DOI: 10.5771/0342-300X-2026-6-470

KI im Personalmanagement besser verstehen und anwenden

Die Rolle von KI-Kompetenzen und Erklärungen

KI-Anwendungen im Personalmanagement können ihre zugrunde liegenden Modelle und Ergebnisse erklären, damit Personaler*innen sie nachvollziehen können. Daneben sind jedoch grundlegende technische, kritische und praktische KI-Kompetenzen notwendig, um eine sichere KI-Nutzung zu fördern und Erklärungen zu verstehen. Die europäische KI-Verordnung verlangt zwar, dass Beschäftigtenkompetenzen geschult werden, konkretisiert die Umsetzung und Inhalte jedoch nicht. Auch wie sich höhere Kompetenzen bei der KI-Nutzung tatsächlich auswirken, ist kaum erforscht. Wir zeigen: KI-Kompetenzen machen KI subjektiv vertrauenswürdiger, aber ihre Interpretation nicht automatisch verlässlicher. Hierzu bedarf es zudem adressaten- und kompetenzgerechter Erklärungen.

YANNICK KALFF, KATHARINA SIMBECK

1 KI-Kompetenzen im Personalmanagement

KI-basierte Software spielt im Personalmanagement eine zunehmend wichtige Rolle. Schon bevor OpenAI sein Sprachmodell ChatGPT der breiten Öffentlichkeit zugänglich gemacht hatte, wurden im Personalmanagement KI-Funktionen und -Anwendungen erprobt und eingesetzt (Böhmer/Schinnenburg 2023). Mittels Techniken des maschinellen Lernens versprechen KI-Anwendungen, das Personalmanagement mithilfe analytischer oder prädiktiver Methoden zu optimieren und durch aktive Handlungsempfehlungen zu gestalten (Tambe et al. 2019). Software untersucht beispielsweise eingegangene Bewerbungsprofile, gleicht sie mit dem Such- und Kriterienkatalog einer offenen Stelle ab und bietet den Personaler*innen eine vorsortierte Liste der geeignetsten Bewerber*innen. Als KI-Systeme verstehen wir im Folgenden algorithmische Systeme, die auf Basis von Daten Muster erkennen, Vorhersagen treffen und kontinuierlich dazulernen können. Sie unterstützen oder automatisieren Entscheidungen mithilfe von Empfehlungen, Sortierungen oder Punktwerten. Dazu zählen maschinelles Lernen, regelbasierte Ansätze sowie generative Modelle, die Texte, Bilder oder andere Medien erzeugen. Konkret unterstützt KI-basierte Software beim Vorsortieren von Bewerbungen, beim Abgleich von Lebensläufen mit offenen Stellen, bei der Gestaltung zielgerichteter, individueller Qualifizierungspläne, beim Nachfolgemanagement sowie bei der Prognose des Personalbedarfs und der Fluktuation (Ncube et al. 2025).

Das Personalmanagement verändert sich dadurch von einem Arbeitsfeld, dem vor allem Menschenkenntnis, Bauchgefühl und Empathie zugeschrieben wurde, zu einem datengetriebenen Feld mit (vermeintlich) analytischer Objektivität (Loscher/Bader 2023). Digitalisierungsprozesse haben bereits seit den 1990er Jahren messbare und objektive Verfahren ermöglicht (Bondarouk/Brewster 2016). KI beschleunigt diese Veränderungen, denn große Mengen an Personaldaten und maschinelles Lernen versprechen datengestützte Evidenz statt subjektiver Entscheidungen (Tambe et al. 2019). Zugleich zeigt sich aber auch, dass KI fehleranfällig ist und dass verzerrte Trainingsdaten Ungleichheit fortschreiben: Aus einer Einstellungspolitik, die in der Vergangenheit vor allem Männer bevorzugt hat, kann eine KI erlernen, dass Männer bevorzugt einzustellen sind (Schellmann 2024; Tripathi 2025, S. 79f.). Menschen können solchen verzerrten KI-Empfehlungen folgen oder sie korrigieren – hierzu benötigen sie jedoch KI-Kompetenzen.

Über die betriebliche Nutzung von KI und die dafür benötigten Fähigkeiten gibt es bislang vor allem Rahmenkonzepte mit verschiedenen relevanten Dimensionen von technischen Kenntnissen bis hin zu reflexiven Fähigkeiten, aber kaum Untersuchungen zum Zusammenhang

© WSI Mitteilungen 2026
Diese Datei und ihr Inhalt sind urheberrechtlich geschützt. Nachdruck und Verwertung (gewerbliche Vervielfältigung, Aufnahme in elektronische Datenbanken, Veröffentlichung online oder offline) sind nicht gestattet.

zwischen KI-Kompetenzen und Verständnis, Vertrauen oder wahrgenommener Nützlichkeit (Benlian/Pinski 2025; Strauß 2024). KI-Kompetenzen (engl. AI literacy) sind jedoch, hierüber besteht Einigkeit, unabdingbar, um KI effizient und verlässlich einzusetzen (Cetindamar et al. 2024). Personaler*innen, die KI nutzen und ihre Ergebnisse verstehen und einordnen müssen, benötigen KI-Kompetenzen, die es ihnen erlauben, die Technologie in ihrem Fachbereich sicher zu handhaben (Bildat/Brouns 2026). Denn KI erzeugt Ergebnisse, die selbst für technische Expert*innen nicht immer genau erklärbar sind. Für das operative Geschäft im Personalmanagement, in dem Entscheidungen begründet und Prozesse nachvollziehbar sein müssen, stellt das ein Problem dar, dem mit KI-Kompetenzen, Transparenz und Erklärbarer KI (engl. *Explainable AI*; kurz: *XAI*), das heißt dem Hinzufügen von Erklärungen, wie KI funktioniert und wie ihre Ergebnisse zustande gekommen sind, begegnet werden soll.

Unser Beitrag untersucht, wie KI-Kompetenzen von Personalmanager*innen die Interpretation KI-basierter Ergebnisse beeinflussen, wobei zwei Fragen im Mittelpunkt stehen: 1. Wie beeinflussen KI-Kompetenzen sowohl die subjektive Wahrnehmung als auch das objektive Verstehen von KI-basierten Softwareprogrammen und ihren Nutzungsoberflächen? 2. Wie wirkt sich die erhöhte Transparenz durch Erklärungen (Erklärbare KI) auf die subjektive Wahrnehmung und das objektive Verstehen von KI-Tools aus?

Wir zeigen, dass KI-Kompetenzen sowie Erklärungen allein nicht ausreichen, um KI-Ergebnisse verlässlich zu interpretieren. Es bedarf einer kombinierten Betrachtung, denn Erklärungen wirken abhängig von den individuellen KI-Kompetenzen der Befragten. Das ist relevant, weil technische Lösungen zur Erklärung von KI nicht automatisch garantieren, dass Menschen die Ergebnisse korrekt verstehen. Um eine Erklärung zu verstehen, warum einem Beschäftigten eine konkrete Weiterbildungsmaßnahme vorgeschlagen wurde, wird grundlegendes Wissen darüber benötigt, wie die KI zu ihren Ergebnissen kommt sowie Fähigkeiten, dieses Wissen auf die KI-Empfehlung anzuwenden, und schließlich Kompetenzen, um die Information kritisch mit Blick auf technische Grenzen, ethische Aspekte und den Arbeitskontext einzuordnen. Sollen Beschäftigte oder die institutionelle Mitbestimmung KI verlässlich, fair und rechtssicher einsetzen, benötigen sie KI-Kompetenzen, um die Limitationen, ethischen Implikationen und technischen Funktionen der Systeme zu erkennen sowie ihre Erklärungen zu verstehen.

Intransparente KI im Personalmanagement ist problematisch. Wenn technische Grundlagen wie maschinelles Lernen, große Datensätze und Algorithmen sowie ihre Limitierungen, Unschärfen, Statistik oder Verzerrungen unsichtbar bleiben, sind informierte und verlässliche Entscheidungen riskant. Auch arbeitspolitisch ist das relevant, da es zu verzerrten oder wirtschaftlich fehlerhaften Entscheidungen kommen kann und KI die Entwicklung des Personalmanagements in Richtung Wertstromorientierung, Kennzahlensteuerung und evidenzbasiertes Management beeinflusst (Coron et al. 2024).

Wir diskutieren zunächst KI-Kompetenzen für nichttechnische Professionen und geben einen Überblick über das Forschungsfeld Erklärbare KI als Methode zur Steigerung der Transparenz von KI-Systemen (Abschnitt 2), bevor wir das Untersuchungsdesign vorstellen (3). Unsere Ergebnisse (4) zeigen, dass KI-Kompetenzen positiv mit der subjektiven Wahrnehmung der Oberflächen zusammenhängen. Allerdings finden wir auch einen negativen Zusammenhang zwischen KI-Kompetenz und dem objektiven Verstehen der Oberflächen. Wir diskutieren abschließend, dass KI-Kompetenzen wichtig sind, jedoch keinen sicheren Umgang mit KI gewährleisten. Erklärbare KI allein löst das Problem ebenfalls nicht vollständig. Vielmehr wirken Erklärungen abhängig vom KI-Kompetenzniveau (5).

2 KI-Kompetenz und Erklärbare KI

2.1 KI-Kompetenz

KI-Kompetenzen werden eine große Bedeutung zugeschrieben, wenn es um die Fähigkeit der Beschäftigten geht, KI-Technologien verlässlich und selbstwirksam einzusetzen.

Rechtlich schreibt die europäische KI-Verordnung KI-Kompetenzen für Nutzer*innen und für Betroffene von KI-Systemen zwar vor, fasst sie jedoch äußerst knapp (KI VO, Art. 4). KI-Systeme im Personalmanagement werden als hohes Risiko klassifiziert, was nicht nur besondere Vorsicht, sondern auch zusätzlichen Kompetenzbedarf bedeutet (KI VO, Art. 6 (2), Anh. III (4)). Dennoch bleibt offen, was konkrete KI-Kompetenzen ausmachen (KI VO, Art. 3 (56)) und wie sie vermittelt werden sollen. Juristische Kommentare der KI-Verordnung haben Kompetenzen ausgelegt, beispielsweise als interne Richtlinien, die durch KI-Beauftragte, Schulungen technischer, sozialer, ethischer und rechtlicher Aspekte, Zertifizierungen oder durch praxisorientiertes Lernen vermittelt werden (Martini/Wendehorst 2024, S. 237ff.). Ob und wie aus der KI-Verordnung unmittelbare Qualifizierungsansprüche abgeleitet werden können, wird eher zurückhaltend bewertet (Donath 2025).

In der wissenschaftlichen Debatte umfasst das Konzept „KI-Kompetenz" die Fähigkeit, künstliche Intelligenz zu verstehen, zu erkennen, anzuwenden, kritisch zu bewerten und ethisch zu reflektieren. Dies geht über klassische digitale Kompetenzen hinaus und ist ein vielschichtiges Konstrukt aus technischem Wissen, anwendungsorientierten Fähigkeiten, reflexiven und psychologischen Kompetenzen und ethischer Sensibilität (Carolus et al.

2023). Das technische Wissen umfasst Grundlagenkenntnisse in den Bereichen Algorithmen, maschinelles Lernen und Datenverarbeitung (Almatrafi et al. 2024), während die Anwendungskompetenz die angemessene Nutzung von KI-Tools zur Problemlösung in spezifischen Kontexten umfasst (Ng et al. 2021). Kritisch-reflexive Elemente beziehen sich auf die Fähigkeit, KI-Systeme und deren Ergebnisse hinsichtlich ihrer Validität, ihrer Verzerrungen und ihrer Zuverlässigkeit zu bewerten (Laupichler et al. 2023). Ethische Reflexionsfähigkeit ermöglicht Anwender*innen, Fairnessprobleme, Datenschutzrisiken und die gesellschaftliche Tragweite zu erkennen, während psychologische Kompetenzen Selbstwirksamkeit, Emotionsregulation und adaptive Problemlösung im KI-Einsatz einschließen (Carolus et al. 2023).

Unsere Untersuchung operationalisiert KI-Kompetenzen nach dem SNAIL-Modell (Laupichler et al. 2023). Es wurde für die Untersuchung von Expert*innen aus nicht-technischen Bereichen entwickelt, die KI-Systeme im Arbeitsalltag nutzen. Das SNAIL-Modell zielt auf die Selbsteinschätzung der Befragten ab und erfasst *erstens* das technische Verständnis dafür, wie KI funktioniert und welche Rolle Daten, Training und maschinelles Lernen spielen. *Zweitens* wird das kritische Vermögen der Befragten erhoben, die Grenzen von KI zu erkennen, den KI-Einsatz auch ethisch abzuwägen und unangemessene KI-Nutzung zu reflektieren. *Drittens* misst es die praktischen Fertigkeiten der Anwender*innen, beispielsweise ob sie KI als solche erkennen oder ob sie den angemessenen KI-Einsatz für bestimmte Aufgaben beurteilen können (Laupichler et al. 2023, S. 6).

Welche Art von KI-Kompetenzen für eine reflexive Nutzung von KI-Technologien in Hochrisiko-Bereichen wie dem Personalmanagement erforderlich sind und welche Rolle sie spielen, ist bislang noch kaum untersucht worden (Benlian/Pinski 2025).

2.2 Erklärbare KI

Beim Thema KI wird häufig das Bild der „Blackbox" (Kalff/Simbeck, i. E.) herangezogen, um zu verdeutlichen, dass KI für nicht-technische Anwender*innen schwer nachvollziehbar ist, denn neben einigen Daten und Ergebnissen bleibt die Transformation durch das Modell unerklärbar. KI-Kompetenzen können bis zu einem gewissen Grad dazu beitragen, dass KI von Nutzer*innen verstanden wird. Allerdings können KI-Systeme auch transparenter gestaltet werden, indem ihre Funktionen und Ergebnisse den Anwender*innen über technische Schnittstellen erklärt werden. Dies ist Gegenstand des Forschungsfeldes „Erklärbare KI" *(Explainable AI, XAI)* (Guidotti et al. 2019; Molnar 2025).

Erklärbare KI zielte zunächst auf Entwicklungsprozesse, denn Entwickler*innen benötigten mathematische Annäherungen an die tatsächliche Funktionsweise ihrer KI-Modelle, um diese zu verstehen und zu korrigieren (Langer et al. 2021). Anwender*innen können auch von Erklärungen profitieren, jedoch unter anderen Voraussetzungen. So zeigt die sozialpsychologische Forschung, dass das traditionelle Verständnis von Erklärung – nämlich genaues Wissen über die Abläufe der KI – bei Anwender*innen an Grenzen stößt. Wissen allein reiche nicht aus, um ein tiefgreifendes Verständnis zu gewährleisten (Ehsan et al. 2024). Stattdessen solle eine Erklärung in sozialen und fachlichen Kontexten gedacht sowie der mentale Zustand der Rezipient*innen berücksichtigt werden (Miller 2019). Die Annäherung an die Funktion sei daher wichtiger als das genaue Verständnis undurchsichtiger Modelle (Páez 2019, S. 456). Kontrastive und selektive Erklärungen seien hierfür besser geeignet: etwa zielgruppenspezifische, im Fachkontext der Rezipient*innen aufbereitete Erklärungen anstelle abstrakter technischer Erklärungen, die Laien in falscher Sicherheit wiegen könnten (Ehsan et al. 2024).

In der Praxis sehen Anwender*innen Erklärungen auf Softwareoberflächen, auf denen beispielsweise mit Bewerbungsunterlagen interagiert wird. Unterschiedliche Erklärungstypen zielen auf verschiedene Ebenen ab (Guidotti et al. 2019): Das Modell kann insgesamt erklärt werden, beispielsweise welche Daten es einbezieht und wie diese Daten gewichtet und verwendet werden. Ist ein Modell hierfür zu komplex, also eine Blackbox, lassen sich Einzelentscheidungen nachträglich erklären, beispielsweise durch relevante Einflussfaktoren auf das Ergebnis *(feature importance)*. In einem Fall könnte eine hohe Gehaltsvorstellung für die Bewertung einer Bewerbung negativ, aber eine lange Berufserfahrung besonders positiv in den berechneten Punktwert einfließen. Mittels kontrafaktischer Erklärungen *(counterfactuals)* lassen sich die Grenzen der Entscheidungsbereiche des Modells verdeutlichen (Cambria et al. 2023; Guidotti et al. 2019). Eine Erklärung zeigt, wie sich einzelne Variablen verändern müssten, damit sich eine Entscheidung ändert.

Erklärbare KI ist für Anwendungen im Personalmanagement hochrelevant, um die Ergebnisse KI-basierter Software zu kontextualisieren und sie damit für nicht-technische Expert*innen nachvollziehbar zu machen. Insbesondere in Hochrisikobereichen wie dem Personalmanagement ermöglichen Erklärungen auf Modell- oder Ergebnisebene darüber hinaus, die Güte und Korrektheit von KI-Vorhersagen einzuschätzen und notfalls zu korrigieren.

2.3 KI-Kompetenz für Erklärbare KI

Erklärungen sind nur wirksam, wenn Nutzer*innen sie auch verstehen und einordnen können. Mit diesem Beitrag untersuchen wir daher den bislang kaum erforschten Einfluss von KI-Kompetenz auf die Wirksamkeit Erklärbarer KI im Personalmanagement (Kalff/Simbeck 2025).

Einige wenige Studien in anderen Bereichen zeigen, dass alltagssprachliche Erklärungen und interaktive For-

mate das Verständnis und das Vertrauen von Laien verbessern können (Weitz et al. 2021). Aber ohne KI-Kompetenz verfehlen selbst technisch ausgereifte Erklärungen ihre Wirksamkeit (Cambria et al. 2023; Kühl et al. 2023). Ohne ausreichende KI-Kompetenzen riskieren Nutzer*innen Fehlinterpretationen oder setzen ungerechtfertigtes Vertrauen in KI-Systeme (Bauer et al. 2023, S. 202). Gleichzeitig können Erklärungen KI-Kompetenz aktiv fördern, indem sie technisches Wissen durch die Visualisierung von Modellentscheidungen vermitteln oder kritische Reflexion anregen (Long et al. 2023).

3 Untersuchungsdesign

Um das Zusammenwirken von KI-Kompetenzen und Erklärbarer KI zu analysieren, haben wir ein Experiment mit einer Befragung konzipiert.[1] Nach einem Pretest im Dezember 2024 wurde die Erhebung im Februar 2025 durchgeführt. Hierfür wurde ein Befragungsunternehmen beauftragt, das eine zertifizierte, wissenschaftlich angemessene Stichprobe und Erhebung gewährleistet. Insgesamt konnten nach einer Plausibilitätsprüfung 410 von ursprünglich 427 Befragten in die Auswertung einfließen, die derzeit oder in den letzten 12 Monaten im Personalmanagement abhängig beschäftigt waren. Das Experiment und die Befragung wurden online durchgeführt und an digitalen Endgeräten erhoben.

3.1 Experimente

Für den Versuchsaufbau des Experimentalteils haben wir Nutzungsoberflächen auf Hersteller- und Entwickler-Homepages recherchiert, die aktiv KI-Funktionen bewerben. Die ausgewählten Oberflächen haben wir leicht bearbeitet und zusätzlich eine zweite Version mit Erklärungen erstellt. Um die Erklärungseffekte bestimmen zu können, zeigten wir den Befragten zunächst die Grundversion, anschließend die ergänzte Version. Ob die Proband*innen die ursprüngliche Software selbst verwenden, wurde nicht abgefragt. Das birgt die Gefahr, dass Befragte bereits Erfahrung damit gemacht haben, die ihr Antwortverhalten beeinflussen könnte.

Zu den dargestellten Nutzungsoberflächen gab es zwei Fragegruppen. Einerseits interessierten uns die *subjektiven Wahrnehmungen* bezüglich der dargestellten Oberflächen: Die Oberfläche ist vertrauenswürdig, hilfreich, verständlich oder transparent (Likert-skaliert von 1 „stimme überhaupt nicht zu“ bis 5 „stimme voll und ganz zu“). Andererseits haben wir das *objektive Verstehen* der Personen hinsichtlich der dargestellten Informationen geprüft, indem wir den Teilnehmenden Aussagen zu den Oberflächen zeigten (z. B. „die zweitplatzierte Person hat passendere Kompetenzen“), die entweder richtig, falsch oder aufgrund der vorhandenen Informationen gar nicht getroffen werden konnten.

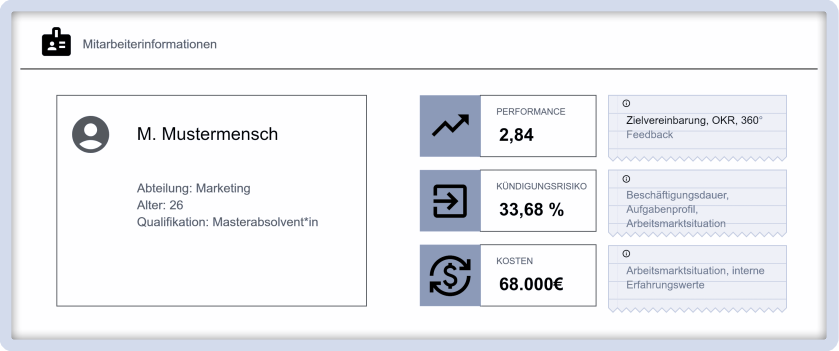
Es wäre nun zu erwarten, dass ein hoher Wert bei der subjektiven Wahrnehmung einer Oberfläche auch zu einer hohen Anzahl objektiv richtiger Antworten führt. In einem zweiten Durchlauf haben wir die um Erklärbare KI-Elemente ergänzten Oberflächen präsentiert. Dadurch können wir untersuchen, wie Erklärungen subjektive Wahrnehmung und objektives Verstehen beeinflussen, und wie sich unterschiedliche KI-Kompetenzniveaus darauf auswirken.

Den Teilnehmenden wurden drei verschiedene Nutzungsoberflächen präsentiert. Das *erste Experiment (Abbildung 1)* zeigt eine schlichte Oberfläche zur Darstellung der Leistung und des Kündigungsrisikos einer fiktiven Person. Die Datengrundlage für die Metriken wird nicht näher erläutert; zudem fehlen Vergleichswerte oder konkrete Skalenangaben. Die Erklärungen im zweiten Durchlauf zielten darauf ab, den Datenursprung der jeweiligen Metriken zu kontextualisieren. *Experiment 2 (Abbildung 2)* zeigt ein Ranking von drei fiktiven Bewerber*innen, deren Namen ohne methodische Hintergedanken gewählt wurden. In der Ausgangsdarstellung sind nur der Rang und ein Indikator zur Passung der jeweiligen Bewerbung angegeben. Platz 1 besaß eine geringere Passung als Platz 2 und 3 – ein Detail, das dem Original entnommen wurde. Zur Bewertung der Erklärungen im zweiten Durchlauf wurden die Teilnehmenden zufällig einem von drei Erklärungstypen zugeteilt. Die erste Gruppe umfasste Erklärungen der relevanten Faktoren *(important features).* Diese zeigen an, wie stark der Wert eines einzelnen Faktors (ein spezifisches Detail aus der Bewerbung) den Output des Modells für eine konkrete Vorhersage oder Einsortierung beeinflusst. Die zweite Gruppe zeigte kontrafaktische Erklärungen: Wie verschiebt sich die Sortierung, wenn sich bestimmte Werte der Bewerbung verändern? Die dritte Gruppe sah die allgemeinen Modellkriterien, nach denen die Sortierung erfolgte – also welche Aspekte das Modell bei allen Bewerbungen als besonders wichtig erachtet. *Experiment 3 (Abbildung 3)* zeigt ein Userinterface im Format einer Handy-App, das die Passung einer einzelnen Bewerberin für eine offene Stelle anhand einer Fünf-Sterne-Skala bewertet. In zwei unterschiedlichen Erklärungen, denen die Teilnehmenden erneut zufällig zugeteilt wurden, wurden relevante Faktoren sowie kontrafaktische Bewertungen dargestellt.

1 Dies geschah im Rahmen des Forschungsprojekts „TRANKI – Standards für transparente KI“, das von der Hans-Böckler-Stiftung gefördert wurde. Förderzeichen: 2022-797-2. Projektlaufzeit 10/2023 bis 9/2026. Projektleitung: Katharina Simbeck. Der zugrunde liegende Datensatz ist veröffentlicht (Simbeck/Kalff 2025).

ABBILDUNG 1

Oberfläche für Experiment 1 – Darstellung der Leistung und des Kündigungsrisikos einer fiktiven Person



Anmerkung: Allen Proband*innen wurde die Oberfläche zuerst ohne, dann mit Erklärungen (rechte Spalte) gezeigt. Die Erklärungen legen relevante Faktoren für die Berechnung von Leistung und Kündigungsrisiko offen.

Quelle: Eigene Darstellung nach https://www.functionhr.com

WSI Mitteilungen

ABBILDUNG 2

Oberfläche für Experiment 2 – Ranking von drei fiktiven Bewerber*innen

Rang / Platz	Name	Skill Match	Einflussfaktoren
1	M. Mustermensch		Gehaltsvorstellung, Berufserfahrung
2	N. Tranh		Kenntnisse, Fähigkeiten, Weiterbildungen
3	O. Çelebi		Gehaltsvorstellung

Rang / Platz	Name	Skill Match	Was wäre wenn?
1	M. Mustermensch		Niedriger positioniert, wenn Gehaltswunsch >49.000€
2	N. Tranh		höher positioniert, wenn Gehaltswunsch <53.000€
3	O. Çelebi		Wäre auf Platz 1, wenn 5 Jahre Berufserfahrung mehr

Rang / Platz	Name	Skill Match	Entscheidungsregel
1	M. Mustermensch		Berufserfahrung > 5 Jahre, Gehaltswunsch < 50.000€, Zusatzqualifikationen vorhanden.
2	N. Tranh		
3	O. Çelebi		

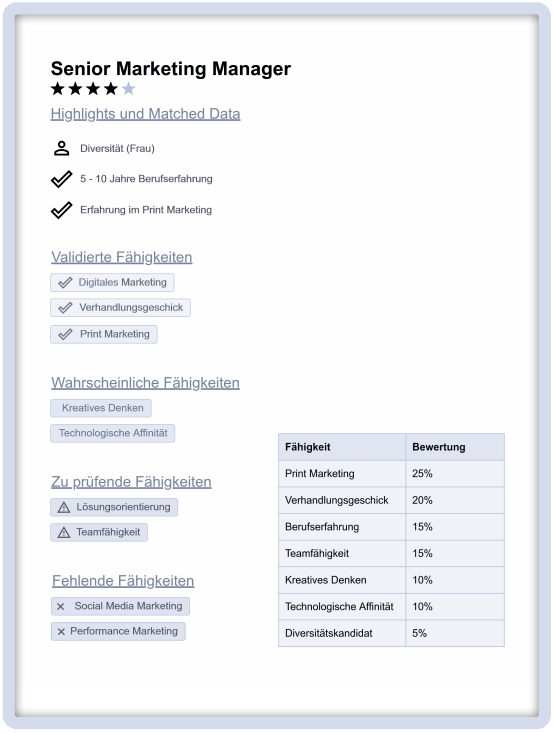
Anmerkung: Allen Proband*innen wurde die Oberfläche zuerst ohne, dann mit Erklärungen (rechte Spalte) gezeigt. Die rechte Spalte zeigt pro Gruppe per Zufallszuteilung jeweils einen Erklärungstyp: „relevante Einflussfaktoren", „kontrafaktische Erklärung" (Was wäre, wenn?) oder „Modellkriterien".

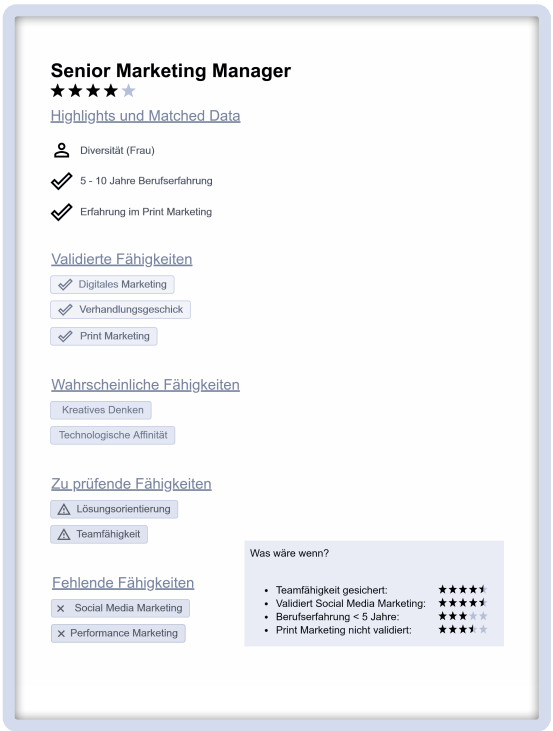
Quelle: Eigene Darstellung nach https://cvviz.com

WSI Mitteilungen

ABBILDUNG 3

Oberfläche für Experiment 3 – Passung einer Bewerberin anhand einer 5-Sterne-Skala





Anmerkung: Allen Proband*innen wurde die Oberfläche zuerst ohne, dann mit Erklärungen gezeigt (rechte Spalten), je nach Gruppe als „relevante Faktoren" oder als „kontrafaktische Erklärungen" (Was wäre, wenn?).

Quelle: Eigene Darstellung nach https://eightfold.ai

WSI Mitteilungen

3.2 Erklärende Variable: KI-Kompetenz

KI-Kompetenzen erheben wir in unserem Studiendesign mithilfe des SNAIL-Instruments von Laupichler et al. (2023). KI-Kompetenzen werden in drei Dimensionen operationalisiert: technisches Verständnis, kritische Kompetenz und praktische Anwendungen. Insgesamt haben wir von den 30 vorgeschlagenen Items 15 (fünf je Dimension) ausgewählt, die für das Personalmanagement besonders relevant sind. Der Indikator für KI-Kompetenz addiert die Werte aller drei Teildimensionen.

Die Items messen die selbsteingeschätzte Kompetenz zuverlässig. *Tabelle 1* stellt Mittelwerte und Standardabweichungen, Cronbachs Alpha sowie die durchschnittliche Korrelation der Items untereinander dar. Cronbachs Alpha und die mittlere Item-Korrelation bestimmen, ob die Items Ähnliches messen und damit für einen Indikator aussagekräftig sind (Reliabilität). Selbsteingeschätzte KI-Kompetenzen sind die selbst wahrgenommenen und sich selbst zugeschriebenen Fähigkeiten und Verständnisse. Der Sache nach können sie verzerrt sein, denn Menschen neigen psychologisch dazu, sich selbst zu über- bzw. zu unterschätzen, insbesondere bei abstrakten technischen Themen (Zell/Krizan 2014). Das heißt: Es wird eine subjektive Einschätzung erhoben – keine objektive Messung der individuellen KI-Kompetenzen.

TABELLE 1

Gütekriterien der verwendeten Indikatoren

Skala	Anzahl Items	Cronbachs α	Mittelwert (SD)	mittlere Item-Korrelation
technisches Verständnis	5	0,92	2,85 (1,28)	0,79
kritische Kompetenz	5	0,90	3,29 (1,22)	0,75
praktische Anwendungen	5	0,89	3,10 (1,24)	0,73

Quelle: Eigene Daten

WSI Mitteilungen

4 Ergebnisse

Die Auswertung unserer Experimente folgt der Annahme, dass Erklärbare KI die subjektive Wahrnehmung von KI-basierten Oberflächen verbessert und Sicherheitsempfinden, Verständlichkeit, Unterstützung und Transparenz für Nutzer*innen fördert. Außerdem gehen wir davon aus, dass Erklärbare KI zum objektiven Verstehen von KI und KI-Ergebnissen beiträgt. Gleichwohl nehmen wir an, dass diese Effekte nicht allen Beteiligten gleichermaßen zugutekommen, sondern dass individuelle KI-Kompetenzen einen Anteil an der Wirksamkeit der Erklärbaren KI-Elemente haben. *Tabelle 2* fasst die statistischen Befunde nach Experiment, Erklärungsart und KI-Kompetenzniveau zusammen. Die Spalten ohne/mit beziehen sich auf Erklärbare KI-Darstellungen: Alle Proband*innen sahen zuerst eine Oberfläche ohne und dann eine Oberfläche mit Erklärungen. Der Tabelle kann man für Experiment 1 entnehmen, dass sich bei niedriger KI-Kompetenz die subjektive Wahrnehmung durch hinzugefügte Erklärungen geringfügig, aber nicht signifikant verschlechtert (von 2,97 auf 2,93) – die Anteile richtiger Antworten im objektiven Verstehen verbessern sich jedoch signifikant (von 53,5 % zu 58,1 %).

Zur Untersuchung der Wirkung von Erklärungen auf die subjektive Wahrnehmung und das objektive Verstehen von KI wurden mehrfaktorielle ANOVAs mit Messwiederholung durchgeführt. Zur weiteren Kontrastierung unserer Analyse haben wir auch das Geschlecht berücksichtigt.[2] Dieses Verfahren ermöglicht die simultane Ana-

TABELLE 2

Auswirkung von KI-Kompetenzen und Erklärungen auf die subjektive Wahrnehmung und das objektive Verstehen von KI

Experiment	Erklärung	KI-Kompetenz	Subjektive Wahrnehmung (Mittel)			Objektives Verstehen (% richtig)		
			ohne	mit	Sig.	ohne	mit	Sig.
1	relevante Faktoren	niedrig	2,97	2,93		53,5	58,1	*
		mittel	3,31	3,34		42,8	48,3	***
		hoch	3,84	3,87		30,6	43,2	***
2	relevante Faktoren	niedrig	2,62	2,83		53,9	49,7	
		mittel	2,99	3,19	**	55,1	59,0	
		hoch	3,57	3,80	*	39,6	67,0	***
	kontrafaktische Erklärungen	niedrig	2,20	2,40		58,6	47,1	*
		mittel	2,76	3,03		51,8	46,5	
		hoch	3,31	3,68		45,7	51,0	
	Modellkriterien	niedrig	2,65	2,69		54,7	43,3	*
		mittel	2,56	2,92	***	53,7	44,0	**
		hoch	3,37	3,61	*	49,6	40,9	**
3	relevante Faktoren	niedrig	3,18	3,18		49,5	51,9	
		mittel	3,35	3,38		54,3	58,5	**
		hoch	3,92	3,89		55,9	63,5	***
	kontrafaktische Erklärungen	niedrig	3,14	3,10		50,6	54,3	
		mittel	3,48	3,50		54,0	59,4	***
		hoch	3,97	3,92		56,6	63,4	***

Anmerkungen: Die subjektive Wahrnehmung entspricht dem Mittelwert einer 5-stufigen Likert-Skala. Das objektive Verstehen wird als prozentualer Anteil korrekter Antworten angegeben. Die statistische Signifikanz (Spalte „Sig.“) wurde mittels des Wilcoxon-Vorzeichen-Rangtests überprüft und gibt an, ob sich Ergebnisse mit Erklärungen signifikant von Ergebnissen ohne Erklärungen unterscheiden (* $p \leq 0{,}05$; ** $p \leq 0{,}01$; *** $p \leq 0{,}001$).

Quelle: Eigene Darstellung

WSI Mitteilungen

2 Geschlecht wurde von uns kategorial als M/W/D abgefragt mit der zusätzlichen Option einer Freitextantwort. Das Sample meldete W: 253, M: 157 zurück.

lyse mehrerer unabhängiger kategorialer Variablen (KI-Kompetenz, Geschlecht, Art der KI-Erklärung) sowie deren Interaktionseffekte auf die abhängigen Variablen. Der Messwiederholungsfaktor (ohne/mit Erklärung) prüft, ob die hinzugefügte Erklärung zu signifikanten Veränderungen führt und ob diese Effekte durch die Zwischensubjektfaktoren – also jene, durch die sich Individuen unterscheiden, wie KI-Kompetenz, Geschlecht usw. – beeinflusst werden. Dadurch lassen sich Haupt- und Interaktionseffekte sowie deren Signifikanz systematisch auswerten.

4.1 Wie Nutzer*innen die Nutzungsoberflächen empfinden

Für die subjektive Wahrnehmung der Proband*innen zeigen alle drei Experimente dasselbe Muster: Wer seine KI-Kompetenzen hoch einschätzt, vertraut der Oberfläche mehr und findet sie hilfreicher. Zusätzliche Erklärungen ändern daran insgesamt wenig. Nur in Experiment 2 verbessern Erklärungen das Vertrauen in die Oberflächen – allerdings nur bei Personen mit mittleren oder hohen KI-Kompetenzen.

Experiment 1: Die Erklärungen allein bewirken nichts. Entscheidend sind die KI-Kompetenzen der Teilnehmenden – sie beeinflussen das Vertrauen stark. Männer mit geringen KI-Kompetenzen stellen einen Sonderfall dar: Sie vertrauen den Oberflächen weniger als Frauen mit demselben Wissensstand.

Experiment 2: Hier wurden verschiedene Erklärungstypen getestet. KI-Kompetenzen bleiben der wichtigste Faktor. Die Erklärungen wirken zwar nicht alle gleich, aber: Bei einer zunächst widersprüchlichen und abstrakten Darstellung steigt das Vertrauen, nachdem Erklärungen zu „relevanten Faktoren“ und „Modellkriterien“ hinzugefügt wurden, statistisch signifikant – jedoch nur bei Personen mit mittleren oder hohen KI-Kompetenzen. Signifikante Geschlechterunterschiede zeigen sich nur bei dem Erklärungstyp „Modellkriterien“, die Frauen signifikant höher bewerten.

Experiment 3: Hier zählen wiederum vor allem KI-Kompetenzen. Geschlecht und Art der Erklärung spielen keine statistisch signifikante Rolle für die subjektive Wahrnehmung.

4.2 Wie gut Nutzer*innen die Oberflächen tatsächlich verstehen

Die Auswertung des objektiven Verstehens liefert uneinheitliche Ergebnisse. In den Experimenten 1 und 2 schneiden Personen mit hoher KI-Kompetenz am schlechtesten ab. In Experiment 3 verstehen Menschen mit hoher KI-Kompetenz die Oberflächen am besten. Erklärungen helfen in den Experimenten 1 und 3, insbesondere bei mittleren und hohen KI-Kompetenzen. In Experiment 2 hingegen *schaden* bestimmte Erklärungen sogar: „Modellkriterien“ verschlechtern das Verständnis bei allen Gruppen, „kontrafaktische Erklärungen“ hingegen besonders bei Personen mit geringer KI-Kompetenz.

Experiment 1: Die Erklärung allein verbessert das Verständnis nicht. Aber sie wirkt je nach KI-Kompetenz unterschiedlich. Frauen mit hoher KI-Kompetenz verstehen die Oberfläche besser durch eine Erklärung. Männer mit mittleren KI-Kompetenzen erzielen weniger Punkte. Bei niedrigen KI-Kompetenzen verbessert der Stimulus die Ergebnisse unabhängig vom Geschlecht.

Experiment 2: Hier zeigen sich starke Unterschiede je nach Erklärungstyp, KI-Kompetenzen und Messzeitpunkt. Männer und Frauen verstehen die Oberflächen gleichermaßen gut. Die Erklärung „relevante Faktoren“ hilft Personen mit hohen KI-Kompetenzen deutlich – sowohl Männern als auch Frauen. Problematisch ist die Erklärung „Modellkriterien“ für Personen mit geringen KI-Kompetenzen und hier insbesondere für Männer: Ihr objektives Verständnis sinkt. Insgesamt verschlechtern „Modellkriterien“ das Verständnis bei allen Gruppen deutlich, während „kontrafaktische Erklärungen“ die Ergebnisse vor allem bei geringen KI-Kompetenzen verschlechtern.

Experiment 3: KI-Kompetenzen beeinflussen das Verständnis deutlich. Die Erklärung allein bringt keine messbare Verbesserung. Erst in der Kombination aus Erklärung und KI-Kompetenz wird statistisch deutlich, dass sich das objektive Verstehen bei mittlerer und hoher KI-Kompetenz verbessert. Ein auffälliges Ergebnis ist, dass bei der Erklärung „relevante Faktoren“ Männer mit wenig KI-Kompetenzen die Oberfläche deutlich besser verstehen als Frauen mit gleichem Wissensstand.

5 Diskussion: KI-Kompetenzen zwischen individueller Befähigung und kollektiver Gestaltung

In allen drei Experimenten zeigt sich ein konsistentes Muster: KI-Kompetenz ist eine verlässliche Vorhersage für die subjektive Wahrnehmung von KI-gestützten Nutzungsoberflächen. Zugleich verdeutlichen die Befunde eine teils widersprüchliche Beziehung zwischen der selbsteingeschätzten Kompetenz, der Erklärungsart und dem objektiven Verstehen. Erklärungen sind nicht universell wirksam: Personen mit hoher KI-Kompetenz profitieren punktuell, während bei niedriger Kompetenz eine Verschlechterung des Ergebnisses möglich ist. Für Beschäftigte und betriebliche Interessenvertretungen folgt daraus, dass Erklärbare KI keine selbsterklärende, technische Lösung für objektives Verstehen ist. KI-Kompetenzen sind hierfür ebenso notwendig, aber nicht hinreichend. Erklärbare KI muss mit KI-Kompetenzen als Teil eines zusammenhängenden soziotechnischen Gestaltungsprojekts verankert werden.

5.1 Zur Bedeutung von KI-Kompetenzen für Personalmanager*innen

Unsere Studie zeigt, dass KI-Kompetenzen zwar die subjektive Wahrnehmung positiv beeinflussen, jedoch nicht automatisch zu einem besseren objektiven Verstehen führen. Gerade in den Experimenten 1 und 2 schneiden Teilnehmende mit hoher KI-Kompetenz in objektiven Tests schlechter ab. Zwei Mechanismen sind relevant: Erstens kann Selbsteinschätzung in Übervertrauen umschlagen – Beschäftigte mit „viel" Kompetenz verlassen sich eher auf vermeintliche Plausibilität, statt Grenzen zu reflektieren. Das deckt sich mit Erkenntnissen zur kognitiven Überlastung durch zu komplexe Erklärungen (Miller 2019) sowie der Gefahr einer „Illusion der Erklärungskraft" (Chromik et al. 2021), bei der Nutzer*innen glauben, mehr zu verstehen, als sie tatsächlich nachvollziehen können. Bei der zweiten Forschungsfrage variiert die Passung zwischen Erklärformaten und Anforderungsniveaus. Kontrafaktische Erklärungen oder abstrakte Modellkriterien helfen ungeübten Nutzer*innen nicht und verschlechtern ihr Verständnis.

Beschäftigte im Personalmanagement – aber auch darüber hinaus – benötigen ein Bündel an KI-Kompetenzen, das technisches Grundwissen, kritische Bewertung und ethische Reflexion integriert (Cetindamar et al. 2024). Da das Personalmanagement ein Hochrisiko-Bereich darstellt, müssen Anwender*innen bei der Nutzung von KI besondere Vorsicht walten lassen, denn sie bearbeiten Personalanliegen, bereiten Entscheidungen vor oder treffen sie mit technischer KI-Unterstützung (Bildat/Brouns 2026). KI-Kompetenzen sind daher unerlässlich, um eine sichere Anwendung der Software sowie die korrekte Interpretation und Bewertung ihrer Ergebnisse zu gewährleisten. Erklärbare KI kann hierbei unterstützen, aber – wie wir gezeigt haben – mit dem Risiko, dass sie Ergebnisse verschlechtert, ohne dass dies den Anwender*innen auffällt. KI-Kompetenzen sind deshalb als kontinuierlicher Entwicklungsprozess zu verstehen, der Meta-Kompetenzen wie Unsicherheitskompetenz, Fehlerkultur oder Emotionsregulation umfasst (Carolus et al. 2023; Li/Kim 2024). Gerade im Personalmanagement bedeutet das, kritische und reflexive KI-Kompetenzen zu fördern, denn bei Entscheidungen, die die Karrieren anderer Menschen beeinflussen, sind sowohl eine Auslagerung von Verantwortung an die Technik als auch ein fehlendes Verständnis problematisch.

Die KI-Verordnung verlangt KI-Kompetenzen von betrieblichen Anwender*innen und von Personen, deren Daten von KI-Systemen verarbeitet werden. Daraus lässt sich jedoch nur indirekt ein rechtlicher Anspruch auf Fort- und Weiterbildung ableiten (Donath 2025). Beschäftigte müssen den KI-Kompetenzaufbau daher aktiv einfordern – was angesichts der exponierten Bedeutung des Personalmanagements für Karriereverläufe, individuelle Entwicklungs- und Lebenschancen sowie berufliche Perspektiven auch ein relevantes Feld und eine klare Aufgabe für die Mitbestimmung darstellt.

5.2 Handlungsperspektiven für Gewerkschaften und Betriebsräte

Die Befunde legen nahe, dass eine individuelle Befähigung zwar notwendig, jedoch nicht ausreichend ist. Erklärbarkeit bietet eine soziotechnische Schnittstelle zur Steigerung der Transparenz der KI, setzt jedoch auch KI-Kompetenzen voraus, um ihre Wirkung zu entfalten (Cambria et al. 2023). Gewerkschaften und Betriebsräte sollten bei der Einführung von KI-Systemen sowohl transparenzsteigernde Maßnahmen als auch Maßnahmen zur Steigerung der KI-Kompetenzen einfordern. KI-Kompetenzen sind als Gegenstand kollektiver Regelung sowie als Ressource der Mitbestimmung zu begreifen, die die Gremien selbst in die Lage versetzen, informiert über KI-Systeme mitzuentscheiden.

Betriebsvereinbarungen sind das Mittel der Wahl, um KI-Systeme betrieblich zu gestalten und neben Datenschutz, Zweckbindung sowie Automatisierungsansätzen auch die Erweiterung menschlicher Fähigkeiten und Qualifizierungsstrategien zu etablieren (Guggemos 2024). Diese könnten beispielsweise Mindeststandards für den KI-Einsatz im Personalmanagement enthalten, die arbeitsplatznahe KI-Kompetenzen sowie zielgruppenspezifische Erklärungsstandards beinhalten (z. B. kompetenzspezifische, domänensprachliche, kontrastive Erklärungen) und verpflichtende Evaluationen festschreiben, die nicht nur die technische Systemqualität, sondern auch das individuelle Verstehen prüfen.

Mitbestimmungsakteure sollten ebenfalls umfassend geschult werden. Auch Betriebsratsarbeit erfordert KI-Kompetenzen, um technische, ethische oder praktische Risiken (Fehlkalibrierung, Drift, verzerrte Trainingsdaten) erkennen und geeignete Maßnahmen verhandeln zu können. Dazu gehören Qualifizierungsrechte, Kontingente für Lernzeiten, Zugang zu Audit-Unterlagen und – im Sinne technischer Sorgfaltspflichten – das Recht auf Einsicht in Implementierungsunterlagen sowie in proprietäre Systeme (Walmsley 2021). KI-Kompetenzentwicklung in Betriebsratsgremien ist daher nicht nur ein „nettes Extra", sondern macht Mitbestimmung in datengetriebenen Prozessen erst wirksam. Gewerkschaften können auf überbetrieblicher Ebene Rahmenbedingungen schaffen, indem sie durch sozialpartnerschaftliche Aushandlung KI-Kompetenzförderung in Tarifverträge integrieren und branchenspezifische Qualifizierungsstandards entwickeln (Molina et al. 2023). Strategien humanzentrierter KI-Gestaltung müssen systematisch um Qualifizierungs- und Transparenzpflichten ergänzt werden (Krzywdzinski et al. 2023), und die institutionelle Mitbestimmung muss KI-Expertise in ihren Strukturen aufbauen, um auf Augenhöhe mit Arbeitgebern und Softwareentwickler*innen algorithmische Systeme zu gestalten.

KI-Kompetenzen sind das zentrale Bindeglied zwischen transparenter Technik und mitbestimmter Arbeitsgestaltung – auch über das Personalmanagement hinaus. Systematischer KI-Kompetenzaufbau im Sinne der Mitbestimmung operationalisiert den europäischen Regulierungsrahmen als Lern- und Kontrollressource und macht die Qualifizierung der Beschäftigten und der Mitbestimmung zur Bedingung verantwortlicher KI-Nutzung. Damit lässt sich verhindern, dass KI eine undurchsichtige „Blackbox" bleibt (Kalff/Simbeck, i. E.). Stattdessen wird sie zum Gegenstand kollektiver Aushandlung – fachlich informiert, institutionell abgesichert und auf die praktische Urteilskraft der Beschäftigten ausgerichtet.

6 Fazit und Ausblick

Die vorliegende Untersuchung zeigt, dass KI-Kompetenzen das Vertrauen in KI-Systeme stärken, jedoch nicht automatisch zu besseren Interpretationsleistungen führen. Für eine verantwortungsvolle Nutzung von KI im Personalmanagement reichen individuelle Qualifizierung oder technische Erklärungen nicht aus. Sie muss durch institutionelle Rahmenbedingungen und kollektive Mitbestimmung flankiert werden, die den systematischen Aufbau von KI-Kompetenzen gewährleisten. Gewerkschaften und Betriebsräte sollten KI-Kompetenzen nicht nur für Beschäftigte einfordern, sondern auch in den eigenen Reihen aufbauen, um die Technologieeinführung und -nutzung kompetent mitzugestalten.

Die Aufgabe zukünftiger Forschung ist es, zu untersuchen, welche Qualifizierungsformate besonders wirksam sind, wie KI-Kompetenzen angemessen und an die jeweiligen Arbeits- und Organisationsprozesse angepasst vermittelt werden können und wie sich die Qualifizierung zwischen Interessenvertretungen, Arbeitgebern und Beschäftigten bei der Einführung von KI-Systemen konkret ausgestalten lässt. Zusätzlich braucht es technische Unterstützung für transparente KI, die jedoch nicht als Allheilmittel für komplexe KI-Modelle missverstanden werden darf: Sie wirkt nur bei kompetenten Beschäftigten.

Unsere Studie hat einige Limitationen, denn die selbsteingeschätzten KI-Kompetenzen können über- oder unterschätzt sein, worauf auch die geschlechtsspezifischen Unterschiede hindeuten. Frauen mit niedrigen KI-Kompetenzen schneiden etwas schlechter ab, während Männer mit niedrigeren Kompetenzen kritischer auf die Oberflächen blicken. Demgegenüber lassen die korrekten Antworten bei Männern mit hoher KI-Kompetenz überdurchschnittlich nach. Objektive Messungen von KI-Kompetenzen sind daher wissenschaftlich wie praktisch wünschenswert und stellen einen Anknüpfungspunkt für zukünftige Untersuchungen dar. ■

LITERATUR

Almatrafi, O. / Johri, A. / Lee, H. (2024): A Systematic Review of AI Literacy Conceptualization, Constructs, and Implementation and Assessment Efforts (2019–2023), in: Computers and Education Open 6, DOI: 10.1016/j.caeo.2024.100173

Bauer, K. / Zahn, M. / Hinz, O. (2023): Expl(AI)ned: The Impact of Explainable Artificial Intelligence on Users' Information Processing, in: Information Systems Research 34 (4), S. 1582–1602, DOI: 10.1287/isre.2023.1199

Benlian, A. / Pinski, M. (2025): The AI Literacy Development Canvas: Assessing and Building AI Literacy in Organizations, in: Business Horizons, DOI: 10.1016/j.bushor.2025.10.001

Bildat, L. / Brouns, T. (2026): Selected or Rejected by Bots? Musings on AI Literacy in Human Resource Management, Upskilling of HR Experts, and a Call for Applied Research, in: Bodrožić-Brnić, K. / Schulte, V. / Thiessen, T. (Hrsg.): Digital Transformation in Knowledge Transfer, Cham, S. 3–28

Böhmer, N. / Schinnenburg, H. (2023): Critical Exploration of AI-driven HRM to build up Organizational Capabilities, in: Employee Relations: The International Journal 45 (5), S. 1057–1082, DOI: 10.1108/ER-04-2022-0202

Bondarouk, T. / Brewster, C. (2016): Conceptualising the Future of HRM and Technology Research, in: The International Journal of Human Resource Management 27 (21), S. 2652–2671, DOI: 10.1080/09585192.2016.1232296

Cambria, E. / Malandri, L. / Mercorio, F. / Mezzanzanica, M. / Nobani, N. (2023): A Survey on XAI and Natural Language Explanations, in: Information Processing and Management 60 (1), DOI: 10.1016/j.ipm.2022.103111

Carolus, A. / Koch, M. J. / Straka, S. / Latoschik, M. E. / Wienrich, C. (2023): MAILS – Meta AI Literacy Scale: Development and Testing of an AI Literacy Questionnaire based on Well-founded Competency Models and Psychological Change- and Meta-Competencies, in: Computers in Human Behavior: Artificial Humans 1 (2), DOI: 10.1016/j.chbah.2023.100014

Cetindamar, D. / Kitto, K. / Wu, M. / Zhang, Y. / Abedin, B. / Knight, S. (2024): Explicating AI Literacy of Employees at Digital Workplaces, in: IEEE Transactions on Engineering Management 71, S. 810–823, DOI: 10.1109/TEM.2021.3138503

Chromik, M. / Eiband, M. / Buchner, F. / Krüger, A. / Butz, A. (2021): I Think I Get Your Point, AI! The Illusion of Explanatory Depth in Explainable AI, in: 26th International Conference on Intelligent User Interfaces, College Station TX USA, S. 307–317

Coron, C. / Scheibmayr, I. / Lescoat, P. (2024): How to do HRM with Numbers? A Performative Lens on HR Metrics, HR Analytics and HR Algorithms, in: New Technology, Work and Employment 40 (1), S. 124–146, DOI: 10.1111/ntwe.12306

Donath, P. (2025): Der Anspruch auf Fortbildung für Beschäftigte im Bereich KI aufgrund der KI-Verordnung, in: Digitales Arbeits- und Sozialrecht 2/2025, S. 61–65

Ehsan, U. / Passi, S. / Liao, Q. V. / Chan, L. / Lee, I.-H. / Muller, M. /et al. (2024): The Who in XAI: How AI Background Shapes Perceptions of AI Explanations, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, Honolulu HI USA, S. 1–32

Guggemos, J. (2024): To Fear or not to Fear – Human Resource Development Professionals' Positioning towards Artificial Intelligence with a Focus on Augmentation, in: Computers and Education: Artificial Intelligence 7, DOI: 10.1016/j.caeai.2024.100260

Guidotti, R. / Monreale, A. / Ruggieri, S. / Turini, F. / Giannotti, F. / Pedreschi, D. (2019): A Survey of Methods for Explaining Black Box Models, in: ACM Computing Surveys 51 (5), S. 1–42, DOI: 10.1145/3236009

Kalff, Y. / Simbeck, K. (2025): Explained, Yet Misunderstood: How AI Literacy Shapes HR Managers' Interpretation of User Interfaces in Recruiting Recommender Systems, in: Kaya, M. / Bogers, T. / Bied, G. / Johnson, C. / Decorte, J.-J. (Hrsg.): Proceedings of the 5th Workshop on Recommender Systems for Human Resources (RecSys in HR 2025), Prague, https://ceur-ws.org/Vol-4046/RecSysHR2025-paper_3.pdf (letzter Zugriff: 18. 06. 2026)

Kalff, Y. / Simbeck, K. (i. E.): Black-Box HR-Analytics? Mythen und Realitäten eines KI-unterstützten Personalmanagements, in: Hütten, M. / Klöpper, M. / Lücking, S. (Hrsg.): Mythen und Realitäten der Digitalisierung, Bielefeld

Krzywdzinski, M. / Gerst, D. / Butollo, F. (2023): Promoting Human-centred AI in the Workplace. Trade Unions and their Strategies for Regulating the Use of AI in Germany, in: Transfer: European Review of Labour and Research 29 (1), S. 23–70, DOI: 10.1177/10242589221142273

Kühl, N. / Meske, C. / Nitsche, M. / Lobana, J. (2023): Investigating the Role of Explainability and AI Literacy in User Compliance, in: SSRN Electronic Journal, DOI: 10.2139/ssrn.4558966

Langer, M. / Baum, K. / Hartmann, K. / Hessel, S. / Speith, T. / Wahl, J. (2021): Explainability Auditing for Intelligent Systems: A Rationale for Multi-Disciplinary Perspectives, in: IEEE (Hrsg.): 29th IEEE International Requirements Engineering Conference Workshops, Piscataway, S. 164–168

Laupichler, M. C. / Aster, A. / Haverkamp, N. / Raupach, T. (2023): Development of the „Scale for the Assessment of Non-experts' AI literacy" – An Exploratory Factor Analysis, in: Computers in Human Behavior Reports 12 (1), DOI: 10.1016/j.chbr.2023.100338

Li, H. / Kim, S. (2024): Developing AI Literacy in HRD: Competencies, Approaches, and Implications, in: Human Resource Development International 27 (3), S. 345–366, DOI: 10.1080/13678868.2024.2337962
Long, D. / Roberts, J. / Magerko, B. / Holstein, K. / DiPaola, D. / Martin, F. (2023): AI Literacy: Finding Common Threads between Education, Design, Policy, and Explainability, in: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg Germany, S. 1–6
Loscher, G. J. / Bader, V. (2023): Creating Accountability Through HR Analytics – An Audit Society Perspective, in: Human Resource Management Review 33 (4), DOI: 10.1016/j.hrmr.2023.100974
Martini, M. / Wendehorst, C. (2024): KI-VO: Verordnung über Künstlichen Intelligenz: Kommentar, München
Miller, T. (2019): Explanation in Artificial Intelligence: Insights from the Social Sciences, in: Artificial Intelligence 267, S. 1–38, DOI: 10.1016/j.artint.2018.07.007
Molina, O. / Butollo, F. / Makó, C. / Godino, A. / Holtgrewe, U. / Illsoe, A. et al. (2023): It Takes Two to Code: A Comparative Analysis of Collective Bargaining and Artificial Intelligence, in: Transfer: European Review of Labour and Research 29 (1), DOI: 10.1177/10242589231156515
Molnar, C. (2025): Interpretable Machine Learning: A Guide for Making Black Box Models explainable, 3. Aufl., München
Ncube, T. R. / Sishi, K. K. / Skinner, J. P. (2025): The Impact of Artificial Intelligence on Human Resource Management Practices: An Investigation, in: SA Journal of Human Resource Management 23, DOI: 10.4102/sajhrm.v23i0.2960
Ng, D. T. K. / Leung, J. K. L. / Chu, K. W. S. / Qiao, M. S. (2021): AI Literacy: Definition, Teaching, Evaluation and Ethical Issues, in: Proceedings of the Association for Information Science and Technology 58 (1), S. 504–509, DOI: 10.1002/pra2.487
Páez, A. (2019): The Pragmatic Turn in Explainable Artificial Intelligence (XAI), in: Minds and Machines 29 (3), S. 441–459, DOI: 10.1007/s11023-019-09502-w
Schellmann, H. (2024): The Algorithm: How AI Decides Who Gets Hired, Monitored, Promoted, and Fired and Why We Need to Fight Back Now, New York
Simbeck, K. / Kalff, Y. (2025): Standards for Transparent AI in Human Resource Management (TRANKI) – Research Data, DOI: 10.5281/ZENODO.17708792
Strauß, S. (2024): CAIL – Critical AI Literacy. Kritische Technikkompetenz für konstruktiven Umgang mit KI-basierter Technologie in Betrieben, Institut für Technikfolgen-Abschätzung der Österreichischen Akademie der Wissenschaften, Wien
Tambe, P. / Cappelli, P. / Yakubovich, V. (2019): Artificial Intelligence in Human Resources Management: Challenges and a Path Forward, in: California Management Review 61 (4), S. 15–42, DOI: 10.1177/0008125619867910
Tripathi, V. M. (2025): The Future of Talent Management: The Promise and Challenges of Artificial Intelligence, Big Data and Machine Learning, in: NHRD Network Journal 18 (1), S. 77–87, DOI: 10.1177/26314541251325706
Walmsley, J. (2021): Artificial Intelligence and the Value of Transparency, in: AI & Society 36 (2), S. 585–595, DOI: 10.1007/s00146-020-01066-z
Weitz, K. / Schiller, D. / Schlagowski, R. / Huber, T. / André, E. (2021): „Let me explain!": Exploring the Potential of Virtual Agents in Explainable AI Interaction Design, in: Journal on Multimodal User Interfaces 15 (2), S. 87–98, DOI: 10.1007/s12193-020-00332-0
Zell, E. / Krizan, Z. (2014): Do People Have Insight Into Their Abilities? A Metasynthesis, in: Perspectives on Psychological Science 9 (2), S. 111–125, DOI: 10.1177/1745691613518075

AUTOR*INNEN

YANNICK KALFF, Dr., wissenschaftlicher Mitarbeiter in der Arbeitsgruppe „Informatik und Gesellschaft" an der Hochschule für Technik und Wirtschaft Berlin. Forschungsschwerpunkte: KI und KI-Kompetenzen in der Arbeit, Transparenz und Erklärbarkeit von KI, Gestaltung des betrieblichen Technikeinsatzes.

yannick.kalff@posteo.de

KATHARINA SIMBECK, Prof. Dr., Professorin und Leiterin der Arbeitsgruppe „Informatik und Gesellschaft" an der Hochschule für Technik und Wirtschaft Berlin. Forschungsschwerpunkte: KI im Personalmanagement, KI und KI-Kompetenzen in der Arbeit, Schul- und Hochschulbildung, Fairness und Auditierbarkeit von KI.

simbeck@htw-berlin.de