



Künstliche Intelligenz und Individuelle Unterschiede

Eine Perspektive der Fachgruppe Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik

Ramona Schödel^{1,2}, Elisa Altgassen³, Mitja D. Back^{4,5}, Martje Buss², Markus Bühner², Julian M. Etzel¹, Benjamin Goecke⁶, Sandra Grinschgl⁷, Eric Grunenbergl⁴, Ole Hätscher⁴, Björn E. Hommel⁸, Tobias L. Kordsmeyer⁹, Sophie Körner¹⁰, Niclas Kuper^{4,5}, Rui Mata¹¹, Patrick Mussel¹², Aaron Petrasch², Beatrice Rammstedt^{13,14}, Richard Rau¹⁵, Cassie Ann Short¹⁶, Cornelia Sindermann^{1,17}, Clemens Stachl¹⁸, Lukasz Stasielowicz¹⁹, Larissa Sust², Jasmin Thelen³, Selina Weiss²⁰, Oliver Wilhelm³, Dirk Wulff²¹, Cornelia Wrzus²² und Matthias Ziegler²³

¹ Charlotte Fresenius Hochschule, University of Psychology, München, Deutschland

² Ludwig-Maximilians-Universität München, Deutschland

³ Universität Ulm, Deutschland

⁴ Universität Münster, Deutschland

⁵ Joint Institute for Individualisation in a Changing Environment (JICE), Universität Münster und Universität Bielefeld, Deutschland

⁶ Hector-Institut für Empirische Bildungsforschung, Universität Tübingen, Deutschland

⁷ Universität Bern, Schweiz

⁸ Universität Leipzig, Deutschland

⁹ Georg-August-Universität Göttingen, Deutschland

¹⁰ Johannes Gutenberg Universität Mainz, Deutschland

¹¹ Universität Basel, Schweiz

¹² Psychologische Hochschule Berlin, Deutschland

¹³ Universität Mannheim, Deutschland

¹⁴ GESIS Leibniz-Institut für Sozialwissenschaften, Deutschland

¹⁵ Institute for Mind, Brain and Behavior, HMU Health and Medical University Potsdam, Deutschland

¹⁶ Carl von Ossietzky Universität Oldenburg, Deutschland

¹⁷ Universität Stuttgart, Deutschland

¹⁸ Institute of Behavioral Science and Technology, Universität St. Gallen, Schweiz

¹⁹ Universität Salzburg, Österreich

²⁰ Institut für Psychologie, Universität Hildesheim, Deutschland

²¹ Max-Planck-Institut für Bildungsforschung Berlin, Deutschland

²² Universität Heidelberg, Deutschland

²³ Humboldt-Universität zu Berlin, Deutschland

Zusammenfassung: In diesem Beitrag beleuchten wir Künstliche Intelligenz (KI) im Kontext der Erforschung und Messung individueller Unterschiede anhand dreier aktueller Themenfelder: KI als Forschungshilfsmittel, die Gestaltung der Mensch-KI-Interaktion sowie KI-gestützte Diagnostik. Anhand ausgewählter Beispiele verdeutlichen wir die vielfältigen Potenziale von KI auf konzeptueller und methodischer Ebene und zeigen zugleich, dass ihr Einsatz neue (Forschungs-)Fragen aufwirft – etwa zur menschlichen Kompetenz im Umgang mit KI, zur „Persönlichkeit“ von KI-Systemen oder zu Qualitätsstandards beim Einsatz KI-basierter diagnostischer Verfahren. Vor diesem Hintergrund diskutieren wir zentrale Voraussetzungen für eine nachhaltige Integration von KI in Forschung und Praxis, darunter kritische Validierung, Transparenz, interdisziplinäre Zusammenarbeit und die systematische Berücksichtigung individueller Unterschiede bei der verantwortungsvollen Gestaltung der Mensch-KI-Schnittstelle.

Schlüsselwörter: Künstliche Intelligenz (KI), Mensch-KI-Interaktion, KI-gestützte Diagnostik, individuelle Unterschiede, Transparenz und Validierung

Artificial Intelligence and Individual Differences. A Perspective of the Differential Psychology, Personality Psychology, and Psychological Assessment Division

Abstract: This paper examines artificial intelligence (AI) in the context of the study and measurement of individual differences across three current topics: AI as a research tool, the design of human-AI interaction, and AI-assisted assessment. Drawing on selected examples, we illustrate the wide-ranging potentials of AI at the conceptual and methodological level, while also showing that its application raises new (research) questions – for instance, regarding human competence in dealing with AI, the “personality” of AI systems, or psychometric standards in the use of AI-assisted assessment procedures. Against this backdrop, we discuss key prerequisites for a sustainable integration of AI into research and practice, including critical validation, transparency, interdisciplinary collaboration, and the systematic consideration of individual differences in the responsible design of the human-AI interface.

Keywords: artificial intelligence (AI), human-AI interaction, AI-assisted assessment, individual differences, transparency and validation

Künstliche Intelligenz (KI) entwickelt sich rasant und prägt zunehmend die individuelle Lebensgestaltung. Ihre wachsende gesellschaftliche Bedeutung fordert die Psychologie mehr denn je, sich wissenschaftlich fundiert mit den Konsequenzen ihrer Nutzung in Wissenschaft und Praxis auseinanderzusetzen. Vor diesem Hintergrund greifen wir als Fachgruppe Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik (DPPD) drei Themenfelder heraus, die Gegenstand unserer aktuellen Forschungsaktivitäten sind und die wir als besonders relevant einschätzen: (1) KI als Hilfsmittel in der Forschung; (2) Gestaltung der Mensch-KI-Interaktion; (3) KI-gestützte Diagnostik. Diese Auswahl ist nicht vollumfassend, sondern spiegelt ausgewählte Schwerpunkte wider, die die Vielfalt der Positionen der 30 beteiligten Expert_innen komprimiert abbilden und unsere fachspezifische Perspektive im interdisziplinären Diskurs sichtbar machen. In unserem Beitrag verstehen wir KI als Sammelbegriff für unterschiedlich komplexe Verfahren, die von klassischen Ansätzen des Maschinellen Lernens (ML) bis hin zu neueren großen Sprachmodellen (Large Language Models, LLMs) reichen.

KI als Hilfsmittel in der Forschung

Methodisch-konzeptuelles Hilfsmittel

Eine Herausforderung der Persönlichkeitsforschung ist die zunehmende Proliferation inhaltlich ähnlicher, teils überlappender Konstrukte, die Vergleiche von Befunden erschwert und kumulativen Erkenntnisfortschritt hemmt. LLMs können Items, Skalen und Bezeichnungen von Konstrukten in numerische Repräsentationen übersetzen,

um Beziehungen zwischen verschiedenen Messinstrumenten sowie zwischen Messinstrumenten und theoretischen Konzepten abzuschätzen (Wulff & Mata, 2025a). So lassen sich irreführende Bezeichnungen (Jingle-Jangle Fallacy) vermeiden und Redundanzen in bestehenden Taxonomien identifizieren.

Zudem können speziell trainierte KI-Modelle psychometrische Kennwerte wie interne Konsistenz, Inter-Item-Korrelationen, Faktorstrukturen sowie Zusammenhänge zwischen Skalen auf Basis von Itemtexten mit bemerkenswert hoher Genauigkeit schätzen (Hommel & Arslan, 2025; Wulff & Mata, 2025b). Dies erlaubt eine ökonomische Vorauswahl und den Vergleich von Instrumenten mit geringerem Validierungsaufwand im Vorfeld und erleichtert die Auswahl geeigneter Messinstrumente (z. B. <https://search.synth-science.org>).

Erschließung neuer Datenformate

KI ermöglicht die Verarbeitung unstrukturierter Daten (z. B. Text- und Bildmaterial), für die bislang aufwendige manuelle Ratings als Goldstandard galten. Auch „natürlich vorkommende“ digitale Datenquellen lassen sich besser erschließen. ML kann komplexe digitale Verhaltensspuren (z. B. Smartphone- oder Social-Media-Logs) strukturieren und psychologisch relevante Variablen extrahieren (Schoedel et al., 2026).

Zudem wird KI zunehmend als Möglichkeit diskutiert, den Aufwand umfangreicher Datenerhebungen zu reduzieren. Aktuelle Arbeiten erproben hierzu sogenannte *Silicon Samples*, bei denen menschliche Studienteilnehmende durch LLMs simuliert werden. Erste Befunde zeigen diesbezüglich sowohl Potenziale (z. B. Effizienzgewinne, Untersuchung sozial sensibler Themen) als auch bislang

unzureichend geklärte Fragen (Cui et al., 2025), die im folgenden Abschnitt näher beleuchtet werden.

Einordnung und Perspektiven für die Forschung

Am Beispiel der Silicon Samples wird ein grundlegender Trade-off des KI-Einsatzes im Forschungsprozess deutlich: Gesteigerte Effizienz bei zugleich bestehenden, noch unzureichend verstandenen Risiken. Studien mit KI-generierten Stichproben zeigen teils größere Effektstärken (Cui et al., 2025). Ob dies auf „sauberere“ Daten oder systematische Überschätzung zurückzuführen ist, bleibt unklar. Hinzu kommen offene Fragen zur Fairness (z. B. Verzerrungen durch Trainingsdaten) sowie zur Robustheit der Ergebnisse, etwa in Abhängigkeit von Prompt-Formulierungen oder inhaltlichen Kontexten (Cummins, 2025). Ein belastbares Gesamtbild zum KI-Einsatz in der Stichprobengenerierung fehlt derzeit. Grundlegende Fragen zu Datenqualität, Nachvollziehbarkeit, Ethik und Robustheit betreffen dabei nicht nur dieses Beispiel, sondern den KI-Einsatz im Forschungsprozess insgesamt.

Für die effiziente Unterstützung menschlicher Expertise im Forschungsprozess und vor dem Hintergrund der Replikationskrise sprechen wir uns daher für hohe methodische Sorgfalt und kritische Reflexion aus, u. a. durch (1) die kritische Überprüfung von KI-Outputs zur Identifikation fachlicher Fehler; (2) einen rigorosen methodischen Zugang, etwa durch Benchmark-Strategien (z. B. Vergleich von *Silicon* und *Human* Samples) sowie Robustheitsanalysen gegenüber analytischer Flexibilität (z. B. Short et al., 2025); (3) transparente Berichterstattung, um Evidenz zu Potenzialen und Risiken zu sammeln und KI-Standards zu entwickeln. Nur so kann KI nachhaltig im Forschungsprozess integriert werden.

Gestaltung der Mensch-KI-Interaktion

Individuelle Unterschiede bei der Nutzung von KI

Menschen unterscheiden sich in ihrer Kompetenz, KI sinnvoll einzusetzen. Diese sogenannte *AI Literacy* ist ein komplexes Konstrukt, das neben der Nutzungskompetenz auch technisches Verständnis, kritische Ergebnisbewertung und die Fähigkeit zur Reflexion von KI-Systemen umfasst (Lintner, 2024). Sie kann auch als Zusammenspiel etablierter Konstrukte verstanden werden. Eine

kompetente, ko-kreative Nutzung erfordert z. B. kognitive Fähigkeiten wie Zielpräzisierung, Kontextualisierung, iteratives Nachfragen und das Abwägen von Alternativen. Dabei spielen vermutlich digitale Vorerfahrung, domänenspezifisches Wissen, epistemische Überzeugungen und Einstellungen gegenüber KI eine zentrale Rolle.

Individuelle Unterschiede in der KI-Nutzung dürften besonders zum Vorschein kommen, wenn KI soziale Funktionen übernimmt (*Social AI*). Überblicksarbeiten zeigen, dass die Mensch-KI-Interaktion der Mensch-Mensch-Interaktion zunehmend ähnelt (Smith et al., 2025). Welche Auswirkungen KI als „sozialer Begleiter“ hat, könnte neben der KI-Gestaltung (vgl. nächster Abschnitt) mit den Merkmalen der Nutzenden zusammenhängen, wie etwa der Tendenz zur Anthropomorphisierung und dem damit verbundenen emotionalen Vertrauen in KI (Folk et al., 2025). Für wen Social AI unter welchen Bedingungen unterstützend oder problematisch wirkt, bedarf weiterer Forschung.

KI-Persönlichkeit und Personalisierung

Auch KI weist individuelle Unterschiede und eine Art „Persönlichkeit“ auf. Jedoch unterscheiden sich LLMs grundlegend vom Menschen: Sie agieren ohne neurobiologische Prozesse und primär über sprachliche Kanäle, während sich menschliche Persönlichkeit multimodal manifestiert. KI lässt sich durch Entwickelnde und Nutzende vergleichsweise leicht prägen und verändern (vgl. Abb. 1 zu Quellen von KI-Persönlichkeit nach Systemverankerung, Dauerhaftigkeit und Veränderbarkeit). Auch wenn „Persönlichkeit“ im KI-Kontext damit nicht 1:1 im menschlichen Sinne zu verstehen ist, zeigen LLMs in einer gegebenen Konfiguration relativ konsistente Interaktionsmuster. Dies führt bei Nutzenden zur Wahrnehmung einer KI-Persönlichkeit, die mehr oder weniger gezielt erzeugt werden kann (vgl. Abb. 1). Selbst ohne explizite Vorgaben zeigen LLMs unterschiedliche Persönlichkeitsprofile (Bhandari et al., 2025). Künftige Forschung sollte diese Befunde mithilfe theoretisch fundierter Taxonomien und standardisierter Prompts replizieren und die zeitliche Stabilität und situationsübergreifende Konsistenz wahrgenommener KI-Persönlichkeit untersuchen. Zudem können LLMs gezielt eingesetzt werden, um spezifische Persönlichkeitseigenschaften zu simulieren und so komplexe Persönlichkeitsdynamiken in Mensch-KI- oder KI-KI-Interaktionen, etwa in sozialen Kontexten, zu untersuchen (Hätscher et al., 2025). Die Kombination agentenbasierter Modellierung mit der Ausdruckskraft von LLMs ermöglicht es, sowohl stabile Eigenschaften als auch sich verändernde Zustände zu erkennen und Reaktionen anzupassen.

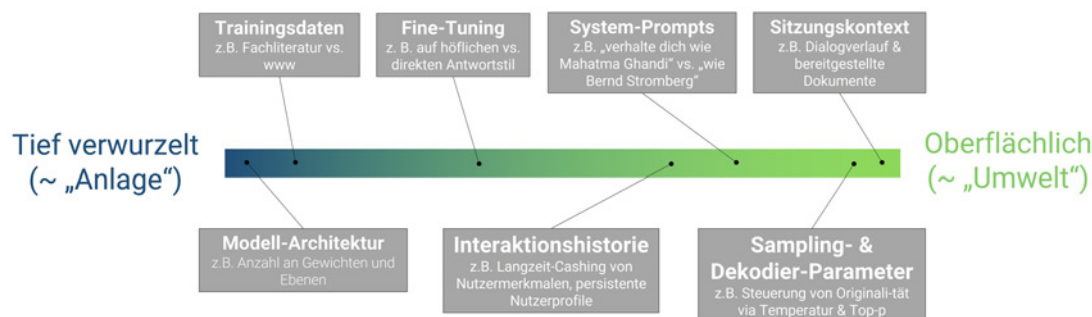


Abbildung 1. Quellen von KI-Persönlichkeit.

LLMs bieten damit interessante Ansatzpunkte für personalisierte Anwendungen: Sie können aktuelles Verhalten *erkennen*, zukünftige Entwicklungen *vorhersagen* und durch sprachbasierte Interventionen gezielt *verändern*. Eine fortlaufende Personalisierung auf Basis längerer Interaktionsverläufe dürfte die Wirksamkeit dieser Schritte erhöhen. Belastbare empirische Evidenz zu Wirksamkeit und Nebenwirkungen fehlt bislang (Stade et al., 2024).

Einordnung und Perspektiven für die Gestaltung der Mensch-KI-Interaktion

Der Einsatz von KI eröffnet neue Forschungsperspektiven zur Gestaltung der Mensch-KI-Schnittstelle. Ein zentraler Fokus sollte dabei auf den für einen verantwortungsvollen KI-Einsatz erforderlichen Fähigkeiten und deren gezielter Förderung liegen. AI Literacy gewinnt bildungspolitisch an Bedeutung und wird in der EU-KI-Verordnung (Europäische Union, 2024) als rechtliche Pflicht verankert. Um konzeptuelle Unschärfen zu vermeiden, sollten individuelle Unterschiede in der KI-Nutzung systematisch in bestehende Modelle eingeordnet und klar abgegrenzt werden; darauf aufbauend sollten psychometrisch fundierte diagnostische Verfahren entwickelt werden.

Ein weiterer wichtiger Punkt ist die nutzerzentrierte Gestaltung von KI. Unsere Fachgruppe kann hier Expertise zu KI-Persönlichkeit und deren Dynamik in Interaktion mit menschlicher Persönlichkeit einbringen, um wirksame Personalisierung etwa in therapeutischen oder arbeitsbezogenen Kontexten zu ermöglichen. Dies erfordert enge Zusammenarbeit mit anderen psychologischen Disziplinen sowie technischen Fächern, da sich die daraus resultierenden Forschungsaufgaben nur interdisziplinär adäquat bearbeiten lassen.

KI-gestützte Testkonstruktion und Diagnostik

Testkonstruktion mit Hilfe von KI

KI kann die Konstruktion von Persönlichkeits- und Leistungstests unterstützen. LLMs erleichtern die aufwendige Entwicklung von Vignetten-Items (z. B. Situation Judgment Tests, Krumm et al., 2024) und die effiziente Erstellung von Parallelformen. LLM-basierte Multi-Agenten-Systeme übernehmen gezielt verschiedene Schritte der Itemgenerierung (Lee et al., 2025). Zudem erlaubt KI die Vorhersage von Itemparametern, wodurch der Bedarf an zeitintensiven Pilotstudien reduziert wird (Sommer & Arendasy, 2025).

Ein weiterer Vorteil von KI liegt in der Entwicklung von Tests mit alternativen Antwortformaten. Insbesondere offene Formate liefern reichhaltige, individuelle Informationen, wurden bislang jedoch selten genutzt, da ihre manuelle Auswertung aufwendig ist. Im Vergleich zu anderen maschinellen Ansätzen, liefern LLMs nach gezielter Feinjustierung konsistentere und qualitativ hochwertigere Bewertungen, etwa in der Kreativitätsdiagnostik (Organisciak et al., 2023). Dies eröffnet neue Anwendungsmöglichkeiten: LLMs können die Eignungsdiagnostik ergänzen, indem sie Umweltmerkmale (z. B. berufliche Anforderungen) kodieren und so die Bestimmung von Person-Umwelt Passung unterstützen.

Automatisierte Merkmalserfassung

Mit der zunehmenden Verfügbarkeit digitaler Daten lassen sich Verhaltensweisen direkt im Alltag erfassen, wodurch klassische Indikatoren aus Persönlichkeits- und Leistungstests ergänzt oder ersetzt werden können. In der Persönlichkeitspsychologie bezeichnet *Personality Computing* Ansätze, die digitale Verhaltensspuren mit Selbst-

berichten kombinieren, um ML-Modelle zu trainieren, die Merkmale wie die Big Five später allein aus Alltagsverhalten ableiten (Phan & Rauthmann, 2021). Dieses Prinzip der automatisierten Merkmalerfassung ist auf weitere Konstrukte übertragbar, etwa Depressivität oder mentale Geschwindigkeit, und dient in Forschungsarbeiten derzeit vor allem der Exploration von Verhaltensmanifestationen. Künftig könnten LLMs hier eine zentrale Rolle spielen, indem sie latente Merkmale direkt aus digitalen Daten erschließen, ohne auf Selbst- oder Fremdberichte bzw. klassische Leistungsmessungen angewiesen zu sein.

Gütekriterien im Zeitalter von KI

In der Praxis werden ML-basierte Testverfahren bereits eingesetzt, ergänzend oder als Ersatz klassischer eignungsdiagnostischer Verfahren (Goretzko & Israel, 2022). Dabei ist die transparente Darstellung der verwendeten Indikatoren zentral. Zudem sollte KI-gestützte Diagnostik anhand klassischer Gütekriterien kritisch reflektiert werden.

Bei der Bewertung der *Objektivität* ergibt sich ein differenziertes Bild: Die Durchführungsobjektivität lässt sich durch konsequente Standardisierung weitgehend sichern. Die Auswertungsobjektivität ist hingegen durch intransparente Entscheidungsprozesse, die Vermischung von Auswertung und Interpretation sowie mögliche Verzerrungen in den Trainingsdaten gefährdet. Ähnliches gilt für die Interpretationsobjektivität. Eine nachvollziehbare, regel- und normbasierte Bewertung KI-generierter Ergebnisse bleibt bislang ungelöst. Daher ist weiterhin eine fachlich qualifizierte menschliche Beurteilung erforderlich.

Die Beurteilung der *Reliabilität* ist besonders herausfordernd. Einige KI-Modelle können bereits bei geringfügigen Änderungen der Eingaben stark unterschiedliche Ergebnisse liefern. Häufige Updates kommerzieller KI-Modelle erfordern regelmäßige Überprüfungen der Messgenauigkeit. Zudem sollten in Szenarien mit KI und menschlichen Beurteilenden (z.B. Assessment Center) Aspekte der Interrater-Reliabilität, etwa der Konsensus zwischen beiden, berücksichtigt werden (Hickman et al., 2023).

Bei der Beurteilung der *Validität* sind alle Validitätsaspekte zu berücksichtigen, auch wenn derzeit häufig die konvergente Validität im Fokus steht (d.h. Übereinstimmung zwischen KI-basierten und tatsächlichen Kriteriumswerten; Hickman et al., 2023). Zur kontinuierlichen Überprüfung der Inhaltsvalidität sollten bei KI-basierten Verfahren erklärbare Methoden sowie geeignete Auditverfahren eingesetzt werden.

Schließlich gewinnen *Nebengütekriterien* an Bedeutung. Verzerrungen in Trainingsdaten erfordern eine gesonderte Beachtung von Messinvarianz und Fairness (Goretzko & Israel, 2022). Zudem sollten KI-basierte Verfahren auf ihre Robustheit gegenüber Manipulationsversuchen geprüft werden, z. B. wenn Teilnehmende bewusst falsche Angaben machen oder versuchen KI-Outputs durch ihre Eingaben zu manipulieren. Silicon Samples, deren Vorteile und Risiken zuerst noch zu klären sind, könnten zur erweiterten Normierung beitragen.

Einordnung und Perspektiven für KI-gestützte Diagnostik

Die Möglichkeiten zur automatisierten Merkmalerfassung auf Basis digitaler Verhaltensspuren hat für die psychologische Diagnostik ein großes Potenzial. Beispielsweise könnten durch den Einsatz von KI niederschwellig kritische mentale Zustände im Alltag erkannt und frühe, maßgeschneiderte Interventionen eingeleitet werden.

Damit diese Potenziale verantwortungsvoll genutzt werden können, sind Qualitätsstandards für KI-basierte Verfahren zentral. Fragen der Messqualität sind essenziell, da diagnostische Befunde auf individueller Ebene weitreichende Konsequenzen für die Lebensgestaltung haben. Ebenso wichtig sind klare Verantwortlichkeitsstrukturen sowie die Einhaltung rechtlicher Rahmenbedingungen wie der EU-KI-Verordnung oder der DIN 33430. Wir plädieren für eine kritische Prüfung KI-gestützter Verfahren sowie für die Weiterentwicklung bestehender Gütekriterien an neue Rahmenbedingungen. Diese Forderung spiegelt sich bereits in der Aufnahme entsprechender Richtlinien in das Bewertungssystem des Diagnostik- und Testkuratoriums (2024) wider. Eine fundierte Aus- und Weiterbildung von Psycholog_innen unter Berücksichtigung bestehender und erweiterter Gütekriterien erachten wir daher als unerlässlich.

Zusammenfassung

Als Fachgruppe beschäftigen wir uns intensiv mit der Messung psychologischer Merkmale und mit der Erforschung individueller Unterschiede. Diese Schwerpunkte spiegeln sich in unserer Perspektive auf KI wider.

Erstens sind Messfehler in der psychologischen Messung unvermeidbar und müssen dementsprechend systematisch berücksichtigt werden. Dieses Grundverständnis sollte auch für den Einsatz von KI gelten. KI eröffnet viele neue Potenziale für Forschung und Diagnostik, kann aber zu Fehlern führen. So wie wir einen neu konstruierten

Test nicht ohne Validierung unkritisch einsetzen, muss auch der Einsatz von KI sorgfältig geprüft und anhand unseres Expert_innenwissens validiert werden.

Zweitens denken wir in individuellen Unterschieden. Die Mensch-KI-Schnittstelle ist kein „One-Size-Fits-All“-Thema; sowohl individuelle Unterschiede zwischen Nutzenden als auch KI-Systemen müssen berücksichtigt werden. Als Fachgruppe können wir wesentlich dazu beitragen, Technik dahingehend verantwortungsvoll zu gestalten, insbesondere in interdisziplinärer Kooperation mit anderen psychologischen Teildisziplinen und Technikwissenschaften.

Der Begriff „Künstliche Intelligenz“ verweist auf eine Metapher für technische Systeme, die versuchen, menschliche intelligente Leistungen nachzuahmen. Wie in diesem Beitrag dargestellt, imitieren aktuelle KI-Anwendungen einzelne, klar abgegrenzte Teilleistungen. Zukünftig könnte technologischer Fortschritt zur Entwicklung zunehmend „stärkerer“ KI führen, d.h. Systemen mit umfassenderen, auf menschlichem Niveau angesiedelten Intelligenzleistungen (Sheikh et al., 2023). Diese langfristige Perspektive ist mit zahlreichen offenen Fragen verbunden und sollte nicht allein Technologieakteuren überlassen werden. Für die verantwortungsvolle Gestaltung des sensiblen Mensch-KI-Gefüges sind psychologische Perspektiven unverzichtbar.

Literatur

- Bhandari, P., Naseem, U., Datta, A., Fay, N. & Nasim, M. (2025). Evaluating Personality Traits in Large Language Models: Insights from Psychological Questionnaires. *Companion Proceedings of the ACM on Web Conference 2025*, 868–872. <https://doi.org/10.1145/3701716.3715504>
- Cui, Z., Li, N. & Zhou, H. (2025). A large-scale replication of scenario-based experiments in psychology and management using large language models. *Nature Computational Science*, 5(8), 627–634. <https://doi.org/10.1038/s43588-025-00840-7>
- Cummins, J. (2025). The threat of analytic flexibility in using large language models to simulate human data: A call to attention. *arXiv*. <https://doi.org/10.48550/ARXIV.2509.13397>
- Diagnostik- und Testkuratorium. (2024). Föderation Deutscher Psychologenvereinigungen: Qualitätssicherung- und -optimierung von Tests: Die vierte Fassung des Testbeurteilungssystems des Diagnostik- und Testkuratoriums. *Psychologische Rundschau*, 75(1), 49–53. <https://doi.org/10.1026/0033-3042/a000657>
- Europäische Union. (2024). *Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (KI-Verordnung)*. <http://data.europa.eu/eli/reg/2024/1689/oj>
- Folk, D., Heine, S. J. & Dunn, E. (2025). Individual differences in anthropomorphism help explain social connection to AI companions. *Scientific Reports*, 15(1), 36548. <https://doi.org/10.1038/s41598-025-19212-2>
- Goretzko, D. & Israel, L. F. (2022). Pitfalls of machine learning-based personnel selection: Fairness, transparency, and data quality. *Journal of Personnel Psychology*, 21(1), 37–47. <https://doi.org/10.1027/1866-5888/a000287>
- Hätscher, O., Pielage, L., Kockwelp, P., Breil, S. M., Junga, A., Dabel, J. et al. (2025). LLM-Based Agent Simulations for Modeling Individual Differences in Social Interactions (LASSI). *PsyArXiv*. https://doi.org/10.31234/osf.io/xrj3c_v1
- Hickman, L., Herde, C. N., Lievens, F. & Tay, L. (2023). Automatic scoring of speeded interpersonal assessment center exercises via machine learning: Initial psychometric evidence and practical guidelines. *International Journal of Selection and Assessment*, 31(2), 225–239. <https://doi.org/10.1111/ijsa.12418>
- Hommel, B. E. & Arslan, R. C. (2025). Language Models Accurately Infer Correlations Between Psychological Items and Scales From Text Alone. *Advances in Methods and Practices in Psychological Science*, 8(4), 25152459251377093. <https://doi.org/10.1177/25152459251377093>
- Krumm, S., Thiel, A. M., Reznik, N., Freudenstein, J.-P., Schäpers, P. & Mussel, P. (2024). Creating a psychological test in a few seconds: Can ChatGPT develop a psychometrically sound situational judgment test? *European Journal of Psychological Assessment*. [Advance online publication]. <https://doi.org/10.1027/1015-5759/a000878>
- Lee, P., Son, M. & Jia, Z. (2025). AI-powered Automatic Item Generation for Psychological Tests: A Conceptual Framework for an LLM-based Multi-Agent AIG System. *Journal of Business and Psychology*, 41, 71–99. <https://doi.org/10.1007/s10869-025-10067-y>
- Lintner, T. (2024). A systematic review of AI literacy scales. *Npj Science of Learning*, 9(1), 50. <https://doi.org/10.1038/s41539-024-00264-4>
- Organisciak, P., Acar, S., Dumas, D. & Berthiaume, K. (2023). Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity*, 49, 101356. <https://doi.org/10.1016/j.tsc.2023.101356>
- Phan, L. V. & Rauthmann, J. F. (2021). Personality computing: New frontiers in personality assessment. *Social and personality psychology compass*, 15(7), e12624.
- Schoedel, R., Sust, L., Sterner, P. & Goretzko, D. (2026). From Digital Data to Psychological Insights: Making Sense of Mobile-Sensing Data through Integrative Preprocessing Pipelines. *Psychometrika*. [Online Publication:1–28]. <https://doi.org/10.1017/psy.2026.10083>
- Short, C. A., Hildebrandt, A., Bosse, R., Debener, S., Özyağcılar, M., Paul, K. et al. (2025). Lost in a large EEG multiverse? Comparing sampling approaches for representative pipeline selection. *Journal of Neuroscience Methods*, 424, 110564. <https://doi.org/10.1016/j.jneumeth.2025.110564>
- Sheikh, H., Prins, C. & Schrijvers, E. (2023). *Artificial Intelligence: Definition and Background*. In H. Sheikh, C. Prins, & E. Schrijvers, *Mission AI* (pp. 15–41). Cham: Springer. https://doi.org/10.1007/978-3-031-21448-6_2
- Smith, M. G., Bradbury, T. N. & Karney, B. R. (2025). Can Generative AI Chatbots Emulate Human Connection? A Relationship Science Perspective. *Perspectives on Psychological Science*, 20(6), 1081–1099. <https://doi.org/10.1177/17456916251351306>
- Sommer, M. & Arendasy, M. (2025). Automatic- and Transformer-Based Automatic Item Generation: A Critical Review. *Journal of Intelligence*, 13(8), 102. <https://doi.org/10.3390/jintelligen13080102>
- Stade, E. C., Stirman, S. W., Ungar, L. H., Boland, C. L., Schwartz, H. A., Yaden, D. B. et al. (2024). Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. *Npj Mental Health Research*, 3(1), 12. <https://doi.org/10.1038/s44184-024-00056-z>

Wulff, D. U. & Mata, R. (2025a). Escaping the Jingle-Jangle Jungle: Increasing Conceptual Clarity in Psychology Using Large Language Models. *Current Directions in Psychological Science*, 35, 59–65. <https://doi.org/10.1177/09637214251382083>

Wulff, D. U. & Mata, R. (2025b). Semantic embeddings reveal and address taxonomic incommensurability in psychological measurement. *Nature Human Behaviour*, 9(5), 944–954. <https://doi.org/10.1038/s41562-024-02089-y>

Onlineveröffentlichung: 21.07.2026

Ethische Richtlinien

Zur Erstellung dieser Arbeit nutzten die Autoren ChatGPT (GPT-5 mini, OpenAI, <https://chat.openai.com/>) zur sprachlichen Kürzung und Formulierungsoptimierung. Nach der Nutzung dieses Werkzeugs/Dienstes überprüften und bearbeiteten die Autoren den Inhalt nach Bedarf und übernehmen die volle Verantwortung für den Inhalt der Veröffentlichung

Open Science

Open Data: Eine englische Übersetzung dieses Artikels ist als Preprint unter PsyArXiv verfügbar unter https://doi.org/10.31234/osf.io/j8szh_v2.

Förderung

Open Access-Veröffentlichung ermöglicht durch die Hochschule Fresenius.

Prof. Dr. Ramona Schödel

Charlotte Fresenius Hochschule
Infanteriestr. 11a
80797 München
Deutschland
ramona.schoedel@charlotte-fresenius-uni.de