

KI im Personalmanagement – ein Zwischenstand mit Fokus auf Mensch-KI-Kollaboration

Von **Prof. Dr. Torsten Biemann** (Universität Mannheim) und **Prof. Dr. Heiko Weckmüller** (Hochschule Koblenz)

Die rasante Entwicklung und zunehmende Verbreitung von Künstlicher Intelligenz (KI) prägt derzeit nahezu alle Aspekte unseres Lebens und revolutioniert die Art und Weise, wie Unternehmen agieren. KI-Anwendungen, insbesondere die jüngsten Fortschritte in der generativen KI wie Chat GPT, haben das Potenzial, etablierte Geschäftsprozesse grundlegend zu verändern und eine neue Ära in der Arbeitswelt einzuläuten. Diese Technologien sind nicht mehr nur isolierte Werkzeuge, sondern entwickeln sich zu universell einsetzbaren Technologien, die in der Lage sind, komplexe Aufgaben zu übernehmen und aus Erfahrungen zu lernen und sich anzupassen. Für Personaler kann dies eine fundamentale Neuausrichtung ihrer Rolle und ihrer Strategien bedeuten.

Das Thema ist sehr bedeutsam und die vorliegende Forschung wächst mit einer Geschwindigkeit, die wir von anderen HR-Themen so nicht kennen. Wir, als Autoren der State-of-the-Art-Reihe, haben in den letzten beiden Jahren immer wieder über die Behandlung dieses Themas nachgedacht, wollten dann aber immer noch ein Quartal auf relevante Ergebnisse warten. Inzwischen sind wir zu der Erkenntnis gelangt, dass es keinen idealen Zeitpunkt gibt und wir beim Thema KI immer über einen Zwischenstand berichten werden.¹ Vor diesem Hintergrund fokussieren wir uns auf die Frage, ob Mensch-KI-Zusammenarbeit zu besseren Ergebnissen führt. Dabei wird sich herausstellen, dass Kontextfaktoren eine große Rolle spielen, weshalb wir anschließend auf die Themen Vertrauen in KI und Präferenzen für KI eingehen.

Wie sollte die Zusammenarbeit zwischen Mensch und KI gestaltet werden?

Der Einsatz von KI kann die Produktivität erhöhen, das mag niemanden überraschen. Noy und Zhang (2023) beispielsweise zeigen, dass für eine Schreibaufgabe durch den Einsatz von Chat GPT die Bearbeitungszeit um 40 Prozent verringert und die Qualität um 18 Prozent gesteigert werden konnte. Diese Verbesserung war bei leistungsschwächeren Studienteilnehmern höher, wodurch also die Ungleichheit der Leistung zwischen den Teilnehmern durch den Einsatz von KI verringert werden konnte. Über dieses Einzelergebnis hinaus ist für die Personalarbeit relevant, wann die Zusammenarbeit zwischen Mensch und KI vorteilhaft ist.

Vaccaro und Kollegen (2024) untersuchen in ihrer Metaanalyse, ob und wann die Zusammenarbeit von Menschen und KI zu besseren Ergebnissen führt. Um Verbesserung zu bestimmen, ist die Bestimmung der Vergleichsgruppe wichtig. Die Autoren vergleichen die Ergebnisse, die aus der Zusammenarbeit zwischen Mensch und KI (Mensch-KI-Team) entstehen mit den isolierten Ergebnissen, wenn nur ein Mensch (Mensch allein) oder nur die KI (KI allein) agiert. Sie analysieren insgesamt 106 Studien, die in der Zeit zwischen Januar 2020 und Juni 2023 veröffentlicht wurden und in denen die Leistung von Menschen allein, KI allein und Mensch-KI-Teams in Experimenten untersucht wurde. Diese beeindruckende Zahl von über hundert Studien zu diesem Thema innerhalb von nur 3,5 Jahren spiegelt die gegenwärtige Bedeutung von KI-Forschung auch in den Sozialwissenschaften wider und bietet eine solide Grundlage für Analysen möglicher Synergieeffekte der Kollaboration von Mensch und KI. Die Forscher kommen zu dem überraschenden Ergebnis, dass Mensch-KI-Teams im Durchschnitt schlechter abschnitten als die jeweils bessere Einzelkomponente (Mensch allein oder KI allein). Die Effektstärke ist mit Hedges' $g = -0,23$ im leicht bis mittleren negativen Bereich, aber allein das negative Vorzeichen ist interessant. Das Ergebnis bedeutet, dass die meisten der untersuchten Mensch-KI-Kombinationen im Durchschnitt nicht in der Lage waren, die Stärken beider Parteien so zu kombinieren, dass ein optimal überlegenes Ergebnis erzielt wurde, das über die Leistung des jeweils besten Einzelakteurs hinausging. Im Durchschnitt konnte also keine Mensch-KI-Synergie nachgewiesen werden.

Allerdings wird mit der Analyse von Mensch-KI-Synergie ein strenger Maßstab angelegt, da die Kollaboration bessere Ergebnisse liefern muss gegenüber den zwei anderen Situationen – Mensch allein und KI allein. Vergleicht man das Mensch-KI-Team nur mit Mensch allein ergibt sich ein anderes Bild. Für diese Art der Leistungsmessung weisen die Autoren

¹ Dem Konzept der State-of-the-Art-Beiträge im PERSONALquarterly folgend, konzentrieren wir uns auf – metaanalytisch abgesicherte – empirische Evidenz. Für theoretische Ansätze und Rahmenmodelle empfehlen wir dem interessierten Leser Chowdhury et al. (2024), Hillebrand et al. (2025) und Karunakaran et al. (2025).

einen positiven Effekt nach ($g = 0,64$). Die Mensch-KI-Kollaboration führt also im Durchschnitt zu besseren Ergebnissen als der Mensch allein. Dieser Vergleich, bei dem die menschliche Leistung durch KI verbessert wird („human augmentation durch KI“) dürfte heute (noch) in vielen Arbeitssituationen der relevante sein, da eine vollständige Automatisierung aus ethischen, rechtlichen oder Sicherheitsgründen nicht möglich oder erwünscht ist.

Jenseits der durchschnittlichen Ergebnisse zeigen sich deutliche Differenzierungen (sog. Moderatoren) bezüglich der Leistungsfähigkeit von Mensch-KI-Teams. So zeigte sich in der Studie, dass die Art der Aufgabe und die relative Leistung von Mensch und KI entscheidend sind: Während Entscheidungsaufgaben oft Leistungsverluste zeigten, führten Aufgaben, die kreatives Schaffen erforderten, zu Leistungssteigerungen.

Weiterhin zeigte die relative Leistungsfähigkeit von Mensch und KI einen signifikanten Moderatoreffekt auf die Human-AI-Synergie: Wenn der Mensch allein die KI allein übertraf, übertraf das kombinierte Mensch-KI-Team im Durchschnitt beide Einzelakteure, mit einer durchschnittlichen Effektstärke für die Mensch-KI-Synergie von $g = 0,46$, was als mittelgroßer Effekt gilt. Dies deutet darauf hin, dass die Zusammenarbeit in solchen Fällen zu einer verbesserten Gesamtleistung führte, die über die des jeweils besten Einzelakteurs hinausging. Wenn die KI allein den Menschen allein übertraf, kam es im kombinierten System zu Leistungseinbußen im Vergleich zur KI allein. Die Effektstärke für die Human-AI-Synergie war hier negativ mit $g = -0,54$, ebenfalls ein mittelgroßer Effekt. Dies

bedeutet, dass die Kombination in diesen Fällen schlechter abschnitt als die KI allein, was darauf hindeutet, dass die menschliche Beteiligung die Leistung der überlegenen KI gemindert hat. Vaccaro und Kollegen liefern als potenzielle Erklärung für dieses Ergebnis, dass in über 95 Prozent der untersuchten Mensch-KI-Teams der Mensch die letztendliche Entscheidung traf, nachdem er den Input von KI-Algorithmen erhalten hatte. Wenn die Menschen insgesamt besser waren als die Algorithmen, waren sie auch besser darin zu entscheiden, wann sie ihren eigenen Meinungen vertrauen und wann sie sich stärker auf die Algorithmen verlassen sollten. Im Gegensatz dazu, wenn also die KI-Algorithmen genauer waren als die Menschen, waren die Menschen oft nicht gut darin, einzuschätzen, wann sie den Algorithmen vertrauen sollten und wann ihrem eigenen Urteilsvermögen, was zu einer geringeren Gesamtleistung des kombinierten Systems führte als die der KI allein. Insofern schließt sich hier die wichtige Frage an, unter welchen Bedingungen Menschen der KI bei ihren Entscheidungen vertrauen und wann nicht.

Welche Faktoren beeinflussen das Vertrauen der Menschen in KI?

In einem Übersichtsartikel diskutieren Glikson und Woolley (2020) empirische Evidenz zu menschlichem Vertrauen in KI. Die Studie unterscheidet zwischen kognitivem Vertrauen, das auf der Bewertung von Sachinformationen, Kompetenz und Nützlichkeit basiert, und emotionalem Vertrauen, das von Affekten oder Emotionen beeinflusst wird.

Tab. 1: Einflussfaktoren auf Vertrauen in KI

	k	Effektstärke (Cohen's d)	95 % CI (untere/obere Grenze)
Alter	2	0,09	-0,08/0,26
Einstellung gegenüber KI	5	1,05	-0,61/2,72
Leistung der KI	22	1,47	1,30/1,64
Aspekte des Interaktionskontextes	25	0,34	0,26/0,42
Risiko der Aufgabe	8	-0,41	-0,80/-0,01
Komplexität der Aufgabe	3	0,19	-0,16/0,54

k = Anzahl der Studien; CI = Konfidenzintervall

Quelle: Auswahl auf Basis von Kaplan et al. (2023)

Für das kognitive Vertrauen sind mehrere Faktoren entscheidend: Transparenz, insbesondere Erklärungen darüber, wie die KI funktioniert oder warum eine bestimmte Entscheidung getroffen wurde, kann das Vertrauen steigern. Zu diesem Aspekt der Erklärbarkeit von KI-Systemen (Explainable AI, „XAI“) haben Atf und Lewis (2025) eine Metaanalyse durchgeführt, die untersucht, ob die Erklärbarkeit von KI-Systemen das Nutzervertrauen signifikant erhöht. In den insgesamt 90 Studien ergibt sich ein schwacher bis mittlerer positiver Zusammenhang von $r = 0,20$ zwischen Erklärbarkeit und Vertrauen, was darauf hindeutet, dass Vertrauen zwar mit zunehmender Erklärbarkeit von KI steigt, dieser Anstieg jedoch nicht substanziell ist.

Eine geringe Zuverlässigkeit der KI verringert das Vertrauen, und die Wiederherstellung des Vertrauens ist schwierig und zeitaufwendig. Das Vertrauen ist in technische Aufgaben, die komplexe Berechnungen erfordern, höher als in sozialen Aufgaben. Darüber hinaus erhöhen unmittelbare Verhaltensweisen der KI, wie Reaktionsfähigkeit und Personalisierung, das Vertrauen, da sie als Zeichen hoher maschineller Intelligenz wahrgenommen werden. Die physische Präsenz („Tangibilität“) von robotergestützter KI kann ebenfalls das Vertrauen erhöhen (Glikson/Woolley, 2020).

Für das emotionale Vertrauen können menschenähnliche Eigenschaften (Anthropomorphismus) positive Emotionen und Vertrauen stärken, jedoch auch Unbehagen oder unrealistisch hohe Erwartungen hervorrufen. Hat ein künstliches Gegenüber menschenähnliche Züge, wirkt aber noch nicht lebensecht, betritt man das „uncanny valley“, welches zwischen positiv bewerteten „echt“ wirkenden und ebenfalls positiver bewerteten nicht menschenähnlichen Robotern liegt.

Generell deuten die Forschungsergebnisse von Glikson und Woolley (2020) darauf hin, dass hohes anfängliches Vertrauen oft sinkt, wenn die KI Fehler macht, und eine Neukalibrierung der Erwartungen entscheidend ist. Die Autoren betonen die Notwendigkeit eines menschenzentrierten Ansatzes bei der KI-Integration, der die Bedürfnisse und Anreize der Mitarbeiter berücksichtigt, um Vertrauen aufzubauen und die Produktivität zu steigern.

Während Glikson und Woolley einen Literaturüberblick liefern, in dem sie zwischen kognitivem und emotionalem Vertrauen unterscheiden, zielt die Metaanalyse von Kaplan und Kollegen (2023) stärker auf eine Quantifizierung der Effektstärken verschiedener Faktoren ab, die das Vertrauen in KI beeinflussen können. Sie unterscheiden in ihrer Analyse drei Hauptkategorien: Merkmale des menschlichen Nutzers, Eigenschaften der KI selbst und Aspekte des Interaktionskontextes. Für alle drei Bereiche können die Autoren Faktoren mit einem signifikanten Einfluss auf das Vertrauen in KI identifizieren, wobei besonders die Eigenschaften der KI auf das Vertrauen der Nutzer beeinflussen (vgl. Tabelle 1). Es mag wenig überraschen, dass gerade die Leistung der KI das Vertrauen in diese

stärkt ($d = 1,47$). Bemerkenswert, wenn auch in den Effekten geringer, sind Merkmale der Nutzer und des Kontextes. So finden die Autoren in den insgesamt 23 Studien, die Merkmale der menschlichen Nutzer untersuchen, einen mittelstarken Effekt von $d = 0,26$ und für Aspekte des Interaktionskontextes ($k = 25$) einen Effekt von $d = 0,31$. Für die Personalarbeit bedeuten diese Ergebnisse, dass bei der Einführung von KI natürlich auf eine leistungsstarke KI, aber eben auch auf Eigenschaften der Mitarbeiter und Kontext geachtet werden sollte, damit ein hohes Vertrauen in die KI bei den Mitarbeitern entsteht.

Wann präferieren Menschen KI und wann lehnen sie KI ab?

Eng verbunden mit der vorherigen Frage nach dem Vertrauen in KI ist die Frage der Nutzungspräferenz: Wenn man zwischen Menschen und KI entscheiden muss, wen bevorzugt man? Dieser Frage gehen Xin Qin et al. (2025) in ihrer Metaanalyse auf Basis von 163 Einzelstudien nach. Im Durchschnitt zeigt sich eine schwache Abneigung der KI (Cohen's $d = 0,27$). Interessant sind auch hier wieder die Kontextfaktoren. Die Autorengruppe schlägt ein Modell vor, in dem die Akzeptanz positiv von der wahrgenommenen Leistungsfähigkeit der KI und negativ von der Notwendigkeit der Personalisierung der Ergebnisse abhängt. Mit Personalisierung ist gemeint, inwieweit die Ergebnisse verallgemeinerbar sind beziehungsweise auf einzelne Personen individuell angepasst werden sollten. Die Empfehlung für einen Film weist beispielsweise einen hohen Personalisierungsgrad auf, da die Geschmäcker als individuell wahrgenommen werden. Die Ergebnisse können das Modell bestätigen: Aversion gegenüber KI tritt auf, wenn entweder Personalisierung oder eine geringe Leistungsfähigkeit der KI vorliegt. Auch diese Analysen zeigen wieder, dass bei der Einführung von KI die individuellen und situativen Kontextfaktoren berücksichtigt werden müssen.

Zusammenfassung und Fazit

Generative KI wie Chat GPT verändert Arbeitsprozesse grundlegend und entwickelt sich stetig weiter. Bisherige Forschung liefert aber schon interessante Erkenntnisse:

- ▶ Mensch-KI-Kollaboration bringt im Durchschnitt leicht schlechtere Ergebnisse als die beste Alternative von Mensch allein beziehungsweise KI allein.
- ▶ Wenn Menschen allein zu besseren Ergebnissen kommen als KI allein, dann führt die Mensch-KI-Kollaboration zu einer weiteren Verbesserung.
- ▶ Aufgabentyp und Rollenverteilung sind entscheidend. Kreative Aufgaben und eine gezielte Steuerung der KI durch den Menschen fördern Synergien, während Entscheidungsaufgaben oder falsche Gewichtung von Mensch- und KI-Beiträgen zu Leistungseinbußen führen können.
- ▶ Vertrauen in KI hängt von mehreren Faktoren ab. Leistung der KI, Transparenz/Erklärbarkeit, Zuverlässigkeit sowie

nutzer- und kontextbezogene Merkmale beeinflussen Vertrauen.

- Technische Kompetenz und nachvollziehbare Entscheidungen stärken kognitives Vertrauen, menschenähnliche Eigenschaften fördern emotionales Vertrauen, bergen aber auch Risiken wie das „uncanny valley“.



PROF. DR. TORSTEN BIEMANN

Lehrstuhl für Personalmanagement und Führung

Universität Mannheim

E-Mail: biemann@bwl.uni-mannheim.de

<http://biemann.bwl.uni-mannheim.de>



PROF. DR. HEIKO WECKMÜLLER

Professur Human Resources

Hochschule Koblenz

Rheinahrcampus Remagen

E-Mail: weckmueller@rheinahrcampus.de

www.hs-koblenz.de

LITERATURVERZEICHNIS

Atf, Z./Lewis, P. R. (2025): Is trust correlated with explainability in AI? A meta-analysis. *IEEE Transactions on Technology and Society*.

Chowdhury, S./Budhwar, P./Wood, G. (2024): Generative artificial intelligence in business: towards a strategic human resource management framework. *British Journal of Management*, 35(4), 1680-1691.

Glikson, E./Woolley, A. W. (2020): Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627-660.

Hillebrand, L./Raisch, S./Schad, J. (2025): Managing with artificial intelligence: An integrative framework. *Academy of Management Annals*, 19(1), 343-375.

Kaplan, A. D./Kessler, T. T./Brill, J. C./Hancock, P. A. (2023): Trust in artificial intelligence: Meta-analytic findings. *Human factors*, 65(2), 337-359.

Karunakaran, A./Lebovitz, S./Narayanan, D./Rahman, H. A. (2025): Artificial Intelligence at Work: An Integrative Perspective on the Impact of AI on Workplace Inequality. *Academy of Management Annals*, (ja), annals-2023.

Noy, S./Zhang, W. (2023): Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187-192.

Qin, X. et al. (2025): AI aversion or appreciation? A Capability–Personalization Framework and a Meta-Analytic Review. *Psychological Bulletin*, 151(5), 580-599.

Vaccaro, M./Almaatouq, A./Malone, T. (2024): When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8(12), 2293-2303.