotierung der Anschreiben wurde die Umfragesoftware Quamp verwendet. Für beide Anwendungen liegen der Technischen Hochschule Wildau Nutzungslizenzen vor.

4.2 Studienvariablen

In der Studie wurden verschiedene Konstrukte durch die Teilnehmenden bewertet (vgl. 4.2). Bis auf das Freitextfeld (Abbildung 15) waren alle Konstrukte als Pflichtfragen konzipiert. Die Messung der Konstrukte wurde an vorhandene Skalen angelehnt und für das vorliegende Experiment angepasst. Dadurch sollte der Zeitaufwand für die Mehrfachbewertung

wiederkehrender Konstrukte bei unterschiedlichen Anschreiben in Grenzen gehalten werden. Nachfolgend werden die Konstrukte näher beschrieben.

4.2.1 Akzeptanz von / Einstellung gegenüber dem Nutzen von KI im Recruiting

Die Abfrage der Akzeptanz von KI im Bewerbungsverfahren wurde aus einer Studie von Lederer et al. (2021) entnommen. Lederer et al. (2021) betrachten die Akzeptanz der Bewerbenden, wenn KI von der Personalabteilung im Bewerbungsverfahren eingesetzt wird. Da im Rahmen der vorliegenden Studie jedoch die umgekehrte Perspektive untersucht werden sollte, waren Anpassungen notwendig. Das Item im Original „Wie empfinden Sie es, wenn KI für die Analyse / Bewertung Ihrer Bewerbung eingesetzt wird“ wurde abgewandelt und in Hinblick auf die Nutzung von KI zur Texterstellung von Anschreiben durch Bewerbende angepasst (Abbildung 3). Die originalen Antwortskalen (positiv, negativ, weiß nicht) von Lederer et al. (2021) wurden dahingehend angepasst, dass die Antwortoption „weiß nicht“ in „neutral“ geändert wurde, um die Akzeptanz breiter abzufragen (Abbildung 3).

Wie empfinden Sie es, wenn Bewerbende **KI-gestützte Anwendungen zur Texterstellung** (z.B. ChatGPT) zum Verfassen der Anschreiben einsetzen?

Negativ Neutral Positiv

☐ ☐ ☐

Abbildung 3. Item zur Akzeptanz.

4.2.2 Augenschein-Validität von Anschreiben unterschiedlicher Herkunft

Die Ausgangsbasis bildet der AKZEPT! Fragebogen von Kersting (2008). Die Augenschein-Validität wird in dem Fragebogen über vier Items erfasst. Für die vorliegende Studie wurde die Augenschein-Validität auf Basis von einem der vier Items erhoben. Dieses Item lautet in der Originalfassung „Dass man mit den Fragen/Aussagen, wie denen des Verfahrens, geeignete Personen für einen Job herausfinden kann, ist zu bezweifeln.“ Da das Konstrukt immer in Relation zum konkreten Auswahlverfahren zu bewerten ist, waren inhaltliche Anpassungen notwendig. In der vorliegenden Studie soll die Augenschein-Validität von Anschreiben unterschiedlicher Herkunft (Mensch / KI) bewertet werden. Daher wurde die Aussage zuerst positiv umformuliert und anschließend inhaltlich auf die traditionell erstellten Anschreiben und die Mithilfe von KI erstellten Anschreiben ausgerichtet (Abbildung 4). Zudem wurde die sechsstufige Antwortskala (von „trifft nicht zu“ bis „trifft genau zu“) aus der Originalversion in eine klassische 5er- Likert-Skala transformiert (Abbildung 4).

	Stimme überhaupt nicht zu.	Stimme eher nicht zu.	teils / teils	Stimme eher zu.	Stimme voll und ganz zu.
Auf Basis von mit KI-gestützten Anwendungen zur Texterstellung (z.B. ChatGPT) verfassten Anschreiben kann man geeignete Personen im Rahmen der Vorauswahl identifizieren.	☐	☐	☐	☐	☐
Auf Basis von traditionellen Anschreiben kann man geeignete Personen im Rahmen der Vorauswahl identifizieren.	☐	☐	☐	☐	☐

Abbildung 4. Items zur Augenschein-Validität.

4.2.3 Person-Job Fit

Der Person-Job Fit, also ob eine Person auf Grundlage ihres Anschreibens zur vorliegenden Stellenausschreibung passen könnte, wurde aus einer Studie der Stereotypenforschung entnommen (Dutz et al., 2022). Die Forschenden hatten ihren Fragenbogen auf den klassischen Studien zur Fit-Theorie (Lauver & Kristof-Brown, 2001) aufgebaut. Im Unterschied zur Studie von Lauver & Kristof-Brown (2001) haben Dutz et al. (2022) den Person-Job-Fit nicht zur Selbsteinschätzung sondern zur Fremdeinschätzung der Bewerbenden eingesetzt. Der Person-Job-Fit umfasst in beiden Studien (Dutz et al., 2022; Lauver & Kristof-Brown, 2001) fünf Items. Von diesen fünf Items fokussieren sich drei Items auf die Fähigkeiten und Fertigkeiten und zwei Items auf die Persönlichkeit einer Person. Daher wurden von den fünf Originalitems aus Dutz et al. (2022) ein Item zu Fähigkeiten/Fertigkeiten und ein Item zur Persönlichkeit ausgewählt und ins Deutsche übersetzt. Die zwei Items (Abbildung 5) wurden zu einer Skala zusammengefasst indem der Mittelwert berechnet wurde. Auch hier wurde nicht die 7er-Likert-Skala verwendet, sondern eine 5er-Likert-Skala (Abbildung 5).

Bitte bewerten Sie die folgenden Punkte

Hier klicken, um den Einleitungstext der Frage zu bearbeiten

	stimme überhaupt nicht zu	stimme eher nicht zu	teils / teils	stimme eher zu	stimme voll und ganz zu
Ich denke, die Person hat die richtigen Fähigkeiten und Fertigkeiten für diese Stelle.	☐	☐	☐	☐	☐
Ich denke, die Persönlichkeit der Person passt gut zu dieser Stelle.	☐	☐	☐	☐	☐

Abbildung 5. Items zum Person-Job Fit.

4.2.4 Aufgebrachte Mühe zur Erstellung einer Bewerbung

Die aufgebrachte Mühe wurde aus einer Studie zur Erforschung der Kreativität künstlicher Intelligenz entnommen (Magni et al., 2024). Die englischen Originalitems “The [person/AI] put a lot effort in coming up with the ideas” und “The [person/AI] worked hard to generate the ideas” wurden zusammengefasst, um die Mühe in einer Frage zu adressieren. Anschließend wurden diese ins Deutsche übersetzt. Zudem wurde der Inhalt an den Bewerbungskontext d.h. die Erstellung von Bewerbungen angepasst. Die originale 7er-Likertskala mit den Extremausprägungen „strongly disagree“ und „strongly agree“ wurde in eine klassische 5er-Likert-Skala transformiert (Abbildung 6).

Bitte bewerten Sie die folgenden Punkte

Hier klicken, um den Einleitungstext der Frage zu bearbeiten

stimme überhaupt nicht zu stimme eher nicht zu teils / teils stimme eher zu stimme voll und ganz zu

Ich denke, die Person hat sich beim Verfassen des Anschreibens sehr viel Mühe gegeben.

Abbildung 6. Item zur Mühe.

4.2.5 Intention eine Intervieweinladung auszusprechen

Die Intention der Personalverantwortlichen, auf Basis einer Bewerbung eine Intervieweinladung auszusprechen, wurde aus einer Studie zur Analyse der Reaktionen von Stellensuchenden auf Bewerbungsprozesse entnommen (Reeve & Schultz, 2004). Das englische Originalitem “How likely would you be to apply for this particular job” wurde auf die Beurteiler-Perspektive der Personalverantwortlichen ausgerichtet, ins Deutsche übersetzt und erneut die 5er Likert-Skala aus den beiden obigen Konstrukten übernommen (Abbildung 7).

Bitte bewerten Sie die folgenden Punkte

Hier klicken, um den Einleitungstext der Frage zu bearbeiten

stimme überhaupt nicht zu stimme eher nicht zu teils / teils stimme eher zu stimme voll und ganz zu

Ich denke, die Person würde ich zu einem Bewerbungsgespräch einladen.

Abbildung 7. Item zur Intention.

4.2.6 Soziodemografische Variablen

Als soziodemografische Variablen wurden drei Standardkategorien (Geschlecht, Alter, höchster Bildungsabschluss) erhoben. Das Geschlecht wurde differenzierter abgefragt, um jeder Person eine Zuordnung zu ermöglichen (Abbildung 8).

Bitte geben Sie Ihr Geschlecht an

Hier klicken, um den Einleitungstext der Frage zu bearbeiten

☒ weiblich

☐ männlich

☐ divers

☐ keine Angabe

Abbildung 8. Variable Geschlecht.

Das Alter wurde über verschiedene Altersgruppen erhoben (Abbildung 9).

Wie alt sind Sie?

Bitte wählen Sie die entsprechende Altersgruppe aus.

☒ 18 - 25 Jahre

☐ 26 - 35 Jahre

☐ 36 - 45 Jahre

☐ 46 - 55 Jahre

☐ älter als 55 Jahre

Abbildung 9. Variable Alter.

Der höchste Bildungsabschluss wurde differenziert nach Ausbildung und Studium abgefragt. Um zu unterscheiden, ob sich die Bildungsabschlüsse der Personalverantwortlichen hinsichtlich des Schwerpunkts Personal unterscheiden, wurde innerhalb von Studium und Ausbildung eine weitere fachliche Differenzierung vorgenommen (Abbildung 10).

Was ist Ihr beruflicher Hintergrund?

Bitte geben Sie Ihren höchsten Abschluss an.

☒ Keine Ausbildung

☐ Ausbildung (ohne Personal-Schwerpunkt)

☐ Ausbildung (mit Personal-Schwerpunkt)

☐ Studium (ohne Personal-Schwerpunkt)

☐ Studium (mit Personal-Schwerpunkt)

Abbildung 10. Variable höchster Bildungsabschluss.

4.2.7 Kontrollvariablen

Neben den soziodemografischen Variablen wurden zudem mehrere Kontrollvariablen überprüft, die aufgrund theoretischer Überlegungen und praktischer Erfahrung mit psychologischen Experimenten ausgewählt wurden. Die erste Kontrollvariable adressierte inwieweit sich die befragten Personen in ihrem Aufgabenbereich mit der Sichtung von

Bewerbungsunterlagen befassen. Der Anteil dieser Aufgabe wurde über die Arbeitszeit erhoben. Diese Information war auch für die Studienteilnahme relevant, da sich die Studie an Personen richtete, die im realen Arbeitsleben an der Sichtung von Bewerbungsunterlagen beteiligt sind.

Abbildung 11. Item zur Arbeitszeit von Recruiting-Aufgaben (Default 50%).

Als weitere Kontrollvariable wurde die Textlänge als mögliches Bewertungskriterium für die Beurteilung der Urheberschaft geprüft.

Abbildung 12. Item zur Textlänge.

Auch die Vorerfahrung mit KI-Experimenten wurde als Kontrollvariable mit einbezogen.

Abbildung 13. Item zur Vorerfahrung.

Auch wenn Probanden in der Studie explizit darauf hingewiesen wurden, dass sie die Signalworte aus der Pseudonymisierung / Anonymisierung (vgl. Buley et al., 2024) nicht als Hinweis auf die Urheberschaft verwenden sollten, wurde dieses Verhalten ebenfalls über eine Kontrollvariable abgefragt.

Haben Sie allgemeine Begriffe wie z.B. *A-Unternehmen* oder *W-Hochschule* als Hinweis dafür genutzt, ob ein Anschreiben mit Hilfe von ChatGPT erstellt wurde?

☐ Ja

☐ Nein

☐ Weiß ich nicht

Abbildung 14. Item zu Signalworten.

Die letzte Kontrollfrage war ein Freitextfeld für weiterführende Hinweise durch die Personalverantwortlichen, um Rückmeldung zu ihren individuellen Entscheidungskriterien zu erhalten. Die Beantwortung dieser Frage war für die Teilnehmenden optional.

Frage 13.

Welche Orientierungspunkte haben Sie ggf. zusätzlich für Ihre Entscheidungen genutzt?

Hier klicken, um den Einleitungstext der Frage zu bearbeiten

Orientierungspunkte/ Informationen

Abbildung 15. Freitextfeld für individuelle Entscheidungskriterien.

4.3 Versuchsmaterial

Für die zwei Experimentalaufgaben war es notwendig geeignetes Versuchsmaterial zu erstellen. Aus der Studie „KI im Recruiting“ (vgl. Buley et al., 2024) lag bereits eine Datenbasis mit 54 Anschreiben (27 KI, 27 Mensch) vor. Da in den zwei Experimentalaufgaben unterschiedliche Varianten der Urheberschaft (siehe 4.1) untersucht wurden, waren auch zwei verschiedene Datensets notwendig. Für die Experimentalaufgabe 1 wurde ein Datenset mit Anschreiben unterschiedlicher Urheberschaft (KI und Mensch) benötigt, da die reale Urheberschaft (Mensch oder KI) im Fokus lag. Für die Experimentalaufgabe 2 wurde ein Datenset benötigt, dass nur von Menschen verfasste Anschreiben enthält, da die KI-Urheberschaft diesmal nur vorgetäuscht werden sollte. Um geeignete Anschreiben auszuwählen wurden mehrere kleine Vorstudien durchgeführt, die getrennt nach den Experimentalaufgaben beschrieben werden (siehe 4.3.1 und 4.3.2). Diese Vorstudien werden im weiteren Verlauf als „Pilotierungen“ bezeichnet.

4.3.1 Material Aufgabe 1: Auswahl der Anschreiben nach Schwierigkeit

Für die Experimentalaufgabe 1 sollte ein Datenset ($n = 8$) zusammengestellt werden, dass Anschreiben mit unterschiedlichen Urheberschaften (KI, Mensch) zu gleichen Teilen enthält. Die Texte mit der Urheberschaft Mensch sind reale Anschreiben von Studierenden. Die Texte mit der Urheberschaft KI sind KI-generierte Versionen dieser realen Anschreiben (vgl. Buley et al., 2024). Die Texte weisen eine natürliche Variation hinsichtlich des Inhaltes und der Sprache auf. Die Zusammenstellung der Anschreiben erfolgte dabei in 3 Schritten, die in Abbildung 16 dargestellt werden.



Abbildung 16. Prozess der Pilotierung der Anschreiben für das Datenset der Aufgabe 1.

Im ersten Schritt wurde überprüft, ob sich die verwendeten Anschreiben aus dem Vorexperiment aufgrund der unterschiedlichen Eingabeaufforderungen beim KI-Prompt (vgl. Buley et al., 2024, Kap. 4.5.3) hinsichtlich ihrer Qualität unterscheiden. Diese Einschätzung erfolgte innerhalb der Forschungsgruppe, wobei die befragten Forschenden ($n = 4$) gebeten wurden, auf Basis der vorliegenden Anschreiben zu entscheiden, ob die Anschreiben mithilfe von KI oder von einem Menschen selbstständig verfasst wurden. Pro Anschreiben sollte auf Grund der Gesamtmenge mind. zwei Urteile vorliegen. Hieraus wurde die Trefferquote ermittelt, also wie viele Anschreiben (einerseits Urheber Mensch, andererseits Urheber KI) als richtig erkannt wurden. Es zeigten sich jedoch keine großen Unterschiede und die finale Entscheidung fiel daher auf solche KI-generierten Anschreiben, deren Eingabeaufforderung Persönlichkeitsmerkmale enthielt. Die Entscheidung basierte auf der Überlegung, dass

Jobsuchende am wahrscheinlichsten Lebenslaufdaten und Angaben zur Persönlichkeit bei ChatGPT angeben, um das Anschreiben mit einer persönlichen Note zu generieren. Als Ergebnis verblieben alle vom Menschen verfassten Anschreiben (27) und die KI generierten Anschreiben wurden auf 27 reduziert.

Im zweiten Schritt wurde überprüft, ob sich die Anschreiben hinsichtlich der Schwierigkeit die Urheberschaft zu erkennen, unterschieden. Im Rahmen einer Pilotierung sollten die Teilnehmenden entscheiden, ob die Anschreiben mithilfe von KI oder von einem Menschen verfasst wurden. Die 54 aus dem ersten Schritt vorliegenden Anschreiben wurden über die Umfragesoftware Quamp randomisiert auf drei Experimentalgruppen mit jeweils 18 Anschreiben (9 KI, 9 Mensch) aufgeteilt. An der Pilotierung nahmen Studierende aus zwei verschiedenen Seminargruppen innerhalb der regulären Lehrveranstaltung teil. Die Teilnahme dauerte ca. 20 Minuten. Die erste Umfrage wurde im November 2023 in der Seminargruppe Bachelor „Wirtschaft & Recht“ durchgeführt. Es nahmen 19 Studierende teil, davon füllten 12 die Umfrage vollständig aus. Die zweite Umfrage wurde im Dezember 2023 in der Seminargruppe Master „Wirtschaft & Recht“ 2022 durchgeführt. Sieben Studierende nahmen teil, davon füllten alle die Umfrage vollständig aus. Die Studierenden erhielten ca. eine Woche nach der jeweiligen Umfrage eine allgemeine Auswertung über ihre Ergebnisse.

Auf Basis der Einschätzung wurde im dritten Schritt ein Schwierigkeitsindex über alle Anschreiben gebildet (Bühner, 2004):

$$P_i = \frac{n_R - \frac{n_F}{k-1}}{n} * 100$$

Abbildung 17. Formel des Schwierigkeitsindexes.

In der obigen Formel (Abbildung 17) entspricht n_R der Anzahl an richtigen Antworten für ein Anschreiben (KI erkannt als KI, Mensch erkannt als Mensch), n_F der Anzahl an falschen Antworten für ein Anschreiben (KI erkannt als Mensch, Mensch erkannt als KI) und n der Anzahl an Bearbeitungen pro Anschreiben. Der in der Formel enthaltene Divisor für n_F von $k-1$ (wobei k die Anzahl an Antwortoptionen darstellte) wurde durch die binäre Wahl (2 Antwortoptionen) zu 1 aufgelöst und entfiel somit.

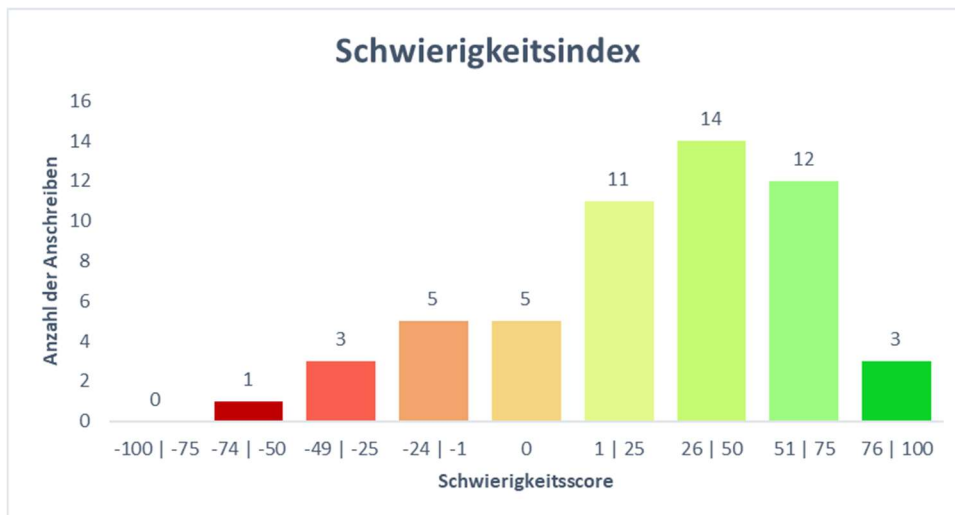


Abbildung 18. Verteilung der Anschreiben auf dem Schwierigkeitsindex.

Der Schwierigkeitsindex umfasst einen Wertebereich von -100 und +100 (Abbildung 18). In der Pilotierung sollten auf Basis theoretischer Überlegungen 2 schwerere (1x KI, 1x Mensch), 4 mittlere (2x KI, 2x Mensch) und 2 einfachere (1x KI, 1x Mensch) Anschreiben ausgewählt werden. Daher wurden Anschreiben mit positiven/negativen Ergebnissen aber gleichem Betrag ausgewählt. Diese lagen bei -20 und +20 vor. Alle Ergebnisse gleich -20 wurden als schwer, Ergebnisse gleich 0 als mittel und Ergebnisse gleich +20 als einfach gewertet. Da bei der Kategorie -20 vier Anschreiben (Ziel: 2 Anschreiben) und bei der Kategorie 0 fünf Anschreiben (Ziel: 4 Anschreiben) vorlagen, wurde eine Zufallsziehung mittels webbasierten Zufallsgenerator (Quelle: zufallsgenerator.net) durchgeführt. Als finales Ergebnis lag ein Datenset im Umfang von 8 Anschreiben vor, dass die Bewertungsgrundlage für die Experimentalaufgabe 1 in der vorliegenden Studie bildet.

4.3.2 Material Aufgabe 2: Auswahl der Anschreiben nach Eignung zur Stelle

Für die Experimentalaufgabe 2 sollte ein Datenset zusammengestellt werden, dass nur reale d.h. von Studierenden verfasste Anschreiben enthält. Die Lebensläufe und Persönlichkeiten der Studierenden unterscheiden sich naturgemäß, sodass die Texte der Anschreiben dahingehend variieren. Somit war es vorab notwendig die Eignung der jeweiligen Person zu bestimmen, um diese als Einflussfaktor kontrollieren zu können. Das bedeutet, dass jedes Anschreiben (n=27) vor dem Hintergrund der Stellenanzeige (Anforderungen, Aufgaben) bewertet werden musste. Dadurch sollte für das Experiment sichergestellt werden, dass die Variation in der abhängigen Variablen (Eignung) auf den Einfluss der unabhängigen Variable (Urheberschaft) zurückgeführt werden kann und nicht auf reale Unterschiede in der Eignung

der Bewerbenden. Um die Eignung zur ausgeschriebenen Stelle zu bestimmen wurde eine zweistufige Pilotierung durchgeführt, die in Abbildung 19 dargestellt wird.

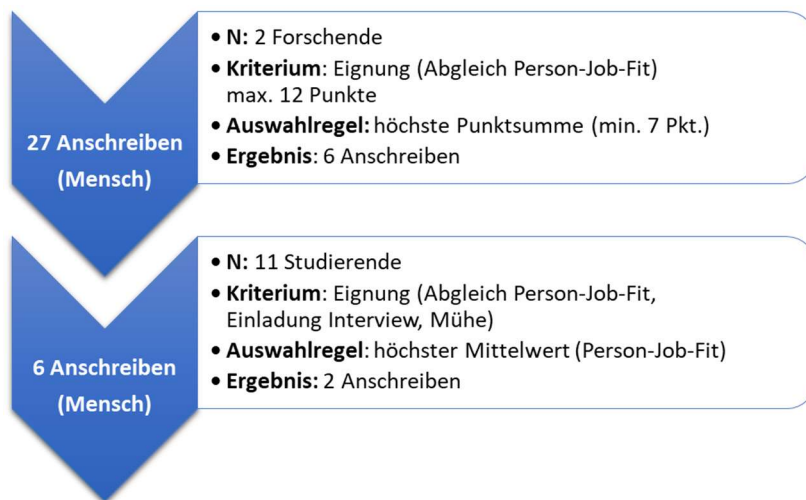


Abbildung 19. Prozess der Pilotierung der Anschreiben für das Datenset der Aufgabe 2.

In der jeweiligen Stufe wurde die Eignung unterschiedlich operationalisiert und von unterschiedlichen Personengruppen beurteilt. In der ersten Stufe wurde die Eignung der Bewerbenden projektintern von zwei Mitarbeiterinnen (MA) beurteilt. Auf Basis der

Freude / Bereitschaft Forschung	gewissenhafter Arbeitsstil	Offenheit datengetriebene Methoden	sichere Office- Kenntnisse	gute Deutschkenntnisse	Projektkoordinatio n	Protokollierung Meetings	Veranstaltung smanagement	Berichte verfassen	Transkription	Erfahrung KI-Tools	externe Kommunikation (E-Mail)
1	1	1	0	1	0	0	0	0	0	0	0
1	0	1	0	1	1	0	0	0	1	0	0
1	1	1	1	1	1	0	0	0	0	1	1
1	0	1	0	1	0	0	0	0	0	0	0

Abbildung 20. Auszug aus dem Bewertungsschema. Gelb sind die Kriterien basierend auf den Anforderungen, Grau sind Kriterien basierend auf den Aufgaben aus der Stellenausschreibung.

Stellenausschreibung (Anforderungen, Aufgaben) wurde ein Bewertungsschema erstellt, das 12 Kriterien umfasst (Abbildung 20). Zusätzlich zu den Anforderungen wurden auch die Aufgaben als Bewertungskriterium herangezogen, da nicht ausgeschlossen werden konnte, dass Beurteilende auch diese Aspekte bei der Eignungsbeurteilung mitberücksichtigen. Die zwei MA beurteilten unabhängig voneinander alle 27 Anschreiben auf Basis dieser Kriterien. Die Ergebnisse der MA wurden verglichen. In den Fällen, wo es zu unterschiedlichen Einschätzungen kam, wurden diese nochmal besprochen und es wurde ein finales Urteil gefällt. Als Untergrenze wurde festgelegt, dass nur abweichende Urteile besprochen werden, wo min. eine von zwei MA 6 Punkte vergeben hatte. Diese Einschränkung wurde vorgenommen, da nur die geeignetsten Anschreiben in das finale Experiment einfließen sollten. Nachdem alle abweichenden Urteile besprochen und die finalen Bewertungen

vorlagen, wurden die geeignetsten Anschreiben ausgewählt. Die geeignetsten Anschreiben wurden definiert als Anschreiben, die final min. 7 von 12 Kriterien erfüllt hatten. Von den ursprünglich 27 Anschreiben haben 6 Anschreiben dieses Kriterium erfüllt (Tabelle 2) und dienen als Datenset für die zweite Stufe (Abbildung 19Tabelle 2).

Anschreiben Nr.	Punktsumme Kriterien (Min 1 Max 12)
1	8
2	8
3	7
4	7
5	8
6	7

Tabelle 2. Punktsummen der Kriterien über die getesteten Anschreiben.

In der zweiten Stufe wurde die Eignung der Bewerbenden von Studierenden (n= 11) beurteilt. Die Pilotierung fand im Rahmen einer regulären Lehrveranstaltung statt. Abweichend von der ersten Stufe wurde die Eignung über den Person-Job-Fit (2 Items, Abbildung 5) erhoben. Diese Items bilden auch die Bewertungsgrundlage für das spätere Experiment. Ziel war es die beiden geeignetsten Anschreiben auszuwählen. Für jedes Anschreiben (n=6) wurde der Person-Job-Fit berechnet. Dafür wurden beide Person-Job-Fit Items (Abbildung 5) zu einer Skala zusammengefasst, indem der Mittelwert gebildet wurde.

Anschreiben Nr.	Mittelwert Person-Job-Fit (Min 1 Max 5)
1	3.27
2	3.27
3	3.95
4	3.86
5	3.14
6	3.36

Tabelle 3. Mittelwerte des Person-Job-Fit über die getesteten Anschreiben.

Final wurden die zwei Anschreiben (Nr. 3 & 4, Tabelle 3) mit dem höchsten Mittelwert ausgewählt und es wurde sichergestellt, dass diese Anschreiben nicht bereits in Aufgabe 1 als

Untersuchungsmaterial verwendet wurde. Als finales Ergebnis liegt ein Datenset im Umfang von 2 Anschreiben vor, dass die Bewertungsgrundlage für die Experimentalaufgabe 2 in der vorliegenden Studie bildet.

4.4 Versuchsablauf

Die Datenerhebung fand im Zeitraum vom Mai bis Juli 2024 statt. Im Ergebnis haben 166 Personalverantwortliche an der Online-Studie teilgenommen. Die Teilnahme war freiwillig und wurde nicht vergütet. Voraussetzungen für die Teilnahme waren Volljährigkeit, gute

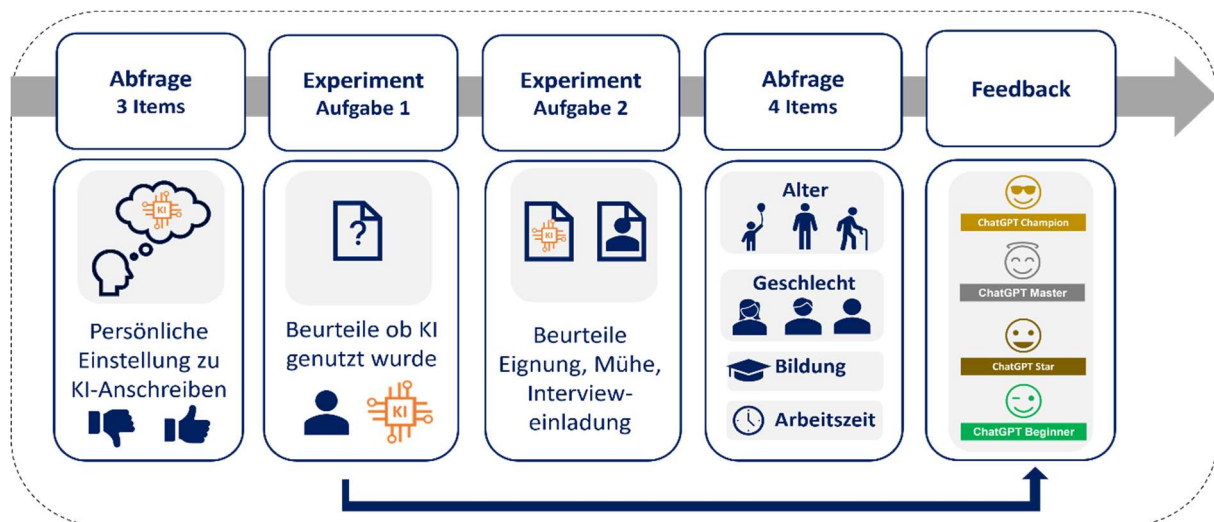


Abbildung 21. Gesamtes Studiendesign.

Deutschkenntnisse und die Beteiligung an der Sichtung von Bewerbungen im Berufsalltag. Die Probandenakquise erfolgte über Unternehmenskontakte/-netzwerke und die sozialen Medien.

Zu Beginn der Online-Studie wurde den Personalverantwortlichen mitgeteilt, dass die Studie sich mit der Nutzung von KI zur Texterstellung im Bewerbungsprozess befasst. Die Personalverantwortlichen beantworteten zuerst zwei Fragen zu ihrer Einstellung bezogen auf den Einsatz von KI zur Erstellung von Anschreiben (Akzeptanz, 3.3.1; Augenschein-Validität, 3.3.2).

Danach wurde auf Basis einer realen Stellenausschreibung ein Personalauswahlverfahren simuliert, das zwei Aufgaben umfasste. In der ersten Aufgabe wurden nacheinander acht verschiedene Anschreiben (davon 50% KI-generiert) dahingehend beurteilt, ob das vorliegende Anschreiben mithilfe von ChatGPT geschrieben wurde (Abbildung 2). In der zweiten Aufgabe wurden nacheinander zwei Anschreiben (davon 0% KI-generiert) beurteilt. Im Unterschied zur ersten Aufgabe wurden beide Anschreiben tatsächlich ohne KI erstellt.

Jedoch wurde mithilfe von zwei verschiedenen Symbolen (Abbildung 21, Tabelle 1) gezielt ein anderer Eindruck vermittelt. Das KI-Symbol täuschte somit nur vor, dass das gekennzeichnete Anschreiben mithilfe von KI geschrieben worden war. Die Reihenfolge der Symbole wurde im Experiment über alle Durchgänge ausbalanciert. Auf Basis des jeweiligen Anschreibens bestand der Auftrag an die Personalverantwortlichen darin zu beurteilen, ob eine Person auf Grundlage ihres Anschreibens zur vorliegenden Stellenausschreibung passen könnte (Person-Job Fit, Kap. 3.3.3), wie viel Mühe sich die Person bei der Erstellung des Anschreibens gegeben hat (Mühe, Kap. 3.3.4) und ob die Person zu einem Interview eingeladen werden würde (Intention, Kap. 3.3.5).

Nachdem die erste Aufgabe absolviert worden war, beantworteten die Personalverantwortlichen zwei Kontrollvariablen (Signalwortnutzung, Textlänge) sowie optional die Frage, an welchen Kriterien sie sich bei Beurteilung der Urheberschaft orientiert hatten (Abbildung 15). Nachdem die zweite Aufgabe absolviert worden war, beantworteten die Personalverantwortlichen soziodemografischen Fragen (Geschlecht, Alter, höchster Bildungsabschluss, 4.2.6) und zwei Kontrollvariablen (Rekrutierungs-Arbeitszeit, Vorerfahrung mit KI, 4.2.7). Am Ende bekamen alle Teilnehmenden ihren persönlichen ChatGPT-Score mitgeteilt, d.h. die Quote der richtig erkannten Anschreiben in der ersten Aufgabe (Tabelle 4). Zudem wurden die Personalverantwortlichen über das Ziel der Studie aufgeklärt (Debriefing).

Medaille	Leistungsfeedback	Wertebereich
 ChatGPT Champion	Wow, Sie sind ein echter KI-Profi! Ihr ChatGPT-Score ist x von 8 Punkten. ChatGPT konnte Sie nicht in die Irre führen.	8 Punkte
 ChatGPT Master	Nicht schlecht! Ihr ChatGPT-Score beträgt x von 8 Punkten. Sie haben schon Einiges drauf, was KI angeht!	6-7 Punkte
 ChatGPT Star	Dieses Mal hat ChatGPT Sie noch überzeugen können. Ihr ChatGPT-Score beträgt x von 8 Punkten. Aber das war schon ein guter Anfang.	5 Punkte

 ChatGPT Beginner	Das war leider noch nicht besser als ein Münzwurf. Ihr ChatGPT-Score beträgt x von 8 Punkten. ChatGPT hat Sie dieses Mal noch in die Irre führen können.	≤ 4 Punkte
--	--	-----------------

Tabelle 4. ChatGPT Score unterteilt nach Leistungsniveau.

Am Ende der Studie erhielten die Personalverantwortlichen zudem die Möglichkeit einzusehen, welche der Anschreiben mithilfe von KI und welche vom Menschen selbst erstellt wurden und wie sie die Frage in Studie beantwortet hatten (Abbildung 22).

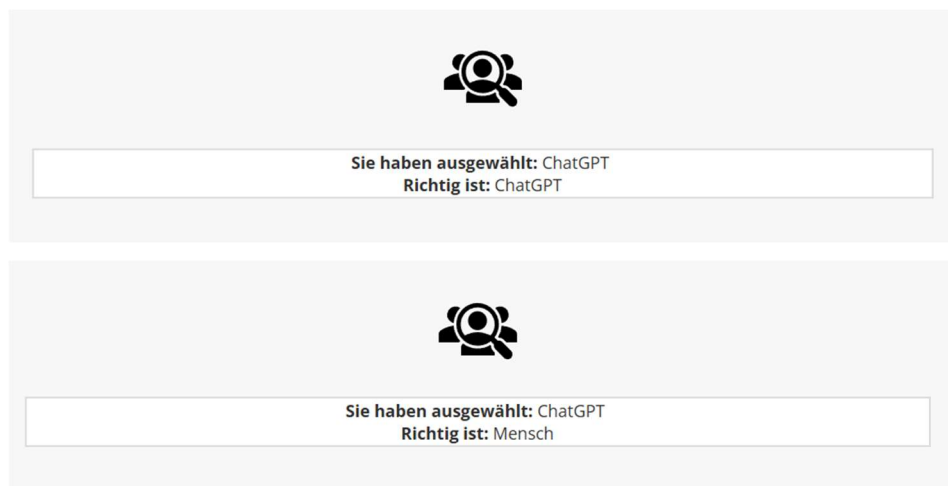


Abbildung 22. Feedbackbeispiele.

4.5 Datenauswertung

Die statistische Auswertung erfolgte mit der Programmiersprache R (Version 4.1.1). Zusammenhänge zwischen zwei metrischen Variablen wurden mit parametrischen Verfahren getestet (Spearman's ρ : stats::cor.test). Zusammenhänge zwischen metrischen und ordinalen/nominalen Variablen wurden mit nicht-parametrischen Verfahren getestet (Wilcoxon- Rangsummen-Test: stats::wilcox.test).

5 Datenanalyse

5.1 Auswertung des Experimentes

Die Auswertung des Experimentes wird getrennt nach deskriptiver Statistik, Teststatistik und qualitativer Inhaltsanalyse berichtet.

5.1.1 Deskriptive Auswertung

Die deskriptive Auswertung bezieht sich dabei auf die Stichprobe, sowie die zwei Einschätzungsvariablen (Akzeptanz und Augenschein-Validität).

5.1.1.1 Stichprobe

An der Online-Studie haben 166 Personen teilgenommen. Die soziodemografischen Fragen, die erst am Ende der Studie abgefragt wurden, haben 113 Versuchspersonen beantwortet. Daher beziehen sich die nachfolgenden Angaben auf die kleinere Stichprobe.

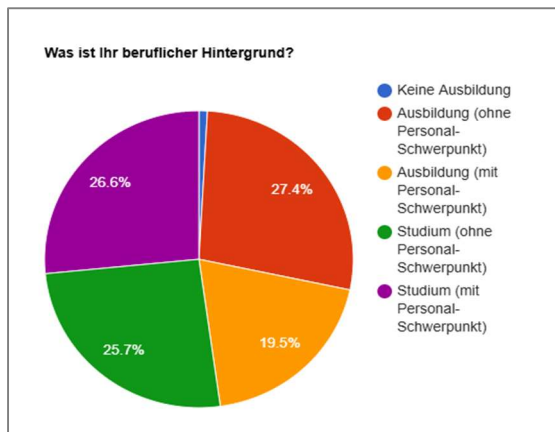


Abbildung 23. Stichprobe unterteilt nach höchstem Bildungsabschluss, n= 113.

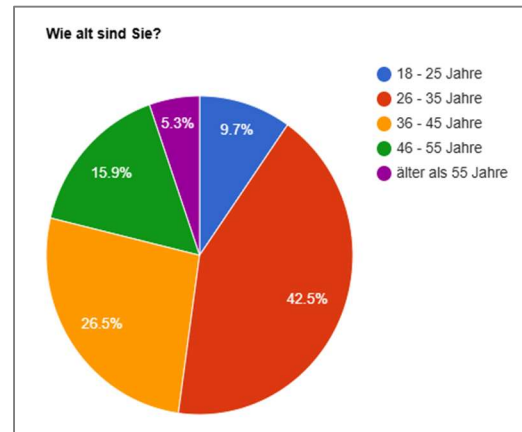


Abbildung 24. Stichprobe unterteilt nach Altersgruppen, n= 113.

Von den Teilnehmenden waren 70% weiblich. Das Alter der Versuchspersonen lag bei 69% der Teilnehmenden im Bereich von 26 bis 45 Jahren (MD = Altersgruppe 26-35 Jahre). Der höchste Bildungsabschluss war bei 47% der Versuchspersonen die Ausbildung (davon hatten 41,5% Personalschwerpunkt) sowie bei 52% das Studium (davon 51% Personalschwerpunkt). Lediglich eine Versuchsperson hatte keinen Bildungsabschluss. Im Durchschnitt verbrachten die Versuchspersonen 47,62% (SD = 23.13) ihrer Arbeitszeit mit der Sichtung von Bewerbungsunterlagen. Wobei es hier eine große Streuung (SD = 23.12) gab. Das könnte mit den unterschiedlichen Aufgabenbereichen bzw. Hierarchien im Personalbereich zusammenhängen. Dazu liegen jedoch keine weiteren Daten vor. Von den Versuchspersonen hatten bereits 4 der Teilnehmenden schon einmal an einer Studie zum Thema KI zur Texterstellung teilgenommen.

5.1.1.2 Akzeptanz

Die Personalverantwortlichen wurden befragt, wie sie es empfinden, wenn Bewerbende KI-gestützte Anwendungen zur Texterstellung (z.B. ChatGPT) zum Verfassen von

Bewerbungsanschreiben einsetzen. Von den befragten Personalverantwortlichen (n=166) finden es 27% positiv, wenn KI genutzt wird. Weitere 58% haben eine neutrale Haltung dazu. Lediglich 15% der Befragten empfinden das als negativ, wenn Bewerbende KI zur Texterstellung für ihr Anschreiben benutzen.

Antworten	Negativ	Neutral	Positiv
Anzahl	25	97	44
Prozent	15.06	58.43	26.51

Tabelle 5. Beurteilung der Akzeptanz.

5.1.1.3 Augenschein-Validität

Die Personalverantwortlichen wurden befragt, inwieweit Anschreiben grundsätzlich in der Vorauswahl geeignet sind, um geeignete Kandidaten zu identifizieren. Dabei wurde differenziert nach Anschreiben, die traditionell versus mithilfe von KI verfasst wurden, gefragt. Die traditionell verfassten Anschreiben (MD = 4) wurden von den Personalverantwortlichen als besser für die Vorauswahl geeignet beurteilt im Vergleich zu den mithilfe von KI (MD = 3) verfassten Anschreiben (Wilcoxon, n= 166, V= 499.5, $p < 0.01$).

5.1.2 Teststatistik

Bei der Teststatistik werden die vier Hypothesen aus dem Kapitel der Forschungsfragen geprüft.

5.1.2.1 Kontrollvariablen: Anzahl Codes, Textlänge & Schwierigkeitsindex

Die nachfolgenden Analysen beziehen sich auf das Material (Anzahl Codes, Textlänge) und die Ergebnisse (Schwierigkeitsindex) aus Aufgabe 1. Die im weiteren Verlauf genutzte Erfolgsrate bezieht sich auf die Quote der richtig erkannten Anschreiben.

Für die Anschreiben (n= 8) wurde ermittelt, ob es einen Zusammenhang zwischen der Anzahl der Codes im Text und der Erfolgsrate gibt. Es zeigt sich kein signifikanter Zusammenhang (Spearman's $\rho = -0.605$, $p > .05$). Des Weiteren wurde geprüft ob es einen Zusammenhang zwischen der Textlänge (d.h. Anzahl der Worte) und der Erfolgsrate gibt. Hier zeigt sich ebenfalls kein signifikanter Zusammenhang (Spearman's $\rho = 0.595$, $p > .05$).

Darüber hinaus wurde geprüft, ob sich KI-generierte versus von Menschen verfasste Anschreiben hinsichtlich der Kontrollvariablen (Anzahl Codes, Textlänge) unterscheiden. Die

Unterschiede hinsichtlich der Anzahl der Codes wurden statistisch geprüft. Es zeigten sich keine signifikanten Unterschiede zwischen Anzahl Codes und Texturheber (Wilcoxon, $W = 6$, $p > 0.05$). Die Unterschiede hinsichtlich der Textlänge wurden statistisch geprüft. Es zeigten sich keine signifikanten Unterschiede zwischen Textlänge und Texturheber (Wilcoxon, $W = 10$, $p > 0.05$).

Anschreiben Nr.	Urheber	Anzahl		Erfolgsrate
		Worte	Codes	
1	KI	185	1	0.440
2	Mensch	265	1	0.416
3	Mensch	428	2	0.520
4	KI	326	3	0.456
5	Mensch	284	0	0.680
6	KI	130	2	0.400
7	KI	302	0	0.752
8	Mensch	297	1	0.672

Tabelle 6. Anschreiben unterteilt nach Urheberschaft, Worte, Codes und Erfolgsrate.

Zusammenfassend zeigt sich, dass weder die Anzahl der Codes noch die Textlänge in Zusammenhang mit der Erfolgsrate in Aufgabe 1 stehen. Zudem zeigen sich keine Unterschiede zwischen den Texten (KI vs. Mensch) hinsichtlich der Codeanzahl und Textlänge. Daher werden diese Variablen nicht weiter in der Auswertung berücksichtigt.

Die in der Studie verwendeten Anschreiben ($n = 8$) wurden auf Basis eines im Vorfeld ermittelten Schwierigkeitsindex ausgewählt (4.3.1). Es wurde angenommen, dass die Schwierigkeitsverteilung der Studierenden auch für die befragten Personalverantwortlichen valide ist (Tabelle 7). Es zeigt sich jedoch kein Zusammenhang zwischen der ermittelten Schwierigkeit von Studierenden und Experten (Spearman's $\rho = 0.077$, $p > .05$). Es ist daher nicht notwendig, die Schwierigkeit als zusätzliche Variable in die Auswertungen aufzunehmen.

Tabelle 7 Schwierigkeitsindex unterteilt nach Stichproben

Anschreiben Nr.	Schwierigkeitsindex	
	Studierende	Personalverantwortliche
1	0	-13.67
2	0	51.18
3	-20	2.22
4	0	-7.46
5	20	34.88
6	20	-21.26
7	-20	-20.29
8	0	34.40

5.1.2.2 Hypothese 1: Urheberschaft unterscheiden

In der ersten Hypothese wurde angenommen, dass Personalverantwortliche nicht gut unterscheiden können, ob ein Bewerbungsanschreiben von den Bewerbenden selbst (vs. mithilfe von KI) geschrieben wurde. In der Tabelle 8 werden die mittleren Erfolgsraten der Personalverantwortlichen aufgeführt. Da bei der Beurteilung der Anschreiben eine 50%-Wahrscheinlichkeit bestand, durch Zufall richtig zu tippen, wurde gegen diesen festen Wert ($\mu = 0.5$) getestet. Der Mittelwert der individuellen Erfolgswahrscheinlichkeiten (MW = 0.54) liegt knapp über dem Zufall ($\mu = 0.5$). Insbesondere ist der Anteil der Personalverantwortlichen, die besser als der Zufall sind, signifikant größer als der Anteil der Personalverantwortlichen, die schlechter als der Zufall sind (Wilcoxon, $n = 125$, $V = 3068$, $p = .01$).

	Level Erfolgsrate								
	0/8 (0%)	1/8 (12,5%)	2/8 (25%)	3/8 (37,5%)	4/8 (50%)	5/8 (62,5%)	6/8 (75%)	7/8 (87,5%)	8/8 (100%)
Anzahl VP	1	1	13	29	27	21	18	13	2

Tabelle 8. Mittlere Erfolgsrate der Teilnehmenden ($n = 125$).

5.1.2.3 Hypothese 2: Sichtbarkeit Urheberschaft & Mühe

In der zweiten Hypothese wurde angenommen, dass Personalverantwortliche angeben, dass sich Bewerbende weniger Mühe gegeben haben, wenn auf den Anschreiben angezeigt wird, dass KI beim Verfassen genutzt (vs. KI nicht genutzt) wurde.

Label	Mühe (Anschreiben 1)					n = 115
	1	2	3	4	5	
KI	2	18	13	20	2	n _{KI} = 55
Mensch	2	7	14	25	12	n _{Mensch} = 60

Tabelle 9. Häufigkeitsverteilung der Mühe unterteilt nach Label (Anschreiben 1).

Im Ergebnis zeigt sich, dass sich Personalverantwortliche bei der Beurteilung, ob sich die Person Mühe gegeben hat, durch das KI-Symbol beeinflussen ließen (Wilcoxon-Test; Anschreiben 1: n= 115, W= 2173, p < .01; Anschreiben 2: n= 113, W= 2328, p< 0.1).

Label	Mühe (Anschreiben 2)					n = 113
	1	2	3	4	5	
KI	5	22	15	12	5	n _{KI} = 59
Mensch	0	5	14	25	10	n _{Mensch} = 54

Tabelle 10. Häufigkeitsverteilung der Mühe unterteilt nach Label (Anschreiben 2).

5.1.2.4 Hypothese 3: Sichtbarkeit Urheberschaft & Eignung

In der dritten Hypothese wurde angenommen, dass Personalverantwortlichen Bewerbende als weniger geeignet beurteilen, wenn auf den Anschreiben angezeigt wird, dass KI beim Verfassen genutzt (vs. KI nicht genutzt) wurde. Die Skala setzt sich aus 2 Items zusammen, sodass aufgrund der Mittelwertberechnung auch Kommawerte resultieren. Im Ergebnis kann man nicht feststellen, dass sich Personalverantwortliche bei der Beurteilung der Eignung vom KI-Symbol beeinflussen ließen (Wilcoxon -Test; Anschreiben 1: n= 115, W= 1713.5, p > 0.05; Anschreiben 2, n=113, W= 1675.5, p> 0.05).

Label	Person-Job-Fit (Anschreiben 1)									n = 115
	1	1.5	2	2.5	3	3.5	4	4.5	5	
KI	0	0	2	2	12	15	19	2	3	n _{KI} = 55
Mensch	0	1	2	4	9	16	19	7	2	n _{Mensch} = 60

Tabelle 11. Häufigkeitsverteilung Person-Job-Fit unterteilt nach Label (Anschreiben 1)

Unterschiede bei der Eignungsbeurteilung kamen eher durch interindividuelle Unterschiede in der Urteilstendenz zustande und weniger durch die Information, ob KI genutzt wurde. Das auf den Anschreiben eingeblendete KI-Symbol hatte somit keinen Einfluss auf die Eignungsbeurteilung.

Label	Person-Job-Fit (Anschreiben 2)									
	1	1.5	2	2.5	3	3.5	4	4.5	5	n = 113
KI	0	2	8	8	8	12	16	2	3	n _{KI} = 59
Mensch	0	1	1	7	16	12	12	2	3	n _{Mensch} = 54

Tabelle 12. Häufigkeitsverteilung Person-Job-Fit unterteilt nach Label (Anschreiben 2)

5.1.2.5 Hypothese 4: Sichtbarkeit Urheberchaft & Intention

In der vierten Hypothese wurde angenommen, dass Personalverantwortliche Bewerbende mit einer geringeren Wahrscheinlichkeit zum Interview einladen, wenn auf den Anschreiben angezeigt wird, dass KI beim Verfassen genutzt (vs. KI nicht genutzt) wurde. Im Ergebnis zeigte sich, dass sich Personalverantwortlichen bei der Beurteilung, ob Sie die Person zum Interview einladen wollen, nicht vom KI-Symbol beeinflussen ließen (Wilcoxon-Test; Anschreiben 1: n= 115, W= 1761.5, p > .05; Anschreiben 2: n= 113, W= 1741, p > .05).

Label	Intention Interview (Anschreiben 1)					
	1	2	3	4	5	n = 115
KI	0	6	11	26	12	n _{KI} = 55
Mensch	2	5	9	27	17	n _{Mensch} = 60

Tabelle 13. Häufigkeitsverteilung Intention für ein Interview unterteilt nach Label (Anschreiben 1).

Das auf den Anschreiben eingeblendete KI-Symbol hatte somit keinen Einfluss auf die Intention, die BewerberInnen zum Interview einzuladen. Das Ergebnis steht im Einklang zur Eignungsbeurteilung.

Label	Intention Interview (Anschreiben 2)					
	1	2	3	4	5	n = 113
KI	1	11	22	17	8	n _{KI} = 59
Mensch	0	9	15	25	5	n _{Mensch} = 54

Tabelle 14. Häufigkeitsverteilung Intention für ein Interview unterteilt nach Label (Anschreiben 2).

5.1.3 Qualitative Inhaltsanalyse Freitextfeld

Die Personalverantwortlichen hatten nach Aufgabe 1 die Möglichkeit, in einem Freitextfeld eine Rückmeldung zu ihren individuellen Entscheidungskriterien zu geben (Abbildung 15). Von den insgesamt 166 Teilnehmenden haben 74 Teilnehmende (44,58 %) die optionale Frage nach ihren Entscheidungskriterien (d.h. woran Sie sich orientiert haben) beantwortet. Da die Rückmeldung als Text vorliegt, war eine qualitative Inhaltsanalyse notwendig, um die relevanten Kriterien zu identifizieren.

5.1.3.1 Methode

Die systematische Analyse der Textdaten wurde mittels qualitativer Inhaltsanalyse nach Mayring (2010) durchgeführt. Mayring (2010) unterscheidet drei Vorgehensweisen bei der qualitativen Inhaltsanalyse: zusammenfassend, explizierend und strukturierend. Für die vorliegende Auswertung wurde die zusammenfassende Inhaltsanalyse gewählt, um einen ersten Überblick zu bekommen, woran sich die Personalverantwortlichen bei der Identifikation von KI-generierten Texten orientiert haben.

Gemäß Mayring (2010, S. 65) besteht das Ziel einer zusammenfassenden Inhaltsanalyse darin, „das Material so zu reduzieren, dass die wesentlichen Inhalte erhalten bleiben, durch Abstraktion einen überschaubaren Corpus zu schaffen, der immer noch Abbild des Grundmaterials ist“. Ein zentraler Bestandteil dieses Verfahrens ist die Bildung von Kategorien, die entweder deduktiv (theoriegeleitet) oder induktiv (aus dem Material heraus) erfolgen kann. Da für die vorliegende Fragestellung bislang keine theoretischen Konzepte existieren, wurde die Kategorienbildung induktiv (Mayring, 2010, S. 83-85), während der Analyse des Materials, durchgeführt.

Die zusammenfassende Inhaltsanalyse nutzt vier Schritte, um das Material aufzuarbeiten (Mayring, 2010, S. 68-70):

- 1) Paraphrasierung der inhaltstragenden Textstellen
- 2) Bestimmung des angestrebten Abstraktionsniveaus und Generalisierung auf dieses Niveau
- 3) Erste Reduktion (Selektion und Streichen bedeutungsgleicher Paraphrasen)
- 4) Zweite Reduktion (Bündelung, Konstruktion und Integration der Paraphrasen auf dem festgelegten Abstraktionsniveau)

Das daraus gewonnene vorläufige Kategoriensystem wird anschließend noch einmal auf Basis des Ausgangsmaterial zurücküberprüft.

5.1.3.2 Datenanalyse

Zunächst wurde das Material gesichtet und mit der Paraphrasierung der Antworten begonnen. Insbesondere Antworten in Fließtextform wurden auf grammatische Kurzformen reduziert. Unwesentliche Ausschmückungen wurden gestrichen. Beispiel: „Einfach aus dem Bauch heraus.“ wurde zu „Bauchgefühl“ paraphrasiert. Das Abstraktionsniveau wurde auf Grundlage der Forschungsfrage festgelegt. Diese befasst sich mit den Kriterien, die Personalverantwortliche als Hinweis auf einen KI-generierten Anschreibentext herangezogen haben. Da es noch kein bestehendes Kategoriensystem gab, war es zunächst notwendig, sich einen möglichst generellen Überblick zu verschaffen. Dafür sollten wenige grobe Kategorien erstellt werden, die ggf. auch für quantitative Analysen geeignet sind (Paraphrasierung, Tabelle 15, Spalte 2). Anschließend wurden Paraphrasen mit gleicher Bedeutung zusammengeführt (z.B. „Bauchgefühl“ und „Intuition“ wurden zu „Intuition“; „Satzbau“ und „Syntax“ wurden zu „Syntax“). Bedeutungsgleiche Paraphrasen wurden gestrichen (1. Reduktion, Tabelle 15, Spalte 3).

<u>Ausgangstext</u>	<u>Paraphrasierung</u>	<u>1. Reduktion</u>
„meine Intuition - es ist schwierig gewesen das an konkreten Anhaltspunkten festzustellen“	• Intuition	• Intuition
„exakte Wiederholungen der Anforderungen und mein Bauchgefühl“	• Wiederholung der Anforderungen • Bauchgefühl	• Wiederholung aus der Stellenanzeige (sprachlich) • Intuition
„Ich habe drauf geachtet, ob die jeweiligen Tätigkeiten in der selben Reihenfolge auch in dem Anschreiben	• Übernahme der Reihenfolge der Tätigkeiten aus der Stellenanzeige	• Wiederholung aus der Stellenanzeige (inhaltlich und sprachlich)

<u>Ausgangstext</u>	<u>Paraphrasierung</u>	<u>1. Reduktion</u>
aufgetaucht sind und so gewirkt haben, als ob diese kopiert sind und dinge in 'Klammern' auch übernommen wurden sind.“		
„Fachbegriffe und persönlichen Bezug“	• Fachbegriffe • persönlicher Bezug	• Fachbegriffe • persönlicher Bezug

Tabelle 15. Auszug aus dem Freitextfeld zur qualitativen Inhaltsanalyse.

Am Ende der ersten Reduktion, lagen noch zu viele ($n = 40$) Kategorien vor. Daher war eine zweite Reduktion nötig, durch die die bereits reduzierten Paraphrasen mit ähnlichem Inhalt in neuen Formulierungen gebündelt wurden. Diese waren nun noch näher an dem festgelegten Abstraktionsniveau. Die Begriffe „Schlüsselworte“ und „Fachbegriffe“ wurden bspw. unter dem Begriff „Wortschatz“ zusammengefasst.

Nach der zweiten Reduktion war erkennbar, dass sich die Antworten grob in fünf Kategorien (Sprache, Inhalt, Sonstiges, Formulierungen, Persönlichkeit) einordnen ließen. Die Kategorie „Sprache“ umfasst dabei Begriffe, die die Art und Weise beschreiben, wie Inhalte vermittelt werden. Dazu gehören vor allem Rechtschreibung und Grammatik, aber auch Wortschatz oder Schreibstil. Die Kategorie Inhalt basiert auf der Definition laut Duden online (Inhalt, 2a, o.D.), der Inhalt als „etwas, was in etwas ausgedrückt, dargestellt wird“ definiert. Das Augenmerk liegt dabei nicht auf der Art und Weise, wie etwas gesagt wird, sondern darauf was gesagt wird. In den Texten zeigte sich, dass viele Teilnehmende „Formulierungen“ als Antwort angegeben haben. Daher wurde zunächst eine eigene Kategorie namens „Formulierungen“ erstellt, in der Antworten wie „allgemeine“ oder „umgangssprachliche Formulierungen“ gesammelt wurden. Zudem wurde die Kategorie „Persönlichkeit“ erstellt, da zunächst davon ausgegangen wurde, dass bei einem direkten Vergleich zwischen Menschen und künstlicher Intelligenz der Faktor der Persönlichkeit eine zentrale Rolle spielt. Die Kategorie „Sonstiges“ bündelte Begriffe, die keiner der vorherigen Kategorien zugeordnet werden konnten (z.B. „eigene Erfahrungen mit ChatGPT“).

Überarbeitung der Reduktion 1	Anzahl	Überarbeitung der Reduktion 2	Anzahl
Doppelungen & Wiederholungen	2	Inhalt	21
Logik & Logikfehler	2		
Oberflächlichkeit	3		
Wiederholungen aus der Stellenanzeige/Reihenfolge (inhaltlich)	4		
persönliche Infos	2		
persönliche Komponenten / Aspekte / Erfahrungen	4		
Erwähnung von KI-Erfahrung	1		
keine Tätigkeitsbeschreibung	1		
Begründung der eigenen relevanten Eigenschaften	1		
konkrete Beispiele	1		
Struktur/Einleitung	5	Sonstiges	23
Struktur/Ende	3		
Beschreibung (Projekte/Tätigkeiten)	2		
eigene Erfahrungen mit ChatGPT	2		
Intuition	4		
multiple choice wahrscheinlichkeit	1		
Vorlage angepasst	1		
(Un-)Persönlichkeit	2		
Fehler	1		
Passagen zu persönlichen Eigenschaften	1		
ungewöhnliche Stichpunkte	1		
Satz- und Textlänge	4	Sprache	85
Formulierungen	23		
Übersetzungsfehler	1		
Asteriks	1		
Wortschatz	15		
sprachliche Fehler	1		
Konjunktiv	1		
Superlative	1		
Syntax	9		
(häufige) Ich-Formulierungen	3		

Redewendungen	1		
Schreibstil	3		
Fließtext	1		
Wiederholungen aus der Stellenanzeige (sprachlich)	9		
Umgangsprache	1		
Grammatik	2		
Rechtschreibung	1		
Kreativität (des Schreibens)	1		
ausführliche Beschreibung (des Werdegangs)	1		
Sprache	1		
Schreibweisen	2		
Struktur	3		
SUMME			129

Tabelle 16 Codeübersicht der 1. und 2. Reduktion.

Anschließend wurde das beschriebene Kategoriensystem anhand des Materials rücküberprüft. Es zeigte sich, dass die Kategorien nicht dem angestrebten Abstraktionsniveau entsprachen, da insbesondere die Kategorien „Formulierungen“ und „Persönlichkeit“ zu spezifisch waren und diesen vergleichsweise wenig Antworten zugeordnet werden konnten. Daher wurden in einer zweiten Runde beide Reduktionsschritte überprüft. Dazu wurde zunächst die Reduktion 1 anhand des Ausgangsmaterials erneut verifiziert. Einige der Begrifflichkeiten wurden geschärft. Anschließend wurde die Reduktion 2 anhand der geänderten Begrifflichkeiten des Reduktionsschrittes 1 verifiziert und diese auf die absolut notwendigen reduziert. Es verblieben folgenden drei Kategorien: Sprache, Inhalt, Sonstiges (siehe Tabelle 16, Spalte 3).

Um die Ergebnisse nicht durch Mehrfachzuordnungen zu verzerren oder die Anzahl der Kategorien unnötig zu erhöhen, wurden der Kategorie „Sonstiges“ nun auch solche Paraphrasen zugeordnet, die aufgrund von Mehrdeutigkeit zwei oder mehreren Kategorien zugeordnet werden konnten: z.B. „Beschreibung der Tätigkeiten“ wurde dieser Kategorie zugeordnet, da deren Interpretation entweder auf die sprachliche Gestaltung („Wie wurde die Tätigkeit beschrieben?“) oder den Inhalt („Es hat eine Beschreibung der Tätigkeit stattgefunden.“) abzielen könnte.

Tabelle 17 Häufigkeiten unterteilt nach Kategorien

Kategorie	Sprache	Inhalt	Sonstiges
Anzahl der Antworten	85 (65,9%)	21 (16,3%)	23 (17,8%)

Auf Basis der Ergebnisse (**Fehler! Verweisquelle konnte nicht gefunden werden.**) kann festgehalten werden, dass mehr als die Hälfte der inhaltlichen Kommentierungen der qualitativen Feedbacks (65,9%) Sprache als Indikator für bzw. gegen den Einsatz von KI nutzten.

5.1.3.3 Limitation qualitative Auswertung

Die größte Schwierigkeit bei der Analyse der Ausgangstexte stellten kurze Antworten da, die sich mehreren Kategorien zuordnen ließen. Als Beispiel lässt sich hier die Antwort „Einleitung, Ende, Schlüsselworte“ heranziehen. „Schlüsselworte“ lässt sich problemlos der Kategorie

„Sprache“ zuordnen. Fraglich ist jedoch, was die Versuchsperson mit „Einleitung“ und „Ende“ meinte. Zum einen könnte es der Kategorie „Sprache“ zugeordnet werden, wenn es darum geht, wie diese Textstellen formuliert sind. Allerdings ist auch eine Kategorisierung bei „Inhalt“ denkbar, wenn es z.B. darum geht, welche Informationen enthalten sind oder ob es überhaupt einen Einleitungs- und Schlussteil gibt. In zukünftigen Untersuchungen könnten alternative Forschungsmethoden (z.B. Interview, Think Aloud) eingesetzt werden, um die Personen direkt befragen zu können.

Die zweite Schwierigkeit bei der Auswertung lag darin, dass nur eine Minderheit der Versuchspersonen angegeben hat, ob die berichteten Kriterien als Hinweis auf KI-generierte oder von Menschen verfasste Texte gedeutet wurden. Zukünftig sollte dies bei der Datenerhebung mit abgefragt werden.

6 Diskussion

Der zunehmende Bekanntheitsgrad von generativer KI (z.B. ChatGPT, Kero et al., 2023) und die zunehmende Nutzung von KI aufseiten der Bewerbenden (softgarden e-recruiting GmbH, 2023) wirft die Frage auf, ob sich die Verwendung von KI nachteilig auf den Bewerbungsprozess auswirken könnte. Ausgehend von der Signaltheorie wurde im Experiment geprüft, ob Personalverantwortliche erkennen können, dass generative KI zur Texterstellung genutzt wurde. Zudem wurde untersucht, wie sich der Hinweis, dass Anschreiben (angeblich) mithilfe von KI verfasst wurden, auf den Bewerbungserfolg (Eignung, Intention, Mühe) der Jobsuchenden auswirkt. Zudem wurden die Personalverantwortlichen zu ihrer Einstellung (Akzeptanz, Augenscheinvalidität) hinsichtlich der Nutzung von generativer KI aufseiten der Bewerbenden in der Personalvorauswahl befragt.

Als erstes Fazit lässt sich aus dem Experiment festhalten, dass die befragten Personalverantwortlichen nicht zuverlässig erkennen können, ob ein Anschreiben mit Hilfe von KI verfasst wurde. Als zweites Fazit lässt sich aus der Befragung festhalten, dass die Mehrheit der befragten Personalverantwortlichen dem Einsatz von KI zur Erstellung von Anschreiben offen gegenübersteht. Im Experiment zeigte sich aber auch, dass die Personaler den Bewerbenden damit weniger Mühe zusprachen, wenn beim Verfassen der Anschreiben vermeintlich KI genutzt worden war. Ob die Personaler die geringere Mühe in den Anschreiben als positiv (z.B. für ausgleichendes technisches Engagement) oder negativ (geringeren Einsatz

für den zukünftigen Job) werten wurde nicht ermittelt. Als drittes Fazit lässt sich aus der Befragung festhalten, dass die Personalverantwortlichen traditionelle Anschreiben für besser geeignet halten, um potenzielle KandidatInnen auszuwählen. Im Experiment allerdings wurden Bewerbende in der Vorauswahl nicht systematisch benachteiligt, wenn sie angeblich KI zur Erstellung der Anschreiben verwendet hatten. Dieser anscheinende Widerspruch kann aber auch daran liegen, dass beide Anschreiben im Experiment tatsächlich ohne KI geschrieben worden waren und somit objektiv keine auf KI-Nutzung zurückführbaren Unterschiede vorhanden waren. Denn für den Untersuchungszweck war lediglich der Anschein erweckt worden, dass eins von zwei Anschreiben mit Hilfe von KI geschrieben worden war. Dieser Zusammenhang, kann in zukünftigen Experimenten (z.B. mit Anschreiben, die tatsächlich mit KI generiert wurden) genauer untersucht werden. In der Inhaltsanalyse zeigte sich, dass sich die Mehrheit der befragten Personalverantwortlichen an der Kategorie Sprache orientierten, wenn Sie ein Urteil über die tatsächliche Urheberschaft (Mensch vs. KI) eines Anschreibens trafen.

Auch wenn die Stichprobe im Experiment relativ klein gegenüber der Grundgesamtheit der Beschäftigten im Personalmanagement und Recruiting ist, kann dennoch folgendes vorsichtig verallgemeinernd festgehalten werden:

- Da die Personalverantwortlichen nicht zuverlässig erkennen können, ob ein Anschreiben mithilfe von KI verfasst wurde und das vermeintliche Wissen über eine Urheberschaft keine negativen Konsequenzen für eine Intervieweinladung hatte, könnte den Bewerbenden empfohlen werden, KI-Tools im Rahmen des Bewerbungsprozesses zu nutzen.
- Da die Personalverantwortlichen nicht zuverlässig erkennen können, ob ein Anschreiben mithilfe von KI verfasst wurde und Bewerbenden mit angeblich KI-generierten Anschreiben weniger Mühe zugesprochen wurde, könnte den Personalverantwortlichen empfohlen werden (a) ihre eigene KI-Kompetenz zu steigern (Upskilling), um auch den Eigenbetrag der Bewerbenden besser einschätzen zu können und (b) die Bewerbenden über die Stellenanzeige zu ermutigen, die Nutzung von KI - Tools im Bewerbungsprozess mit anzugeben (Transparenz).
- Da die Personalverantwortlichen einerseits KI-Tools offen gegenüberstehen und andererseits traditionelle Anschreiben für die Personalauswahl bevorzugen, könnte

den Personalverantwortlichen empfohlen werden, diesen Widerspruch und seine Bedeutung für die Personalarbeit workshopbasiert im Kollegium zu reflektieren.

7 Literaturverzeichnis

- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. Addison Wesley / Pearson Benelux.
- Buley, C., Hannah Poppe, Roedenbeck, M., & Stollhoff, R. (2024). *KI-Verfahren im Recruiting: Analyse von Persönlichkeit, Ernsthaftigkeit, Urhebererschaft und Konsistenz* (S. 56) [Zwischenbericht].
- Celiktutan, B., Klesse, A.-K., & Tuk, M. A. (2024). Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations. *International Journal of Research in Marketing*.
- Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S., & Smith, N. A. (2021). All that's human's not gold: Evaluating human evaluation of generated text. *arXiv preprint arXiv:2107.00061*.
- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling Theory: A Review and Assessment. *Journal of Management*, 37(1), Article 1. <https://doi.org/10.1177/0149206310388419>
- Dutz, R., Hubner, S., & Peus, C. (2022a). When agency "fits" regardless of gender: Perceptions of applicant fit when job and organization signal male stereotypes. *Personnel Psychology*, 75(2), 441–483. <https://doi.org/10.1111/peps.12470>
- Dutz, R., Hubner, S., & Peus, C. (2022b). When agency "fits" regardless of gender: Perceptions of applicant fit when job and organization signal male stereotypes. *Personnel Psychology*, 75(2), Article 2.
- Kero, S., Akyürek, S. Y., & Flaßhoff, F. G. (2023). Bekanntheit und Akzeptanz von ChatGPT in Deutschland. *Meinungsmonitor Künstliche Intelligenz, Factsheet Nr. 10*.

- Kersting, M. (2008). Zur Akzeptanz von Intelligenz-und Leistungstests. *Report Psychologie*, 33(9), 420–433.
- Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of Experimental Social Psychology*, 40(1), 91–98.
- Lauver, K. J., & Kristof-Brown, A. (2001a). Distinguishing between Employees' Perceptions of Person–Job and Person–Organization Fit. *Journal of Vocational Behavior*, 59(3), 454–470. <https://doi.org/10.1006/jvbe.2001.1807>
- Lauver, K. J., & Kristof-Brown, A. (2001b). Distinguishing between Employees' Perceptions of Person–Job and Person–Organization Fit. *Journal of Vocational Behavior*, 59(3), 454–470. <https://doi.org/10.1006/jvbe.2001.1807>
- Lederer, M., Müller-Jungnickel, A. M., & Pirkel, S. (2021). Künstliche Intelligenz in HR-Prozessen: Anwendungsfälle und Akzeptanzstudie für die Personaleinstellung. *Künstliche Intelligenz erfolgreich umsetzen: Praxisbeispiele für integrierte Intelligenz*, 41–53.
- Magni, F., Park, J., & Chao, M. M. (2024). Humans as Creativity Gatekeepers: Are We Biased Against AI Creativity? *Journal of Business and Psychology*, 39(3), 643–656. <https://doi.org/10.1007/s10869-023-09910-x>
- Mayring, P. (2010). *Qualitative Inhaltsanalyse* (11. Aufl.). Beltz Verlagsgruppe. <https://content-select.com/de/portal/media/view/519cc17d-6158-4e6c-9944-253d5dbbeaba>
- Reeve, C. L., & Schultz, L. (2004). Job-Seeker Reactions to Selection Process Information in Job Ads. *International Journal of Selection and Assessment*, 12(4), 343–355. <https://doi.org/10.1111/j.0965-075X.2004.00289.x>
- softgarden e-recruiting GmbH. (2023). *Künstliche Intelligenz in der Bewerbung*.
- Spence, M. (1978). Job market signaling. In *Uncertainty in economics* (S. 281–306). Elsevier.

Weiss, D., Liu, S. X., Mieczkowski, H., & Hancock, J. T. (2022). Effects of using artificial intelligence on interpersonal perceptions of job applicants. *Cyberpsychology, Behavior, and Social Networking*, 25(3), 163–168.