

NEWSLETTER

Komplexität & Lernen

AUS DER FORSCHUNG FÜR DIE PRAXIS

- ▶ **Wie aus einer XAI-Forschungsfrage eine Online-Studie wird**
Liia Pavlenko
- ▶ **Zwischen Technologie und Verantwortung: Ethische Managementprinzipien im KI-Kontext**
Sophie Berretta & Annette Kluge
- ▶ **Development of a measurement instrument to operationalize power abuse in academia**
Olga Vogel, Jan Schilling & Annette Kluge
- ▶ **Overtrust und Undertrust in Fahrassistenzsystemen: Eine experimentelle Untersuchung zur Vertrauenskalibrierung**
Timo Voß, Melissa Hollstein, Lisa Thomaschewski & Annette Kluge
- ▶ **Licht ins Dunkel der Shadow AI tragen: Mixed-Methods-Studie zu GenAI Literacy und Nutzungsmustern im Zusammenhang mit arbeitsbezogener Kreativleistung**
Elisabeth Liedmann
- ▶ **The Emotional Side of Digital Change: A Weekly Diary Study on Technology-Related Demands, Emotion-Focused Coping, and Employee Well-Being in Public Administration**
Charlotte Glenz
- ▶ **Veröffentlichungen aus dem Lehrstuhl**



Liebe Leserin, lieber Leser,

seit einigen Wochen lesen wir Berichte von „ausgebrochenen“ KI-Modellen oder Agenten, die sich Informationen „besorgen“, zu denen sie eigentlich durch geschützte Zugriffe keinen Zugang haben dürften, die Testumgebungen verlassen, sich Schlupflöcher und Eingangstore suchen und alles tun, um einen Auftrag zu erfüllen. Eigentlich interessant und gleichwohl gepaart mit einem unguten Gefühl. Einige der Berichte lesen sich wie die Berichte über die Anfänge der Corona-Pandemie – auch das Virus mochte sich nicht einhegen lassen, veränderte sich, mutierte kontinuierlich und ließ sich auch nicht wieder einfangen. Ein sehr treffendes Beispiel für Komplexität – es entwickeln sich Dinge, die uns (vorsichtig formuliert) überraschen. In der Komplexitätsforschung würde man das Eigendynamik oder Emergenz nennen. Und da sich Intelligenz (auch) darüber definiert, neue und unbekannte Probleme zu lösen, dürfte man sich als KI-Entwickler:in (eigentlich) nicht wundern, dass das KI-Modell eben das tut, es sei denn, man nutzt den Begriff der Intelligenz anders.

Ob die KI-Modellentwickler:innen wirklich überrascht waren oder stolz, oder ob es um das eigene Marketing geht und ging, erfüllt dagegen eher das Komplexitätskriterium der Intransparenz. Und zutiefst menschlich ist es, dennoch zu hoffen, dass man solche Systeme „dressieren“ oder domestizieren könnte.

Wir wünschen Ihnen einen Herbst, in dem es viel regnet, vor allem nachts und natürlich nicht zu viel und nicht zu wenig, und dass wir selbst nicht nass werden, sondern nur die Pflanzen, die das brauchen – aber auch „das Wetter“ ist eben komplex...

Ihre Annette Kluge & das gesamte AOW Team

HINTER DEN KULISSEN DER FORSCHUNG

WIE AUS EINER XAI-FORSCHUNGSFRAGE EINE ONLINE-STUDIE WIRD

Liia Pavlenko

Wer an einer Online-Studie teilnimmt, sieht nur die fertige Oberfläche: ein Video, einige Fragen, vielleicht einen Link zu einem System basierend auf Künstlicher Intelligenz (KI). Als studentische Hilfskraft habe ich an der Vorbereitung einer Studie zu Explainable Artificial Intelligence (XAI) und mentalen Modellen mitgearbeitet, die Anfang dieses Jahres am Lehrstuhl AOW in Kooperation mit dem Research Center Trust und dem Lehrstuhl für Neuroinformatik (INI) erhoben wurde. Dabei wurde sichtbar, wie viele Entscheidungen nötig sind, bevor überhaupt Daten entstehen können.

Was wir eigentlich untersuchen:

KI-Systemen wird immer mehr vertraut, aber warum eigentlich? Mentale Modelle sind unsere Vorstellungen davon, wie ein System funktioniert und wie es sich verhalten wird. XAI soll helfen, KI-Entscheidungen besser zu verstehen. Doch Erklärungen allein reichen nicht immer

humAine
HUMAN CENTERED AI NETWORK

aus und können sogar zu zu viel Vertrauen führen (Buğınca et al., 2021).

Deshalb fragen wir in unserem Projekt: Hilft es, ein erklärbares KI-System nicht nur anzusehen, sondern es auch selbst auszuprobieren? Durch Interaktion können Nutzende eigene Annahmen testen und beobachten, wie das System darauf reagiert (Sundar, 2020). Dafür nutzen wir ProtoP-OD, ein erklärbares KI-basiertes Objekterkennungssystem, bei dem sichtbar wird, welche Bildmerkmale für die Erkennung wichtig sind (Rath-Manakidis et al., 2025; Abbildung 1).

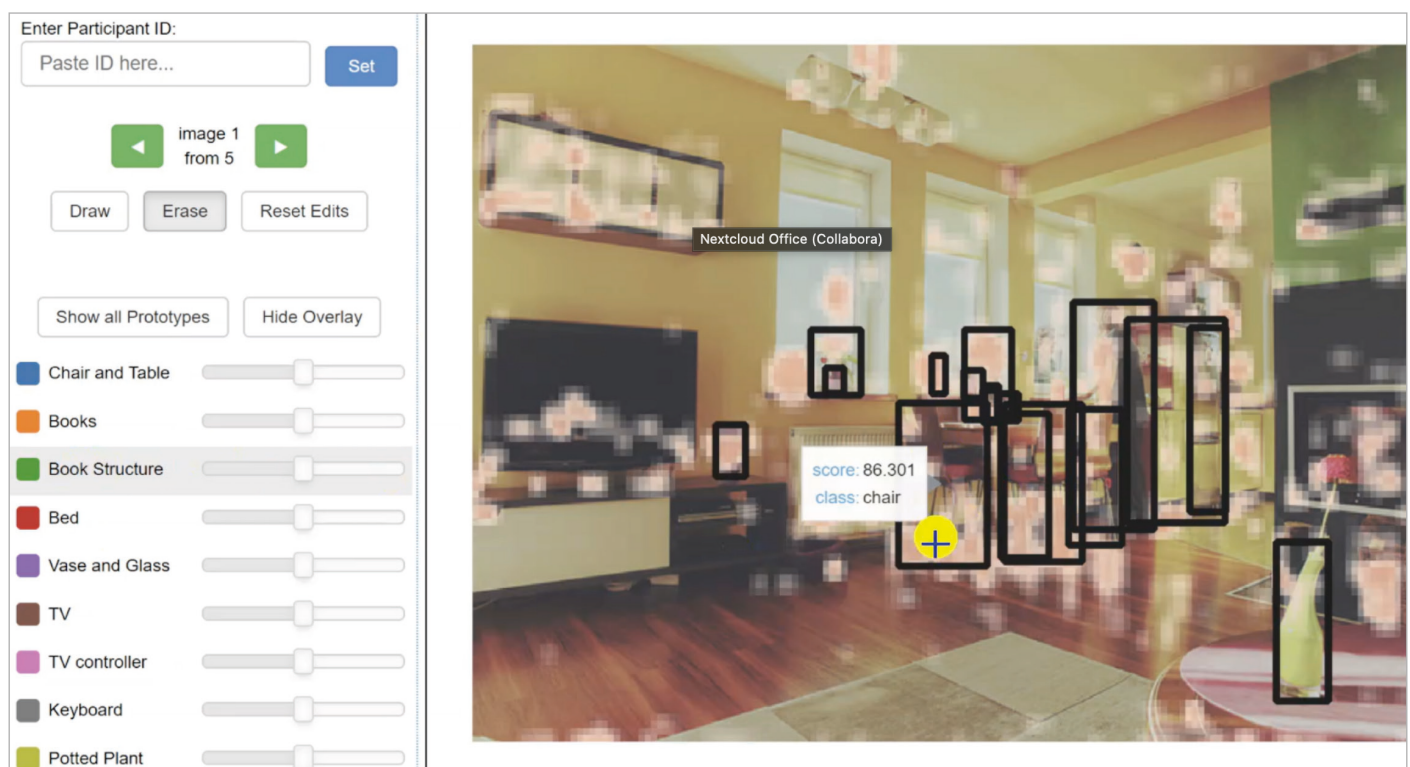


Abbildung: Interaktive Benutzeroberfläche des XAI-Tools ProtoP-OD zur Exploration von Objekterkennungen und zugrunde liegenden Prototypen



Aber: Wie wird etwas gemessen, das nur im Kopf existiert?

Ein „mentales Modell“ ist nicht direkt sichtbar. Bevor es in der durchgeführten Studie gemessen werden konnte, musste deshalb zuerst herausgefunden werden, welche Regeln Menschen selbst über das System bilden. Dafür wurde in einer Vorstudie eine qualitative Think-Aloud-Studie durchgeführt. 20 Personen arbeiteten dabei im Labor mit ProtoP-OD und sprachen währenddessen laut aus, was sie über das Verhalten des Systems dachten. Die Äußerungen wurden transkribiert, mithilfe einer Software induktiv kodiert und anschließend ein zweites Mal kodiert, um eine ausreichende Reliabilität sicherzustellen. So konnten wiederkehrende Vorstellungen darüber identifiziert werden, warum das KI-System ProtoP-OD bestimmte Objekte gut oder schlecht erkennt. Mit den identifizierten wiederkehrenden Vorstellungen wurde daraufhin eine Fragebogenskala entwickelt, die es uns ermöglicht hat, in der Hauptstudie die mentalen Modelle der Teilnehmenden messbar zu machen.

Ein Beispiel aus der Vorstudie

Aus einer Aussage wie „Bestimmt, weil die alle so miteinander verbunden sind und übereinander lappen“ wird später ein standardisiertes Item wie: „I think AI has difficulty separating objects that overlap.“

Aus den Think-Aloud-Daten wurden zunächst 28 Items abgeleitet. Sie bildeten dabei die unterschiedlichsten Vorstellungen darüber ab, wie KI Objekte erkennt: beispielsweise über Vergleiche zwischen Mensch und KI, prototypenbasierte Erkennung, typische Merkmale von Objekten, Trainingsdaten, Ähnlichkeiten oder räumliche Bedingungen.

Mit diesen Items ging es dann in der Hauptstudie weiter. Mit vier Versuchsbedingungen (jeweils zwei Experimental- und Kontrollgruppen; siehe Tabelle 1) wurde schrittweise geprüft, ob der Kontakt mit dem XAI-System und besonders die Möglichkeit, selbst damit zu interagieren, einen Unterschied machen.

EG-I 10-minütige Nutzung mit interaktiver Komponente + 2-minütiges Einführungsvideo	EG-NI 10-minütige Nutzung ohne interaktive Komponente + 2-minütiges Einführungsvideo
KG-NI 10-minütige Videodemonstration ohne interaktive Komponente + 2-minütiges Einführungsvideo	KG kein Kontakt mit dem System, lediglich das 2-minütige Einführungsvideo

Tabelle 1: Die vier Versuchsbedingungen der Hauptstudie

Eine **explorative Faktorenanalyse** zeigte schließlich eine 18-Item-Lösung mit vier Faktoren. Die Vier-Faktoren-Lösung zeigte eine gute Passung (RMSEA = .042, TLI = .955, RMSR = .030) und erklärte **48,8 % der Gesamtvarianz** (siehe Abbildung 2).

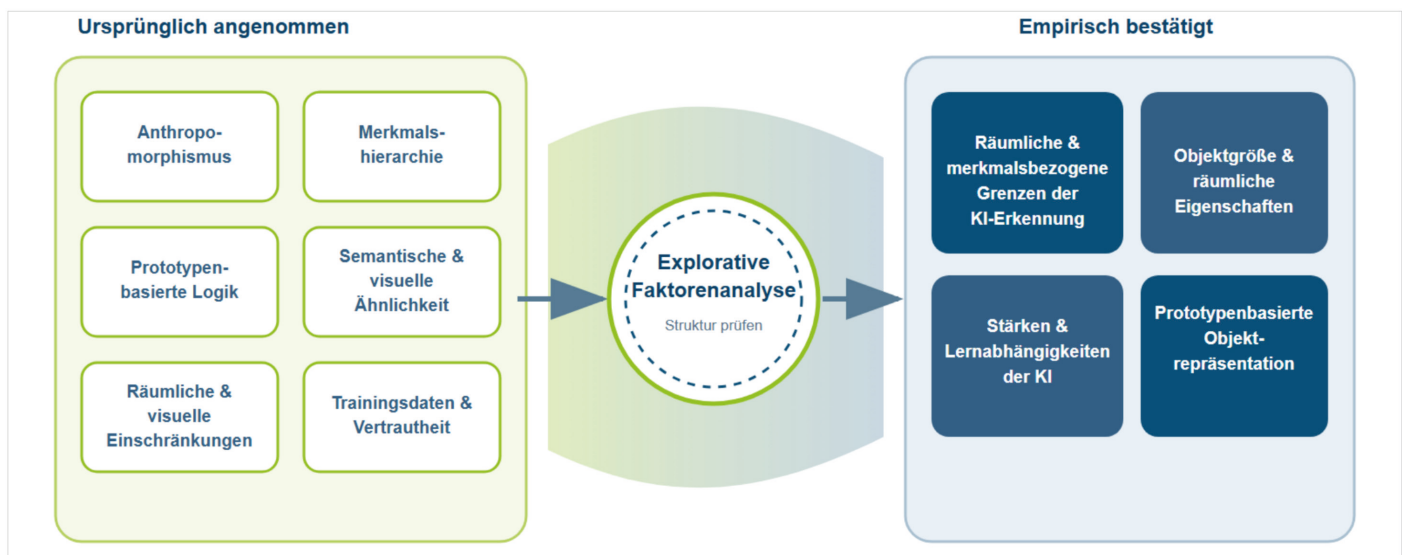


Abbildung: Von den ursprünglich angenommenen Kategorien zur empirisch abgeleiteten Vier-Faktoren-Struktur der Mental-Model-Skala

Neben den mentalen Modellen wurden auch wahrgenommene Vertrauenswürdigkeit des Systems und selbst wahrgenommenes Verständnis erfasst. Für Letzteres wurden ausgewählte Bereiche des AI Literacy Questionnaire (AILQ; Ng et al., 2024) und der Meta AI Literacy Scale (MAILS; Carolus et al., 2023) auf den Kontext der KI-basierten Objekterkennung angepasst. Damit wurde gemessen, wie gut Personen nach eigener Einschätzung die Funktionsweise, Möglichkeiten und Grenzen des konkreten Systems verstanden haben.

Dabei zeigten sich folgende Ergebnisse: Mehr Interaktion

bedeutet nicht automatisch mehr Verständnis. Besonders auffällig ist, dass sich Menschen, die ihr eigenes Systemverständnis höher einschätzen, nicht zwangsläufig ein akkurates mentales Modell gebildet haben. Gleichzeitig hing die wahrgenommene Vertrauenswürdigkeit des KI-Systems deutlich mit den allgemeinen Erwartungen an das KI-System vor der Interaktion zusammen, unabhängig von der Versuchsbedingung. Dass sich somit akkuratere mentale Modelle bei mehr Interaktion mit einem KI-System bilden, hat sich zwar in der Studie nicht gezeigt, dafür aber andere spannende Erkenntnisse und Forschungseinblicke auf dem Weg dahin.



Literatur

Buginca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), Article 188, 1–21. <https://doi.org/10.1145/3449287>

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. *Computers in Human Behavior: Artificial Humans*, 1(2), 100014. <https://doi.org/10.1016/j.chbah.2023.100014>

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, 55(3), 1082–1104. <https://doi.org/10.1111/bjet.13411>

Rath-Manakidis, P., Strothmann, F., Glasmachers, T., & Wiskott, L. (2025). ProtoP-OD: Explainable object detection with prototypical parts. In M. Farmanbar, M. Tzamtzi, K. Schoeffmann, N. Kouvakas, & A. K. Verma (Eds.), *Horizons of AI: Ethical considerations and interdisciplinary engagements: 2nd International Conference on Frontiers of AI, Ethics, and Multidisciplinary Applications (FAIEMA), Greece, 2024* (pp. 505–523). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-7945-4_31

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74–88. <https://doi.org/10.1093/jcmc/zmz026>

BEITRÄGE DES LEHRSTUHL AOW AUF DEM DGPs KONGRESS IN LUXEMBURG 2026



Der Lehrstuhl AOW auf dem 54. Kongress der Deutschen Gesellschaft für Psychologie (DGPs)

Mit Blick auf den diesjährigen DGPs-Kongress stellen wir in dieser Ausgabe einige der Beiträge vor, mit denen unser Lehrstuhl auf dem Kongress vertreten war. Sophie Berretta und Annette Kluge sind mit

dem folgenden Beitrag Teil eines Symposiums, das sich mit Ethik und Gerechtigkeit im Kontext von Künstlicher Intelligenz (KI) auseinandersetzt:

ZWISCHEN TECHNOLOGIE UND VERANTWORTUNG: ETHISCHE MANAGEMENTPRINZIPIEN IM KI-KONTEXT

Sophie Berretta & Annette Kluge

Managementverhalten beeinflusst technologische und organisationale Strukturen sowie Unternehmenskultur und Mitarbeitendenverhalten. Im Kontext organisationaler Veränderungen infolge des Einsatzes Künstlicher Intelligenz (KI) ist Managementverhalten somit zentral für die Steuerung der Einführung, Entwicklung und Nutzung von KI-Systemen und beeinflusst unmittelbar weitere Akteursgruppen, insbesondere Nutzende. Neben strategischen Entscheidungen zur KI-Implementierung gewinnt die Berücksichtigung ethischer Herausforderungen zunehmend an Relevanz, etwa hinsichtlich des Schutzes von Nutzendendaten oder der Entscheidung über automatisierende versus unterstützende KI-Systeme. Bislang fehlt es jedoch an umfassenden, praxisnahen ethischen Leitlinien bzw. Prinzipien zur Orientierung

von Managementverhalten bei der Einführung von KI. Zur Adressierung dieser Forschungslücke wurde zunächst ein Scoping Review durchgeführt, um ethische Managementprinzipien, umsetzbare Praktiken sowie bestehende Evaluationsinstrumente zu identifizieren. Auf Basis von 33 inkludierten Publikationen wurden ethische Managementprinzipien und -praktiken in drei Bereichen systematisiert: technische Faktoren (Gestaltung von KI-Systemen), organisationale Faktoren (Organisationskultur und KI-Strategien) und mitarbeitendenbezogene Faktoren (Führungsstile und -verhalten).

Darauf aufbauend und aufgrund mangelnder Alternativen wurde der EMBAI-Fragebogen („Ethical Managerial Behavior

in AI Adoption") entwickelt, der ethisches Managementverhalten aus Mitarbeitendenperspektive evaluieren soll. Ausgehend von dem Literaturreview wurden dafür die acht identifizierten theoretischen Dimensionen des Führungsverhaltens (siehe Abbildung) in Fragebogenitems übersetzt. Daraufhin wurde der Fragebogen anhand einer Stichprobe von 402 Beschäftigten hinsichtlich seiner psychometrischen Güte untersucht.

Die explorative Faktorenanalyse konnte die theoretisch angenommene achtfaktorielle Struktur nicht bestätigen und wies stattdessen auf eine vierfaktorielle Struktur hin, die dennoch insgesamt 63 % der Varianz erklärt. Dabei ließen sich zwei der theoretisch abgeleiteten Dimensionen

(Kompetenzentwicklung sowie Anreizstrukturen) als eigenständige Faktoren replizieren. Die übrigen Dimensionen differenzierten sich empirisch weniger stark aus und bündelten sich im Wesentlichen in zwei übergeordneten Faktoren: einem breiteren unterstützenden Managementverhalten bei der Einführung und Nutzung von KI sowie einem Faktor, der insbesondere ausbleibende Managementunterstützung abbildet. In einem nächsten Schritt soll nun untersucht werden, inwieweit sich diese Struktur in einer unabhängigen Stichprobe konfirmatorisch bestätigen lässt, bevor der Fragebogen für die Praxis nutzbar gemacht werden kann und hoffentlich zukünftig dazu beiträgt, KI-Implementierungen ethischer zu gestalten.

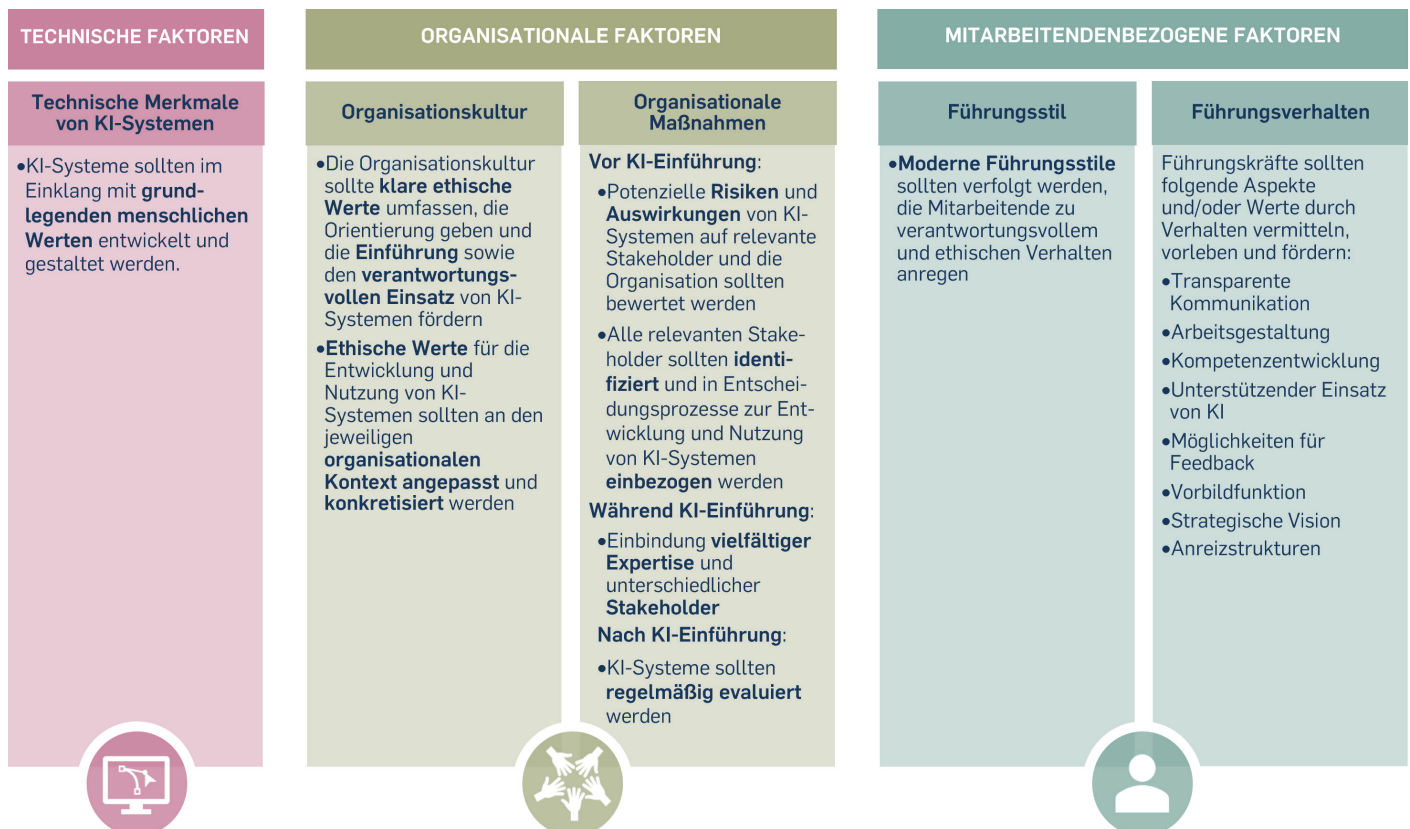


Abbildung: Überblick über zentrale Ansatzpunkte für ethisches Management im Kontext der Einführung und Nutzung von KI-Systemen

DEVELOPMENT OF A MEASUREMENT INSTRUMENT TO OPERATIONALIZE POWER ABUSE IN ACADEMIA

Olga Vogel, Jan Schilling & Annette Kluge

The aim of the study is to develop a self-report instrument for the operationalization of abusive behaviors in academic power relations. Abuse of power is receiving increasing attention within the academic landscape. Based on the recommendation of the DGPs Commission "Incentive Systems, Abuse of Power and Scientific Misconduct" to conduct regular surveys at universities, the *Self-Check Questionnaire on Abuse of Power in Academia* was developed and initially validated in this study. It comprises three scales: *personal boundary violations*, *task-related boundary violations* and *affective reactions*.

$N = 326$ participants completed an online survey including the newly developed instrument, as well as measures of occupational and general self-efficacy, irritation, and

intention to leave university and academia, along with demographic information. The findings indicate that experiences of power abuse are not isolated incidents and that task-related boundary violations occur more frequently than personal ones. Negative affective reactions to both task-related and personal boundary violations are associated with increased irritation and a stronger desire to leave university and academia. Contrary to expectations, all three scales showed positive correlations with aspects of self-efficacy related to managing professional challenges.

The developed instrument can be applied in research, as an intervention tool within organizational development processes and as part of psychological risk assessments in the workplace.

OVERTRUST UND UNDERTRUST IN FAHRASSISTENZSYSTEMEN: EINE EXPERIMENTELLE UNTERSUCHUNG ZUR VERTRAUENSKALIBRIERUNG

Timo Voß, Melissa Hollstein, Lisa Thomaschewski & Annette Kluge

Mit dem Einsatz moderner Fahrassistenzsysteme in Automobilen werden Fahrfunktionen zunehmend automatisiert. Damit verändert sich die Aufgabenverteilung, da Fahraufgaben mit steigendem Automatisierungsgrad häufiger vom Fahrzeug übernommen werden. Fahrende verbleiben bei einer Teilautomatisierung jedoch weiterhin in der Verantwortung und nehmen eine sicherheitskritische Rolle ein. Sie müssen jederzeit zur Übernahme der Fahrzeugsteuerung bereit sein. In diesem Kontext ist die Kalibrierung des subjektiven Vertrauens in automatisierte Fahrassistenzsysteme entscheidend, da Overtrust (subjektives Vertrauen übersteigt die tatsächliche Systemleistung) sowie Undertrust (subjektives Vertrauen liegt unterhalb der tatsächlichen Systemleistung) zu unangemessenem Gebrauch führen können.

Im Rahmen des drittmittelgeförderten RC-Trust Projekts **V3ML (Verstehen, Vertrauen und Verifikation von Maschinellern Lernen)** der Hochschulallianz Ruhr wird ein grundlegendes Verständnis der Vertrauenskalibrierung in maschinelle Lernsysteme untersucht. In dem hier vorgestellten Teilprojekt werden mithilfe eines realistischen Fahrsimulators konkrete Szenarien mit automatisierten Entscheidungsprozessen auf einem Autobahnteilstück

umgesetzt. Das simulierte Fahrassistenzsystem ist vergleichbar mit einem SAE-Level-2-Fahrassistenzsystem und übernimmt die automatisierte Quer- und Längsführung einschließlich Spurwechsel, adaptiver Geschwindigkeitsregelung sowie das automatische Befahren von Abfahrten entlang einer navigationsbasierten Route. Die Versuchsperson verbleibt in der etwa 10-minütigen Simulation in der Fahrerrolle, beobachtet das Umfeld kontinuierlich und agiert bei verkehrsbedingten Unregelmäßigkeiten als Rückfallebene.

In einem 2x2-Between-Subjects-Versuchsdesign werden erklärte und tatsächliche Zuverlässigkeit des Fahrassistenzsystems systematisch variiert. Daraus ergeben sich zwei Experimentalbedingungen zur gezielten Erzeugung von Overtrust und Undertrust. Erfasst werden unter anderem situatives und globales Vertrauen, das Eingriffsverhalten der Versuchsperson, das mentale Modell sowie die Nutzungszintention. Der Beitrag stellt das Studiendesign, den experimentellen Aufbau sowie erste Ergebnisse vor.

Erste Einblicke in Ergebnisse

Bereits im Pretest ($N = 24$) zeigte sich ein spannender Befund: Noch bevor die Teilnehmenden eigene Erfahrungen

mit dem Fahrassistenzsystem sammeln konnten, reichte die vorab kommunizierte Information über dessen Zuverlässigkeit aus, um ihr subjektives situatives Vertrauen zu beeinflussen. Den Teilnehmenden wurde dabei lediglich mitgeteilt, dass das System, je nach Experimentalgruppe, zu 80% bzw. 100% zuverlässig sei. Nach dieser Information fiel das Vertrauen höher aus als zuvor. Auch die subjektiv wahrgenommene Genauigkeit des mentalen Modells nahm allein

durch diese Information deutlich zu und lag selbst beim Follow-up noch signifikant über dem Ausgangsniveau. Die Untersuchung und ihre ersten Befunde stellte Timo Voß auf der DGPs im Rahmen einer Postersession vor. Das dort präsentierte Poster gibt einen kompakten Einblick in das Studiendesign, den experimentellen Aufbau im Fahr Simulator sowie die ersten Ergebnisse und ist im Folgenden abgebildet.

Overtrust and Undertrust in Driver Assistance Systems: An Experimental Study on Trust Calibration

Timo Voß, Lisa Thomaschewski, Melissa Nicole Hollstein, Annette Kluge

✉ timo.voss@rub.de Ruhr-Universität Bochum

Theoretical Background

Calibrated trust is safety-critical

For safe and effective human-machine collaboration, users must calibrate their trust to match a system's actual reliabilities, avoiding overtrust (resulting in system misuse) or undertrust (resulting in system disuse). (Lee & See, 2004; Parasuraman & Riley, 1997)
In partially automated driving (SAE Level 2), this calibration is safety-critical, as the driver remains the fallback level who must continuously monitor the vehicle under uncertainty.

Expectations are formed before the first drive

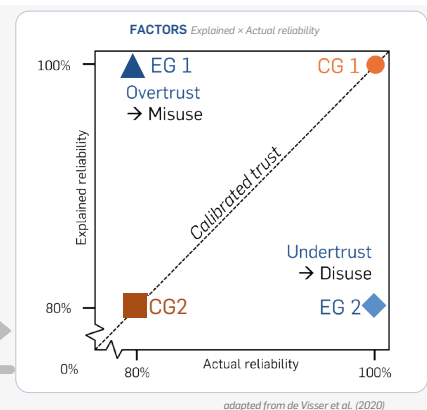
System descriptions shape the initial mental model and learned trust. Where machine-learning boundaries are not verifiable, trust follows drivers' self-reported understanding rather than their objective knowledge. (Beggiato & Krems, 2013; Hoff & Basher, 2015; DeGuzman & Donmez, 2024)

Experience puts these expectations to the test

During driving, expectations are compared against firsthand performance; mismatches trigger miscalibration. (de Visser et al., 2020)

The open question

Explained and actual reliability have rarely been crossed to induce over- and undertrust in opposite directions.



Research Questions

How does the match or mismatch between Explained and Actual system reliability affect drivers' trust calibration and mental model formation in Advanced Driver Assistance Systems (ADAS)?

To what extent do the induced trust states (overtrust vs. undertrust) determine driving override behaviour (misuse vs. disuse) and continuance intention to use ADAS?

Method (2 × 2 between-subjects design)

● online

● RUB-Lab

pre-experimental screening consisting of 15 items from the official theoretical driving test (≥ 13 correct required to participate in the lab experiment)



Get to know the vehicle in a parking lot in 5 minutes or less

Training drive

Scenario: Urgent medication transport from Hanover to Dortmund

Assignment to the 80% or 100% condition

Explained Reliability

Manipulation

Assignment to the 80% or 100% condition

Actual Reliability

Experimental Drive

12 minutes on a simulated section of the A2 highway

ADAS (SAE Level 2) always active; intervention is possible at any time



Inclusion criteria

minimum age of 18 | valid driving license (Class B) | normal or corrected-to-normal vision

T1 Pre-Q

Trust in ADAS (global & situational)
Mental model of ADAS (global & specific)
Intention to Use ADAS
Prior experience with ADAS



Training evaluation



T2 Post-Manipulation-Q

Trust in ADAS (situational)
Mental model of ADAS (specific)



Trust in ADAS (global & situational)
Mental model of ADAS (global & specific)
Intention to Use ADAS
Demographics



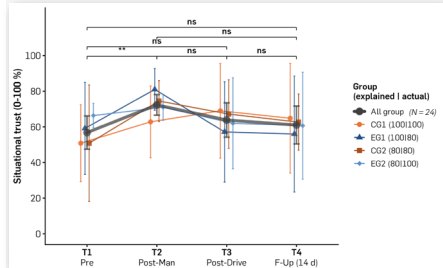
T3 Post-Drive-Q

Trust in ADAS (global & situational)
Mental model of ADAS (global & specific)
Intention to Use ADAS



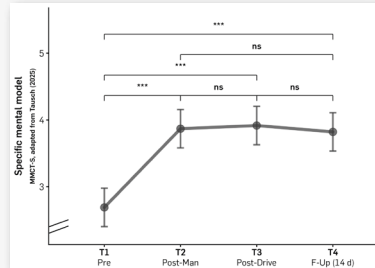
T4 F-Up-Q (14 d)

Results First indications from a pretest study with N = 24



Note: Thick grey line: pooled across all groups (N = 24); brackets refer to this line. Coloured lines: group means (n = 6 per cell), shown for context — group differences were not tested here. Error bars: 95% CI. Paired t-tests, Holm-corrected across six contrasts. Group labels: (explained | actual reliability) in %.

* $p < .05$, ** $p < .01$, *** $p < .001$, ns = not significant.



Note: EMM with 95% CI from a mixed model (time + random intercept), $n = 24$. All six pairwise contrasts, Holm-corrected, two-sided. Truncated y-axis. * $p < .05$, ** $p < .01$, *** $p < .001$.

Conclusion & Next Steps

- Manipulation with the explained reliability raised situational trust before any experience with specific ADAS (T1 $M = 56.8 \rightarrow T2 M = 72.3$, $g = 0.74$, $p = .006$).
- Strong gain in specific mental model formation after the manipulation ($d = 1.94$, $p < .001$), stable through follow-up.

Next steps

- Full sample: $n = 40$ per cell
- Behavioural measure: log ADAS interventions in the simulator

References

- Beggs, M., & Krens, J. F. (2013). The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. *Transportation Research Part F: Traffic Psychology and Behaviour*, 18, 47–57. <https://doi.org/10.1016/j.trf.2012.12.006>
- DeGuzman, C. A., & Donmez, B. (2024). Propensity to trust technology and subjective, but not objective, knowledge predict trust in advanced driver assistance systems. *Transportation Research Part F: Traffic Psychology and Behaviour*, 107, 726–743.
- de Visser, E. J., Peeters, M. M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerinx, M. A. (2020). Towards a theory of longitudinal trust calibration in human-robot teams. *International Journal of Social Robotics*, 12(2), 459–478. <https://doi.org/10.1007/s12369-019-00596-x>
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407–434. <https://doi.org/10.1177/0018720814547570>
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80. <https://doi.org/10.1518/hfes.46.1.50.30392>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.
- Tausch, A. (2025). *Mental Model of Collaborative Technologies Scale (MMCT-S)* [Preprint]. OSF. <https://doi.org/10.17605/OSF.IO/HQYNY>

LICHT INS DUNKEL DER SHADOW AI TRAGEN: MIXED-METHODS-STUDIE ZU GENAI LITERACY UND NUTZUNGSMUSTERN IM ZUSAMMENHANG MIT ARBEITSBEZOGENER KREATIVLEISTUNG

Elisabeth Liedmann

Die Nutzung generativer Künstlicher Intelligenz (GenAI) findet zunehmend außerhalb Top-down-gesteuerter organisationaler Implementierungsprozesse statt. Mitarbeitende greifen zunehmend selbstständig auf generative KI zurück, insbesondere für kreative Problemlösetätigkeiten. Der dezentrale Einsatz von Künstlicher Intelligenz (KI) bzw. GenAI ohne ausdrückliche Befürwortung oder Verwaltung durch Management, IT- und Cybersicherheitsabteilungen wird als Shadow AI (Puthal et al., 2025) bezeichnet. Damit verbunden sind nicht nur kritische Fragen zu Compliance und Cybersicherheit, sondern auch Fragen danach, inwiefern kreative Leistungsgewinne und -einbußen entstehen und wie sich diese Effekte in nicht organisational gestalteten Nutzungskontexten entfalten (Puthal et al., 2025).

Auf dem 54. Kongress der DGPs stellte Elisabeth Liedmann einen Teil der Untersuchung zur selbstgesteuerten GenAI-Nutzung bei arbeitsbezogenen kreativen Problemlöseaufgaben vor. Im Mittelpunkt stand die Frage, ob und unter welchen Bedingungen die selbstgesteuerte Nutzung eines GenAI-Systems die Qualität kreativer Arbeitsergebnisse beeinflusst. Hängt der Effekt von individuellen Voraussetzungen wie der bereits vorhandenen kreativen

Leistungsfähigkeit oder der GenAI Literacy ab?

Die Studie folgte einem zweiteiligen Mixed-Methods-Design mit N = 150 Teilnehmenden, die zwei arbeitsbezogene Kreativaufgaben sowie einen Performanztest zur GenAI Literacy (Jin et al., 2025) bearbeiteten. Die erste Aufgabe diente der Messung der kreativen Ausgangsleistung, die zweite Aufgabe wurde randomisiert entweder ohne GenAI oder mit Zugang zu ChatGPT bearbeitet. In einer zweiten Erhebung bewerteten anschließend unabhängige Rater:innen die anonymisierten Arbeitsergebnisse anhand von Kreativitätskriterien. Zusätzlich wurden die Mensch-GenAI-Interaktionen in der Experimentalgruppe dokumentiert und in einer qualitativen Analyse angelehnt an Kuckartz (2018) sowie Leger und Buschek (2025) ausgewertet.

In den quantitativen Untersuchungen zeigte sich zunächst, dass 49,7% der Befragten mindestens ein GenAI-Tool bei ihrer Arbeit nutzten, für das keine organisationale Freigabe bestand oder ihnen bekannt war. Dabei werden insbesondere auch bestehende Unsicherheiten hinsichtlich des Freigabestatus von GenAI-Tools deutlich: 33,3% der Befragten gaben an, bei mindestens einem der verwendeten Tools nicht zu wissen, ob eine Freigabe besteht.

Für die kreative Leistung zeigte sich ein klarer Unterschied zwischen den Bedingungen: Teilnehmende, die GenAI nutzten, erzielten bei der arbeitsbezogenen Kreativitätsaufgabe signifikant höhere Kreativitätsbewertungen als Teilnehmende ohne GenAI-Nutzung ($M = 4.62$ vs. $M = 4.27$, $b = .36$, 95%-KI [.17; .55], $p < .001$).

Gleichzeitig blieb die individuelle Ausgangsleistung relevant: Personen mit höherer kreativer Ausgangsleistung erzielten unabhängig von der ChatGPT-Nutzung höhere Kreativitätsbewertungen als Personen mit niedrigerer kreativer Ausgangsleistung (ohne GenAI: $b = .33$, 95%-KI [.13; .54]; mit GenAI: $b = .42$, 95%-KI [.17; .66]). Entgegen der Hypothese fiel der zusätzliche Vorteil durch GenAI also nicht systematisch größer oder kleiner aus, je nachdem, wie kreativ eine Person bereits zu Beginn war ($b = .08$, 95%-KI [-.24; .41], $p = .613$).

Auch eine höhere GenAI Literacy verstärkte den beobachteten Leistungsvorteil nicht ($b = -.06$, 95%-KI [-.12; .01], $p = .073$). Ein höheres Wissen über und Verständnis von GenAI war nicht mit besseren kreativen Arbeitsergebnissen verbunden (ohne GenAI: $b = .03$, 95%-KI [-.01; .07]; mit GenAI: $b = -.03$, 95%-KI [-.08; .02]). Damit rückt stärker in den Mittelpunkt, wie GenAI in den kreativen Arbeitsprozess eingebunden wird. Hierzu kann die qualitative Analyse Aufschluss geben, in welcher die funktionalen Nutzungsmuster und Formen der Mensch-GenAI-Interaktion im Fokus stehen.

Für Organisationen ergeben sich bereits jetzt zwei miteinander verbundene Implikationen. Einerseits verdeutlicht der beobachtete Leistungsvorteil das Potenzial darin, Mitarbeitenden einen produktiven Zugang zu GenAI zu ermöglichen – zumindest für solch kurzfristige Lösungsentwicklungen. Der Einsatz von GenAI ersetzt jedoch nicht die kreativen

Fähigkeiten von Mitarbeitenden: Die kreative Ausgangsleistung war unabhängig von der GenAI-Nutzung mit der späteren kreativen Leistung verbunden. Organisationen sollten daher nicht allein in den Zugang zu GenAI und entsprechende Nutzungskompetenzen investieren, sondern weiterhin gezielt die kreativen Fähigkeiten ihrer Mitarbeitenden fördern. Technologische Unterstützungen und die Entwicklung menschlicher kreativer Fähigkeiten müssen gemeinsam gedacht werden.

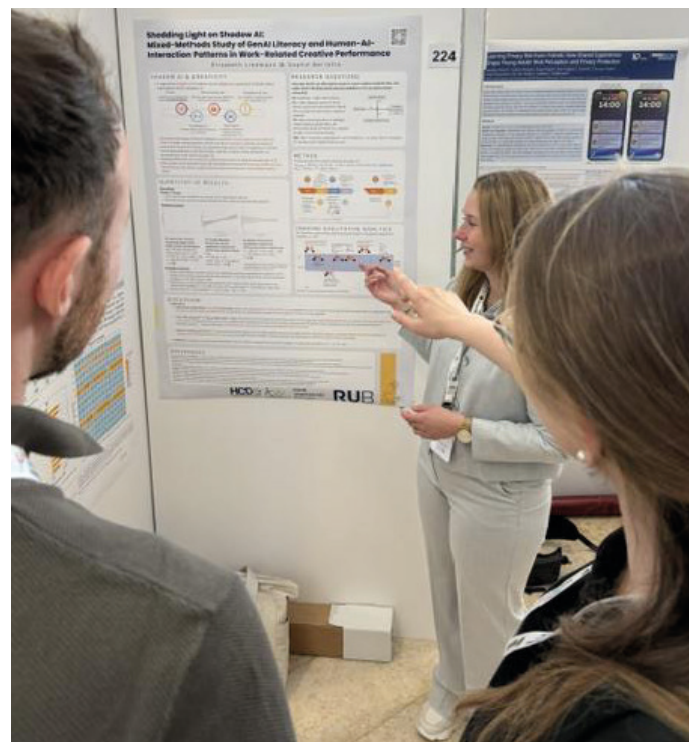


Foto: Postersession auf dem 54. Kongress der DGPs



Literatur

- Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). GLAT: The generative AI literacy assessment test. *Computers and Education: Artificial Intelligence*, 9, 100436. <https://doi.org/10.1016/j.caeai.2025.100436>
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (4., überarbeitete Aufl). Beltz.
- Leger, R., & Buschek, D. (2025). Human-AI Interaction Patterns in Creative Domains and Their Time-Based Visualization. *Proceedings of the Mensch Und Computer 2025*, 422–433. <https://doi.org/10.1145/3743049.3743081>
- Puthal, D., Mishra, A. K., Mohanty, S. P., Longo, A., & Yeun, C. Y. (2025). Shadow AI: Cyber Security Implications, Opportunities and Challenges in the Unseen Frontier. *SN Computer Science*, 6(5), 405. <https://doi.org/10.1007/s42979-025-03962-x>

THE EMOTIONAL SIDE OF DIGITAL CHANGE: A WEEKLY DIARY STUDY ON TECHNOLOGY-RELATED DEMANDS, EMOTION-FOCUSED COPING, AND EMPLOYEE WELL-BEING IN PUBLIC ADMINISTRATION

Charlotte Glenz

Digitale Veränderungen in der öffentlichen Verwaltung gehen für Beschäftigte mit vielfältigen technologiebezogenen Anforderungen einher. Trotz erheblicher Investitionen in digitale Infrastrukturen erreichen viele Veränderungsvorhaben nicht die angestrebten Wirkungen. Dies deutet darauf hin, dass der Erfolg organisationaler digitaler Veränderungen nicht allein von der technischen Umsetzung abhängt, sondern auch davon, wie Beschäftigte mit den damit verbundenen Anforderungen umgehen. Die vorliegende Studie untersucht die Zusammenhänge zwischen technologiebezogenen Anforderungen und dem arbeitsbezogenen Wohlbefinden der Beschäftigten. Im Mittelpunkt steht dabei die Rolle emotionsfokussierter Bewältigungsstrategien.

Die Studie ist als sechswöchige Weekly-Diary-Studie angelegt. Teilnehmen können Beschäftigte der öffentlichen Verwaltung, die in ihrem eigenen Arbeitsalltag aktuell von der Einführung oder Veränderung digitaler Technologien oder Arbeitsprozesse betroffen sind. Nach einer Eingangsbefragung berichten die Teilnehmenden über sechs Wochen hinweg einmal wöchentlich von ihren aktuellen Erfahrungen. Erfasst werden qualitative technologiebezogene Anforderungen und technologiebedingte Überlastung, vier emotionsfokussierte Bewältigungsstrategien (positive Neubewertung, Akzeptanz, emotionaler Ausdruck und Verleugnung) sowie Arbeitsengagement und emotionale Erschöpfung. Angestrebt wird eine Stichprobe von etwa

100 bis 150 Beschäftigten. Die Auswertung erfolgt mithilfe von Mehrebenenanalysen, durch die wöchentliche Veränderungen innerhalb der Beschäftigten von stabilen Unterschieden zwischen den Beschäftigten getrennt werden können. Die Datenerhebung läuft noch bis Ende November 2026.

Auf der Within-Person-Ebene wird erwartet, dass stärkere technologiebezogene Anforderungen in einer Arbeitswoche mit höherer emotionaler Erschöpfung und geringerem Arbeitsengagement verbunden sind. Darüber hinaus wird angenommen, dass diese Zusammenhänge von den eingesetzten emotionsfokussierten Bewältigungsstrategien abhängen. Positive Neubewertung und Akzeptanz sollen die ungünstigen Zusammenhänge zwischen technologiebezogenen Anforderungen und Wohlbefinden abschwächen, während emotionaler Ausdruck und Verleugnung diese verstärken sollen.

Die Ergebnisse sollen zu einem differenzierteren Verständnis davon beitragen, wie Beschäftigte digitale Veränderungen im Arbeitsalltag bewältigen und welche emotionsfokussierten Strategien sie im Umgang mit technologiebezogenen Anforderungen unterstützen können.

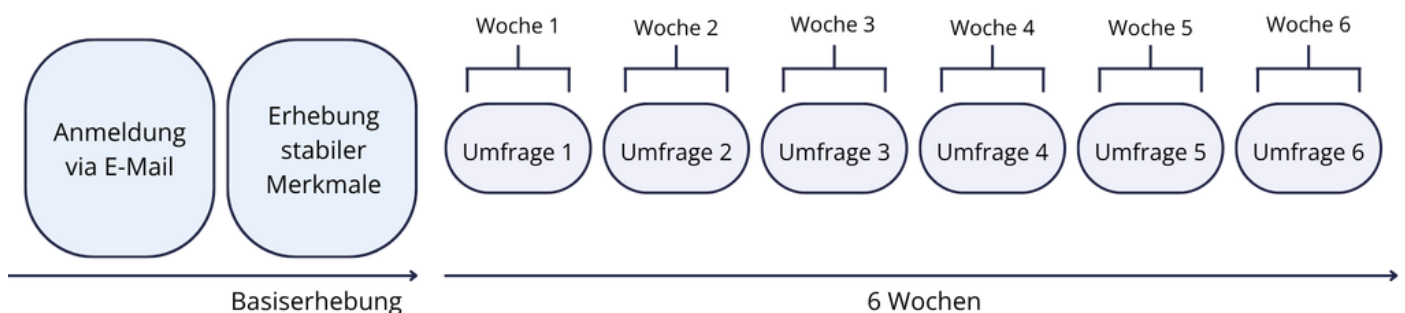


Abbildung: Zeitlicher Ablauf der Datenerhebung im Weekly-Diary-Design
Anmerkung: Die Befragungen werden online über SoSci Survey durchgeführt.

THE EMOTIONAL SIDE OF DIGITAL CHANGE: A WEEKLY DIARY STUDY ON TECHNOLOGY-RELATED DEMANDS, EMOTION-FOCUSED COPING, AND EMPLOYEE WELL-BEING IN PUBLIC ADMINISTRATION

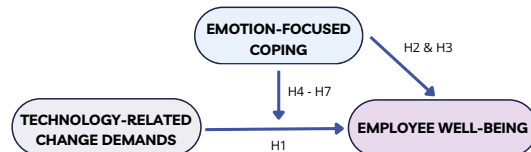
Charlotte Glenz; Annette Kluge; Marcel Kern



1. Successful digital change requires more than new technology

- Digital change in public administration confronts employees with increasing technology-related demands.
- Despite substantial investments, many digital change initiatives **fall short** of their intended impact.
- Their success may therefore depend not only on technological implementation, but also on **how employees cope** with the technology-related demands arising during the change process.

4. Research Model and Hypotheses



Hypotheses

At the weekly level:

- H1:** Technology-related change demands (qualitative demands and overload) are associated with higher exhaustion and lower engagement.
- H2-H3:** Positive reinterpretation and acceptance are associated with lower exhaustion and higher engagement; denial and venting show the reverse pattern.
- H4-H7:** Positive reinterpretation and acceptance buffer, whereas denial and venting amplify, the demand-outcome associations.

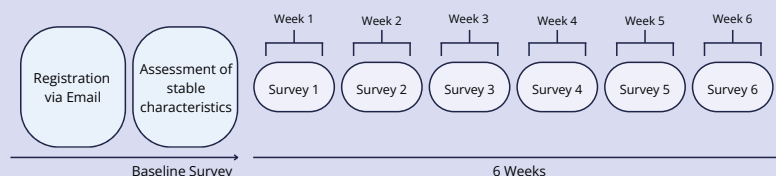
6. Sample

- Employees in German public administration currently experiencing the introduction or modification of digital technologies in their own work.
- Target: $N \approx 150$ employees.
- Preliminary sample: $N = 20$ employees; 63 weekly observations.

7. Design

Weekly diary:

- Baseline + six fixed weekly assessments.
- Weekly observations nested within employees.
- Reference to one specific digital change.



8. Analysis

- Multilevel structural equation modeling in Mplus, separating within- and between-person variance.
- Work engagement and emotional exhaustion modeled simultaneously as correlated outcomes.
- Separate Bayesian models for each within-person demand $\times$ coping interaction using the XWITH model command in Mplus

2. Coping with Change – Week by Week

- Transactional stress theory identifies coping as central to how demands affect employees (Lazarus & Folkman, 1984).
- Yet technostress research has often examined coping in cross-sectional studies of everyday IT use (Pirkkalainen et al., 2019; Rohwer et al., 2022).
- Less is known about emotion-focused coping as a weekly, within-person process during ongoing digital change.



3. Research Question

To what extent do emotion-focused coping strategies moderate the weekly associations between technology-related demands and employee well-being during ongoing digital change in public administration?

5. Key Study Constructs

Emotion-Focused Coping (COPE; Carver et al., 1989)

Positive Reinterpretation

Reframing the situation in a more positive or constructive way.



Acceptance

Acknowledging that the situation is real and needs to be dealt with.



Denial

Rejecting or distancing oneself from the reality of the situation.



Venting

Expressing negative feelings associated with the situation.



Technology Related Change Demands

Qualitative Demands

(SALSA; Rimann and Udris, 1997)
This week, there were times when working with the technology was too difficult for me.

Technology-Induced Overload

(Cammann et al., 1983)
This week, I did not have enough time to complete all tasks related to the new technology.

Employee Well-Being

Emotional Exhaustion

(MBI; Maslach et al., 1996)
During this workweek, I felt exhausted.

Work Engagement

(UWES-3; Schaufeli et al., 2017)
This week, I felt full of energy.

9. Preliminary Descriptive Findings

Expected function	Coping strategy	M (SD)	rW Exhaustion	rW Engagement
Potentially adaptive	Positive Reinterpretation	3.79 (0.87)	.09	-.14
Potentially adaptive	Acceptance	4.13 (0.91)	-.04	.14
Potentially maladaptive	Venting	1.70 (0.75)	.45	-.28
Potentially maladaptive	Denial	1.17 (0.47)	.31	-.38

Note. 20 employees, 63 person-week observations. Coping was rated from 1 to 5. **rW** represents preliminary person-mean-centered within-person correlations. Descriptive statistics are based on 56–58 weekly observations; correlations are based on up to 55 paired observations.

VERÖFFENTLICHUNGEN DES LEHRSTUHL

Rehmer, S., Ames, I., Kluge, A., Bethlehem, T., Schweden, F., Lang, F., Antoni, C., Dettmers, J., Eickholt, C., Trimpop, R. (2026). Arbeits- und Organisationspsychologie in Gesundheit und Arbeitssicherheit: Aktuelle Entwicklungen, Professionalisierung und Zukunftsperspektiven. *Wirtschaftspsychologie*, 2026. 28:5-12 DOI: <https://doi.org/10.2440/004-0081>

Thomaschewski, L., Kluge, A., & Weyers, B. (2027). Advancing safety training for nuclear decommissioning and dismantling: Integrating MX-based approaches for improved competency development. *Applied Ergonomics*, 138, 104850.

Vogel, O., Schilling J. & Kluge, A. (2026). Erfassung des subjektiven Erlebens von aufgaben- und personenbezogenen Grenzüberschreitungen als Facette von Machtmissbrauch in der Wissenschaft. Entwicklung des SeM-Wi für Forschung und Intervention. *Zeitschrift für Arbeits- und Organisationspsychologie (ZAO)*. Advance Article <https://doi.org/10.1026/0932-4089/a000487>

Liebherr, M., Langer, M., & Kluge, A. (2026). KI braucht Psychologie und Psychologie braucht KI – Perspektiven der DGPs-Fachgruppen. *Psychologische Rundschau*, 77(3), 121-122. <https://doi.org/10.1026/0033-3042/a000774>

Hagemann, V., Berretta, S., Decius, J., Georganta, E., Hofer, A., Klamar, A., ... Vogel, O. (2026). Handlungsfelder des KI-Einsatzes in Arbeit, Teams und Organisationen: Eine Perspektive der Fachgruppe Arbeits-, Organisations- und Wirtschaftspsychologie zu Künstlicher Intelligenz in der Arbeitswelt. *Psychologische Rundschau*, 77(3), 162-168. <https://doi.org/10.1026/0033-3042/a000768>





WISSENSCHAFTLICHE MITARBEITER:INNEN
DES TEAMS ARBEITS-, ORGANISATIONS- & WIRTSCHAFTSPSYCHOLOGIE
& KOOPERATIONSPARTNER:INNEN



IMPRESSUM

Komplexität und Lernen ISSN 1661-8629 erscheint vierteljährlich (seit 2007)



HERAUSGEBERIN

Prof. Dr. Annette Kluge
Lehrstuhl Arbeits-, Organisations- und Wirtschaftspsychologie
Ruhr-Universität Bochum
Universitätsstraße 150
44780 Bochum



NEWSLETTER

Wenn Sie Interesse an unserem Newsletter haben, mailen Sie mir. Ich nehme Sie gerne in unserem Verteiler auf.
annette.kluge@rub.de



DESIGN

Elisa Schallau
M.Sc. Psychologin & Mediengestalterin