

Wie verändert ChatGPT die Zukunft der Personalauswahl?

■ LUCIANA VON RÖMER, AMELIE KLEINMANN &
DR. ANNA-LENA JOBMANN

*Nach dem überaus erfolgreichen Start von ChatGPT in der breiten Öffentlichkeit ist viel über Nutzen und Risiken dieser Anwendung der künstlichen Intelligenz geschrieben und diskutiert worden. Da ChatGPT offensichtlich Schulabschlüsse und Zertifizierungen bewältigen kann, steht für uns die Frage im Raum, ob Bewerber*innen ChatGPT nutzen können, um Online-Eignungstests oder Online-Interviews (besser) zu bestehen. Dazu haben wir einen Selbstversuch unternommen und in diesem Beitrag dokumentiert. Welche Auswirkungen ergeben sich konkret für die Online-Personalauswahl und welche präventiven Maßnahmen für Arbeitgeber gibt es, um möglichen Mißbrauch zu verhindern.*

Seitdem im November 2022 ChatGPT öffentlich und kostenfrei zugänglich wurde, häufen sich Artikel, Berichte und Tutorials darüber, wie ChatGPT Schulabschlüsse und Zertifizierungen erlangen sowie Auswahlverfahren bestehen kann. Schlagzeilen wie: „ChatGPT KI besteht bayerisches Abitur mit Bravour“, „ChatGPT besteht IQ-Test“ oder „ChatGPT besteht den Test zum Wirtschaftsprüfer“ (Bültermann, 2023; Schiffer & Gawlik, 2023; Triesch, 2023) erwecken den Eindruck, dass Online-Prüfungen, Tests und Auswahlverfahren mit ChatGPT einfach zu manipulieren sind. Ist es wirklich so, dass ChatGPT scheinbar mühelos Prüfungsaufgaben besteht, die in vielen Kontexten angewendet werden? Können Bewerber*innen sich durch ChatGPT bei Interviews helfen lassen? Was bedeutet das konkret für die Personalauswahl – insbesondere für Online-Verfahren in der Personalauswahl?

Wir möchten mit unserem Artikel einen Beitrag dazu leisten, diese Fragen zu beantworten. Zunächst beschäftigen wir uns mit einigen Grundlagen von ChatGPT. Anschließend zeigen wir für Eignungstests sowie Assessment-Center, wie Bewerber*innen hier ChatGPT nutzen könnten und welche Folgen das mit sich bringt. Wir schließen mit Vorschlägen zu Maßnahmen der Prävention.

Was ist ChatGPT?

ChatGPT ist zurzeit einer der bekanntesten und leistungsfähigsten Chatbots. In der ersten Woche nach Veröffentlichung abonnierten über eine Millionen Menschen dieses neue KI-Werkzeug. Doch was ist ChatGPT? ChatGPT basiert auf einem Large Language Model (zu dt. großem Sprachmodell). Ein Large Language Model ist eine Art KI, die Deep Learning (eine Form des maschinellen Lernens) nutzt, um Text zu verarbeiten und

in möglichst natürlicher Sprache zu generieren. Diese Modelle werden mit einer unglaublichen Menge an Texten trainiert, damit sie in der Lage sind, die Komplexität und die Facetten menschlicher Sprache zu erlernen (Susnjak, 2022). Das genaue Trainingsmaterial für ChatGPT ist unbekannt. Bestandteile waren aber vor allem Bücher, Zeitungsartikel, Wikipedia-Seiten und Online-Unterhaltungen. Aus diesen Daten entsteht ein riesiges Wahrscheinlichkeitsmodell davon, wie Wörter von Menschen in der Regel in verschiedenen Kontexten benutzen werden (West, 2023). Es entsteht der Eindruck, dass Text, der durch ChatGPT generiert wird, sich von menschlich produziertem Text nicht unterscheidet. Eine Interaktion zwischen Nutzer*in und ChatGPT wirkt daher wie ein Gespräch zwischen zwei Menschen, da die gewählten Formulierungen natürlich und intuitiv wirken (Rudolph et al. 2023).

ChatGPT kann auf Basis der Trainingsdaten viele Fragen beantworten und Aufforderungen nachkommen. Dabei produziert ChatGPT sehr häufig, aber nicht zwangsläufig immer korrekte Antworten. Durch die geschickte Formulierung der Texteingaben bzw. Anweisungen, der sogenannten Prompts, kann ChatGPT für viele Aufgaben unterstützend benutzt werden und „trainiert“ werden, um Aufgaben zu lösen oder bei der Lösung zu unterstützen. Ein guter Prompt sollte klar und präzise formuliert sein und dem Modell ausreichend Informationen liefern, um eine gezielte Antwort zu erhalten. Die Qualität des Prompts beeinflusst somit maßgeblich die Qualität der Antwort des Chatbots. Zusätzlich zeichnet sich der Chatbot durch die Fähigkeit aus, Informationen aus vorherigen Interaktionen zu speichern und diese in seinen Antworten zu berücksichtigen.

Für die Zukunft der Arbeit und des Lernens ist die Nutzung von KI-Tools wie ChatGPT nicht mehr wegzudenken. Auf diese Entwicklung müssen nun viele Bereiche reagieren und sich Strategien überlegen, wie sie einerseits KI in Prozesse integrieren und sinnvoll nutzen und sich andererseits vor möglicherweise negativen Auswirkungen von KI schützen.



Wie heißt die erste Bundeskanzlerin?



Die erste Bundeskanzlerin Deutschlands war Angela Merkel. Sie bekleidete das Amt von 2005 bis 2021 und war die erste Frau, die dieses Amt in Deutschland innehatte. Bitte beachten Sie, dass sich diese Information auf meinen Wissensstand bis September 2021 bezieht, und es möglicherweise Entwicklungen oder Änderungen seitdem gegeben haben könnte.

Was ändert sich für die Personalauswahl?

Auf Grund der Möglichkeiten von ChatGPT zur Unterstützung stellt sich die Frage, ob und wie Bewerber*innen ChatGPT in Online-Auswahlverfahren nutzen könnten. Wenn Bewerber*innen ungewollt zur Lösung von Aufgaben in Online-Auswahlverfahren ChatGPT nutzen, dann gefährdet das die Validität des Auswahlverfahrens. Das bedeutet, die Schlüsse, die aus gezeigten Leistungen der Bewerber*innen gezogen werden können, sind nicht mehr eindeutig. Anders ausgedrückt: Eine gute Leistung im Auswahlverfahren ist nicht mehr unbedingt auf die Fähigkeiten und Eignung des/der Bewerber*in zurückzuführen, sondern auf die geschickte Nutzung von ChatGPT. Auch wenn es sicherlich für

viele Jobs hilfreich ist, ChatGPT sinnvoll nutzen zu können, so sollen Auswahlverfahren Aussagen über Bewerber*innen im Allgemeinen und deren Fähigkeiten und Eigenschaften machen können. Für die Berufsleistung ist in den allermeisten Fällen die Fähigkeiten zum logischen Denken oder die Sozialkompetenz der Bewerber*innen wichtiger als die Fähigkeiten im Umgang mit ChatGPT. Aufgrund dieser Überlegungen ist die Gefährdung der Validität von Auswahlverfahren durch ChatGPT ernst zu nehmen.

Deswegen wird im nächsten Abschnitt des Artikels im Selbstversuch untersucht, welche Aspekte bei der Planung und Durchführung von Online-Eignungstests sowie Online-Interviews zu berücksichtigen sind, insbesondere im Hinblick auf mögliche Täuschungsversuche von Bewerber*innen mit ChatGPT. Die Selbstversuche wurden mit der zu diesem Zeitpunkt aktuellen kostenlosen Version ChatGPT 3.5 durchgeführt. Abschließend schlagen wir angemessene Maßnahmen vor, um Täuschungsversuche zu reduzieren.

Teil I: Online-Eignungstests

Im US-amerikanischen Raum wurde bereits mehrfach getestet, wie ChatGPT bei standardisierten Tests abschneidet. In unterschiedlichen Wissensgebieten ist ChatGPT imstande, die Tests zu bestehen, wenn auch nicht alle Aufgaben zu lösen (Choi et al., 2023; Geerling et al., 2023; Gilson et al., 2023; West, 2023). Während ChatGPT in den Bereichen Medizin, Recht und Physik gerade vergleichbare Ergebnisse wie durchschnittliche Studierende erzielt (Choi et al., 2023; Gilson et al., 2023; West, 2023), so schneidet es besonders gut im Bereich Wirtschaft ab (Geerling et al., 2023). Choi (2023) konnte zeigen, dass ChatGPT bei Essayaufgaben im Bereich

Recht bessere Ergebnisse erzielte als bei Multiple-Choice-Aufgaben. Alle Studien hatten bei der Beurteilung von ChatGPT eine ähnliche Vorgehensweise. Die Aufgaben aus den jeweiligen standardisierten Tests wurden einzeln Schritt für Schritt ChatGPT zur Lösung vorgelegt. Die meisten Aufgaben bestanden aus einem Multiple-Choice-Format. Es wurden aber im Bereich Recht auch Essayaufgaben untersucht. Durch unterschiedliche Prompts wurde weiterhin versucht, die Treffergenauigkeit von ChatGPT zu erhöhen. Die unterschiedlichen Prompts hatten keinen signifikanten Einfluss auf die Leistung von ChatGPT.

Keine der vorliegenden Studien hat jedoch das Abschneiden von ChatGPT unter realen Prüfungsbedingungen inklusive Zeitbeschränkungen getestet. Es ist dadurch weiterhin offen, ob die Benutzung von ChatGPT während einer Prüfung zu einer deutlichen Leistungsverbesserung führt. Aus diesem Grund haben wir einen Selbstversuch unternommen und den modularen Online-Test für Verwaltungsberufe (MOT-V) der dgp mit Hilfe von ChatGPT unter realen Bedingungen absolviert. Beim MOT-V handelt es sich um einen Online-Test, der aus zwei Basismodulen besteht, nämlich einerseits dem Modul zu numerischer und andererseits dem Modul zu verbaler Verarbeitungskapazität. Zu den Basismodulen können weitere optionale Module wie z.B. Rechtschreibung, Allgemeinwissen oder Kenntnistests aus verschiedenen Wissensgebieten hinzugefügt werden. Weitere Informationen zu unseren Online-Eignungstests finden Sie auf unserer Homepage unter [dgp.de/test-konfigurator](https://www.dgp.de/test-konfigurator)

Reale Bedingungen bedeutet in diesem Fall: Mit einem Zugang zu dem Online-Test konnte der Test mit den Zeitbeschränkungen und

den üblichen Sicherheitsmechanismen bearbeitet werden. Dazu gehört eine Copy & Paste-Sperre, der Aufgabentext kann also nicht einfach herauskopiert werden. Trotz einer Copy & Paste-Sperre gibt es Möglichkeiten, Aufgaben (automatisiert) in ChatGPT einzuspeisen. Wir haben uns dafür entschieden, in einem ersten Durchlauf mit einem Mobiltelefon Items abzufotografieren, automatische Texterkennung zu nutzen und die Aufgaben in ChatGPT zu kopieren sowie in einem anderen Durchlauf, die Aufgaben in ChatGPT abzuschreiben. ChatGPT hat dann, ggf. mit Anpassung der Prompts, so viele Aufgaben pro Testbereich gelöst, wie zeitlich möglich waren.

Auf Basis von Zeitungsberichten haben wir damit gerechnet, dass unser Versuch zum Erfolg wird, und wir ein gutes Ergebnis beim Online-Test erreichen werden. Insgesamt war das Vorgehen in beiden Fällen jedoch eher aufwändig und wirkte nur bei wenigen Aufgabentypen flüssig. Tatsächlich haben unsere Test-Bewerber*innen mit Hilfe von ChatGPT nur Testwerte 79 (1. Durchlauf) und von 77 (2. Durchlauf) erreichen können und blieben damit deutlich unterdurchschnittlich auf einer Standardwertskala mit einem Mittelwert von 100 und einer Standardabweichung von 10. Dabei konnten sie die meisten Aufgabentypen so „prompten“, dass ChatGPT korrekte Antworten erzielen konnte: Die Eingaben der Aufgabenkategorien „Textrechnen“, „Schlüsse vergleichen“, „Grundrechnen“, „Zahlenreihen“ und „Wortklassifikationen“ führten zu korrekten Antworten bei den meisten eingegebenen Aufgaben. Dennoch erreichten sie mit Hilfe von ChatGPT nicht die möglichen Höchstwerte, da die Zeitbegrenzung meist vor Beenden der letzten Aufgabe abließ. Einige Aufgaben konnte ChatGPT nicht oder nur teilweise gut lösen (besonders

im Bereich der Textanalyse, bei Tabellen & Statistiken sowie bei Analogien). An zwei Aufgaben zeigen wir beispielhaft, zu welcher Lösung ChatGPT kommt und warum hier dennoch kein perfektes Ergebnis erzielt werden konnte.

→ Wortklassifikationen: In diesem Aufgabentyp müssen Bewerber*innen aus einer Wortfolge von 4 Wörtern jenes markieren, welches nicht zu den anderen passt. Die Aufgaben zeichnen sich dadurch aus, dass sie meist nicht schwierig sind, aber schnell gelöst werden müssen. Dabei scheiterten wir mit ChatGPT, denn wir konnten nur circa ¼ der Aufgaben in der zur Verfügung stehenden Zeit überhaupt in ChatGPT übertragen.



Welches Wort passt nicht zu den anderen?
greifen fassen tasten halten



Das Wort, das nicht zu den anderen passt, ist „halten“.

Die anderen Wörter („greifen“, „fassen“ und „tasten“) beziehen sich alle auf das Berühren oder Anfassen von Gegenständen oder Dingen mit den Händen. „Halten“ hingegen bezeichnet das Festhalten oder die Beibehaltung einer Position oder eines Zustands, ohne unbedingt ein aktives Berühren zu beinhalten.

→ Textanalyse: In diesem Aufgabentyp lesen Bewerber*innen Texte und müssen anschließend aus fünf Antwortoptionen diejenige identifizieren, die den Inhalt des Textes korrekt wiedergibt. Einige der Aufgaben kann ChatGPT gut lösen, bei anderen, z.B. mit juristischen Texten, kann

ChatGPT die korrekte Antwort nicht identifizieren. Zudem gibt es hier – trotz deutlich mehr Bearbeitungszeit – beim Abtippen der Texte das Problem, dass dieses Vorgehen deutlich zu zeitaufwändig ist.

Zwischenfazit: Worauf müssen Arbeitgeber bei Online-Tests achten, um Täuschungen durch ChatGPT zu verhindern?

Prinzipiell kann ChatGPT bei der Lösung vieler Aufgaben helfen. Einfache Schutzmechanismen, wie die Copy & Paste Sperre und eine Zeitbeschränkung verhindern jedoch, dass die Nutzung von ChatGPT zum Erfolg für die Testteilnahme wird. Bei einigen Aufgabentypen kann ChatGPT die korrekten Antworten nicht produzieren. Es ist jedoch zu vermuten, dass mit entsprechender Vorbereitung und Formulierung von Prompts, eine Verbesserung bei diesen Aufgaben erreicht werden könnte. Wir empfehlen folgende Schutzmaßnahmen für Online-Eignungstests:

→ Zeitbegrenzung: Eine sinnvolle Zeitbegrenzung, die nicht dazu führt, dass Bewerber*innen unter zu engen zeitlichen Begrenzungen den Test bearbeiten müssen, jedoch einen gewissen Zeitdruck haben, sollte genutzt werden. Es ist empfehlenswert, die Zeit nicht bekannt zu geben, damit kein aktives Zeitmanagement betrieben werden kann.

→ Copy & Paste-Sperre: Die Copy & Paste-Sperre ist effektiv. Text abzutippen oder anders in ChatGPT zu kopieren, ist zeitaufwändig und in Kombination mit einer Zeitbegrenzung ein nicht attraktives Vorgehen bei einem Täuschungsversuch mit ChatGPT.

→ Proctoring: Da es dennoch Aufgaben gibt, die sich leicht abtippen lassen, oder auch Wege, die Copy&Paste-Sperre zu um-

gehen (z.B. Abfotografieren & Texterkennung), ist eine Beaufsichtigung (Proctoring) der Online-Tests eine empfehlenswerte Ergänzung.

TEIL II: Online-Interviews

Technologievermittelte Interviews über Telefon, Videokonferenz oder als zeitversetzte asynchrone Video-Interviews sind keine „Alternative“, sondern für viele Organisationen bereits der Standard bei Vorstellungsgesprächen.

Bei der dgp werden vorwiegend Videokonferenz-Interviews durchgeführt, bei denen die Auswahlkommission und die Bewerber*innen in einer Videokonferenz zusammensitzen. Es entsteht eine ähnliche Gesprächssituation wie im Face-to-Face-Format, nur mit einem großen Unterschied: Die Kommission und die Bewerbenden sind physisch nicht im gleichen Raum.

Selbstversuch: Feldforschung Videokonferenz-Interview

Durch die o.g. Fähigkeit von ChatGPT, Informationen aus vorherigen Interaktionen zu speichern, können auch fiktive biographische Elemente authentisch nachgeahmt werden. In unserem Versuch haben wir den Chatbot mit relevanten Kontextinformationen zu einem bestimmten Verfahren sowie einem fiktiven Bewerber-Lebenslauf mit -Motivationsschreiben gefüttert. Dann haben wir den Prompt folgendermaßen formuliert:

„*Sie sind in einem Bewerbungsinterview für die Stelle XY bei Z, was würden Sie auf [Frage] antworten?*“

Bei unserem Selbstversuch nutzten wir das Format des teilstrukturierten Videokonferenz-Interviews. Hier werden den Bewerbenden Fragen gestellt, die im Vorfeld formuliert wurden. Grundsätzlich lassen sich fünf unterschiedliche Fragentypen unterscheiden:

1. **Biographische Fragen:** Werden gern zur Einleitung genutzt
2. **Standardfragen:** Zu grundlegenden Werten, Motivation und Haltung
3. **Fachfragen:** Hinterfragen fachbezogene Anforderungen an die Stelle
4. **Situativfragen:** Schildern Situationen aus dem Arbeitsalltag
5. **Abschlussfragen:** Runden das Interview am Ende ab

Die Teilstrukturierung hat den Vorteil, dass standardisierte Nachfragen von der Moderation gestellt werden können. Bei dem Selbstversuch wurde der Chatbot im Vorfeld mit Informationen aus einem Lebenslauf und Motivationsschreiben gefüttert. Die Fragen wurden dann, wie im obigen User-Prompt dem Chatbot gestellt.

Folgende Beobachtungen kamen bei unserem Selbstversuch heraus:

- In Interviews fällt es bei geübten Personen, die das Tippen beherrschen und gleichzeitig auf den Bildschirm schauen können, kaum auf, wenn sie während des Gesprächs unauffällig tippen.
- Bei der Beantwortung von **biographischen** Fragen war ChatGPT mit der richtigen Voreinstellung sehr hilfreich. Da der Chatbot Informationen in Form von Listen darstellt, waren diese besonders einfach zu benutzen. Der Bot konnte relevante Kompetenzen zusammenfassen und sogar konkrete – wenn auch erfundene – Beispiele nennen.



Wann und wodurch ist Ihr Interesse an der Stelle entstanden?



Mein Interesse an dieser Position wurde durch meine langjährige Erfahrung in Behörden und meinem tiefen Verständnis für die verschiedenen Aufgabenbereiche geweckt. Über die Jahre hinweg habe ich eine Expertise in administrativen Prozessen, der Anwendung von Gesetzen und Vorschriften sowie im Umgang mit komplexen Sachverhalten entwickelt. Diese Stelle bietet die Möglichkeit, meine Fähigkeiten und Erfahrungen optimal einzusetzen, um effiziente und qualitativ hochwertige Ergebnisse zu erzielen. Ich schätze die Verantwortlichkeiten, die mit dieser Position einhergehen, und sehe die Gelegenheit, einen nachhaltigen Beitrag zu leisten.

- Bei **Standardfragen** war es einfach, während des Anhörens der Frage wichtige Punkte oder Fragen in ChatGPT einzugeben und eine prägnante Stichpunktantwort zu erhalten. Selbst Personen, die langsam tippen, konnten wesentliche Aspekte und Schlüsselbegriffe problemlos in den Chatbot eingeben. Standardfragen sind in der Regel kurz und weisen, wie Studien zeigen, bereits ohne ChatGPT ein höheres Faking-/Impression-Management-Risiko auf (Strobel, A., Franke-Barthold, L. Püttner, I., Kersting, M. (2018)).
- Bei **Fachfragen** hängt die Genauigkeit der Antworten von der Art der Frage ab. Bei faktualen – d.h. auf Wissensabfrage ausgerichteten – Fragen konnte ChatGPT einige Themen gut erklären, wobei die Komplexität des Fachgebiets eine Rolle spielte. Es

ist wichtig zu beachten, dass das Wissen von ChatGPT zum Zeitpunkt des Selbstversuchs nur bis 2021 reicht und eher oberflächlich ist, zudem erfindet der Chatbot gelegentlich Fakten. Weitere Entwicklungen und Verbesserungen des Chatbots sind jedoch zu erwarten. Eine kostenpflichtige Version von ChatGPT hat bereits Zugang zum Internet und somit zu tagesaktuellen Datenbanken.

- Bei **Situativfragen**, die mehrere Informationen enthalten und konkrete Situationen beschreiben, gestaltet sich die Formulierung eines geeigneten Prompts etwas schwieriger. Zudem haben wir festgestellt, dass ChatGPT bei Fragen, die kritische Entscheidungen und ein gewisses Maß an Urteilsvermögen erfordern, keine große Hilfe ist. Die Antworten auf viele Situativfragen waren eher vage. Auch auf Nachfragen war es schwer, konkrete Antworten mit dem Chatbot zu generieren. Bewerber*innen, die mit dem Umgang des Chatbots vertraut sind, können jedoch relevante Teillinformationen aus den Antworten ziehen. Es bleibt zu beachten, dass ChatGPT die Fragen beantworten kann, aber mit Einschränkungen in Bezug auf spezifische Situationen und kritische Entscheidungen.
- Auch bei den **Abschlussfragen** zeigt sich, dass die Qualität der Antworten entscheidend von einem geeigneten Prompt abhängt. Auch in diesem Fall gelingt es ChatGPT, qualitativ hochwertige Antworten zu liefern.

Zwischenfazit: Worauf müssen Arbeitgeber bei Videokonferenzinterviews achten, um Täuschungen durch Chat GPT zu verhindern?

Im Vorfeld...

- Planen Sie teilstrukturierte Interviews, mit Möglichkeit der Nachfrage zu konkreten Beispielen oder zum Lebenslauf – besonders bei Standardfragen.
- Stellen Sie viele tätigkeitsrelevante Situativfragen. Auch fachliche Fragen können in eine Situation umgemünzt werden und somit zu fachlichen Situativfragen werden.
- Geben Sie bei der Einladung zum Videokonferenz-Interview technische Hinweise, welche Täuschungsversuchen unter Verwendung von ChatGPT vorbeugen. Bitten Sie die Bewerber*innen z.B., ihre Kamera so zu platzieren, dass man Gesicht, Oberkörper und Hände im Laufe des Interviews sieht.
- Vermeiden Sie kurze sehr pauschale Standardfragen sowie kurze sehr allgemeine Fachfragen.

Am Tag des Interviews...

- Geben Sie klare Verfahrenshinweise zum Anfang des Interviews: Bitten Sie z.B. alle Bewerbenden, nichts aufzuschreiben, auch nicht zu tippen und stellen Sie sicher, dass die Bewerbenden allein im Raum sind.
- Erläutern Sie, dass die Fairness des Verfahrens das oberste Prinzip ist.
- Führen Sie keine Interviews ohne funktionierende Kamera durch.

Es ist von großer Bedeutung, angemessene Kontrollmechanismen einzusetzen, ohne dabei den Eindruck zu erwecken, den Bewerbenden generell Misstrauen entgegenzubringen. Vielmehr sollten Sie im Hinterkopf behalten, dass eine wahrgenommene Fairness im Interview mit einer geringeren Anzahl von Täuschungsversuchen seitens der Bewerber*innen einhergeht (Bourdage et al., 2018).

Fazit

In unserem Selbstversuch zeigte sich, dass ChatGPT den Bewerbenden in vielen Bereichen neue Möglichkeiten für Täuschungsversuche bietet. ChatGPT kann bei schriftlichen Eignungstests Logik- und Kenntnisaufgaben lösen. Zusätzlich kann es im mündlichen Teil gute Antworten auf unterschiedliche Fragentypen generieren und auf den persönlichen Lebenslauf Bezug nehmen.

Aus unserer Sicht wäre die Herangehensweise, Fragen so zu formulieren, dass ChatGPT sie nicht lösen kann, ein verlorenes Rennen. Die Entwicklung von KI schreitet so schnell voran, dass eine rasche Anpassung zu erwarten ist. Zur Sicherung der Validität sowie der Fairness ist es aus unserer Sicht sinnvoller, sicherzustellen, dass die Technologie gar nicht erst zum Einsatz kommt.

In den Selbstversuchen zeigte sich, dass es einige Stellschrauben gibt, um den Einsatz von ChatGPT einzugrenzen. So können bei Online-Tests Copy & Paste-Sperren sowie Zeitbegrenzungen den Nutzen von KI-Tools deutlich verringern. Tätigkeitsrelevante Situativfragen sind in videobasierten Interviews gut geeignet, um den Einfluss von ChatGPT zu beschränken. Das wichtigste Instrument, um den Einsatz zu unterbinden, ist aber die Beaufsichtigung der Online-Test mit Proctoring.

Auch wenn wir dafür plädieren, die Nutzung von ChatGPT in Auswahlverfahren zu verhindern, steht für uns außer Frage, dass der angemessene Umgang mit KIs wie ChatGPT in unserer Gesellschaft zunehmend an Bedeutung gewinnt und eine wichtige Kompetenz darstellt. Durch die vorgestellten Stationen eines Onlineauswahlverfahrens sollen allerdings die Kompetenzen der Bewerber*innen gemessen werden. Über diese möchten wir mit den Auswahlverfahren Aussagen treffen können – dafür müssen wir die Nutzung von KI-Tools in Auswahlverfahren kontrollieren können.

Für die Zukunft mag es sinnvoll sein, die Kompetenzen im Umgang mit KI-Tools in einer separaten Aufgabe zu erfassen.

Literatur

- Basch, J.M., Melchers, K.G. (2023). Interviews per Telefon, Videokonferenz oder Videoaufzeichnung – inwiefern macht dies einen Unterschied?. In: Schulte, S., Hiltmann, M. (eds) Eignungsdiagnostische Interviews. Springer Gabler, Wiesbaden. https://doi.org/10.1007/978-3-658-38720-4_12
- Bourdage, J. S., Roulin, N., & Tarraf, R. (2018). „I (might be) just that good“: Honest and deceptive impression management in employment interviews. *Personnel Psychology*, 71(4), 597–632. <https://doi.org/10.1111/peps.12285>
- Bültermann, S. (2023, März 30). ChatGPT besteht IQ-Test. *Computer Bild*. <https://www.computerbild.de/artikel/cb-News-Internet-ChatGPT-besteht-IQ-Test-35523461.html>
- Choi, Jonathan H. and Hickman, Kristin E. and Monahan, Amy and Schwarzc, Daniel B., ChatGPT Goes to Law School (January 23, 2023). *Journal of Legal Education (Forthcoming)*. <http://dx.doi.org/10.2139/ssrn.4335905>
- Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. (2023). ChatGPT has Aced the Test of Understanding in College Economics: Now What? *The American Economist*, 0(0). <https://doi.org/10.1177/05694345231169654>
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med Educ*, 9, e45312. <https://doi.org/10.2196/45312>
- Strobel, A., Franke-Barthold, L., Püttner, I., Kersting, M. (2018). Eignungsinterviews/direkte mündliche Befragung (eds) Personalauswahl kompetent gestalten. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-53772-5_2
- Lukacik, E.-R., Bourdage, J. S., Roulin, N. (2022). Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews. *Human Resource Management Review*, 32(1), Article 100789. <https://doi.org/10.1016/j.hrmr.2020.100789>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching*, 6(1). <https://doi.org/10.37074/jalt.2023.6.1.9>
- Schiffer, C., & Gawlik, P. (2023, Mai 25). ChatGPT: So gut hat die KI das bayerische Abitur bestanden. *BR24*. <https://www.br.de/nachrichten/netzwelt/chatgpt-ki-besteht-bayerisches-abitur-mit-bravour,TFB3QBw>
- Susnjak, T. (2022). ChatGPT: The end of online exam integrity?. *arXiv preprint arXiv:2212.09292*.
- Triesch, A. (2023, Juni 5). ChatGPT besteht Test zum Wirtschaftsprüfer. *Rheinische Post*. https://rp-online.de/nrw/staedte/duisburg/uni-duisburg-essen-chatgpt-besteht-test-zum-wirtschaftspruefer_aid-91626531
- West, C. G. (2023). AI and the FCI: Can ChatGPT project an understanding of introductory physics?. *arXiv preprint arXiv:2303.01067*.

Kontakt

Luciana von Römer

M.Sc. Psychologie
M.Sc. Neuroscience
Deutsche Gesellschaft für
Personalwesen e.V.



Luciana von Römer ist Psychologin sowie Neurowissenschaftlerin und als Co-Leiterin der dgp-Geschäftsstelle Berlin/Leipzig tätig. Schwerpunktmäßig arbeitet sie im Bereich der Personalauswahl sowie der Personalentwicklung. Ihr Interesse gilt insbesondere der Transformation und dem Wandel von Systemen und Strukturen.

✉ roemer@dgp.de

Amelie Kleinmanns

M.Sc. Psychologie
Deutsche Gesellschaft für
Personalwesen e.V.



Amelie Kleinmanns ist Mitarbeiterin der Stabstelle Forschung und Entwicklung der dgp e.V. Ihre Schwerpunkte sind die Entwicklung und Überprüfung von kognitiven Eignungstests sowie die Entwicklung von Persönlichkeitsinventare auf Basis bewährter statistischer Testmodelle.

✉ kleinmanns@dgp.de

Dr. Anna-Lena Jobmann

Diplom Psychologin
Deutsche Gesellschaft für
Personalwesen e.V.



Dr. Anna-Lena Jobmann ist Mitarbeiterin der Stabstelle Forschung und Entwicklung der dgp e.V. Ihre Schwerpunkte sind die Entwicklung und Überprüfung von kognitiven Eignungstests auf Basis bewährter statistischer Testmodelle, inklusive adaptiven Testens, Messung sozialer Kompetenzen und Fähigkeiten sowie Fairness von Eignungstests.

✉ jobmann@dgp.de