Smarter Workplaces: The Role of AI in Promoting Occupational Safety and Health

Potsdam, 16 September 2025
Proceedings

baua: Focus

These are the proceedings of the AI_4POSH session – Smarter Workplaces: The Role of AI in Promoting Occupational Safety and Health, held as part of the 48th German Conference on Artificial Intelligence in Potsdam, Germany, 16-19 September 2025. The session was organized by PD Dr. Thea Radüntz, head of the unit "Artificial Intelligence in Work Systems" of the Federal Institute for Occupational Safety and Health.
The AI_4POSH session highlights the diverse and growing role of Artificial Intelligence (AI) in shaping healthier, safer, and more human-centered workplaces. The presented papers span applications in workplace health promotion, occupational medicine, mental health, algorithmic management, risk assessment, and AI-assisted operations. Contributions explore both the potential and challenges of AI from technical innovations like sensor networks and image analysis to human-centered concerns such as digital sovereignty, engagement, and workload management. Together, these works reflect an interdisciplinary effort to harness AI for promoting occupational safety and health while critically examining its impact on workers and organizations. By addressing how intelligent systems reshape the modern workplace, the AI_4POSH session contributes to the development of safer, more responsive, and ethically sound working environments in the age of AI.

Content

Federal Institute for Occupational Safety and Health

KIM Künstliche Intelligenz und mentale Gesundheit: Ergebnisse einer qualitativen Studie in der Versicherungsbranche

Stefanie André
IU Internationale Hochschule

Zusammenfassung

Die zunehmende Integration Künstlicher Intelligenz (KI) in die Arbeitswelt stellt nicht nur technologische, sondern auch ethische und gesundheitsbezogene Herausforderungen. Im Forschungsprojekt KIM – Künstliche Intelligenz und Mentale Gesundheit – wurde im Zeitraum von Februar bis April 2025 in Kooperation mit dem Bundesverband der Assekuranzführungskräfte (VGA) die erste qualitative Phase eines Mixed-Methods-Designs durchgeführt. Ziel war es, erste empirische Einblicke in psychosoziale Belastungen, ethische Spannungsfelder und organisationale Schutzfaktoren im Umgang mit KI-Systemen zu gewinnen. Insgesamt wurden 17 halbstrukturierte Online-Interviews mit Führungskräften aus der Versicherungsbranche geführt und inhaltsanalytisch nach der zusammenfassenden Inhaltsanalyse nach Mayring mit MAXQDA ausgewertet. Die qualitative Analyse liefert ein vertieftes Verständnis darüber, wie Führungskräfte den KI-gestützten Wandel erleben, welche Rolle sie sich selbst zuschreiben und welche organisationalen, individuellen und strukturellen Bedarfe daraus hervorgehen. Das Salutogenese-Modell (Antonovsky, 1997) dient dabei als richtungsweisender Referenzrahmen, da der Zusammenhang zwischen einem starken Kohärenzgefühl und der Gesundheit von Beschäftigten vielfach empirisch belegt ist. Die Studie diskutiert darüber hinaus die Weiterentwicklung gesetzlicher Gefährdungsbeurteilungen, den Umbau von Führungsrollen im digitalen Wandel sowie die Einführung einer interdisziplinären Fachkraft für KI-gestützte Arbeitsgestaltung.

Einleitung

KI-gestützte Systeme verändern zunehmend die Struktur von Arbeit, Entscheidungsfindung und Führung. Während potenzielle Effizienzgewinne im Fokus unternehmerischer Interessen stehen, rücken Fragen der ethischen Verantwortlichkeit, psychischen Gesundheit und organisationalen Resilienz erst langsam in das Zentrum wissenschaftlicher und betrieblicher Aufmerksamkeit. Insbesondere in wissensbasierten Dienstleistungsbranchen wie der Versicherungs- und Finanzwirtschaft werden starke Einschnitte durch den Einsatz von KI erwartet (Tenner & Diaz, 2024). Dennoch fehlen bislang empirische Befunde zu psychosozialen Belastungen, die sich aus dem konkreten Einsatz von KI ergeben. Die Dringlichkeit dieses Forschungsfeldes wird auch durch die WHO-Leitlinien zur psychischen Gesundheit am Arbeitsplatz unterstrichen, die Führungskräften eine zentrale Rolle bei der Förderung mentaler Gesundheit zuweisen (WHO, 2022).

Theoretischer Hintergrund

Das Projekt KIM basiert auf einem interdisziplinären Bezugsrahmen aus Gesundheitsmanagement, Arbeits- und Organisationspsychologie und Ethik. Die psychische Gesundheit wird im Sinne der Weltgesundheitsorganisation (WHO, 2021) als ein Zustand des Wohlbefindens verstanden, in dem Individuen ihre Fähigkeiten ausschöpfen, normale Lebensbelastungen bewältigen, produktiv arbeiten und zu ihrer Gemeinschaft beitragen können. Im gesundheitswissenschaftlichen Diskurs stützt sich das Projekt auf das Konzept der Salutogenese von Antonovsky (1997), das insbesondere den Zusammenhang zwischen Kohärenzerleben – bestehend aus Verstehbarkeit, Handhabbarkeit und Sinnhaftigkeit – und gesundheitlicher Stabilität in komplexen Kontexten betont (Eriksson & Lindström, 2006). Damit verbunden ist das Job-Demands-Resources-Modell (Bakker & Demerouti, 2007), das verdeutlicht, wie Arbeitsanforderungen, wie technologische Überforderung und Ressourcen und Führungsunterstützung im Wechselspiel die Gesundheit beeinflussen. Zur konzeptionellen Rahmung folgt das Projekt zunächst einer funktionalen, international anerkannten Arbeitsdefinition von Künstlicher Intelligenz, wie sie von der Europäischen Kommission und der OECD vorgeschlagen wurde: Künstliche Intelligenz bezeichnet Systeme, die durch maschinelles Lernen, regelbasierte Verfahren oder andere rechnergestützte Techniken in der Lage sind, Daten zu analysieren, Entscheidungen zu treffen oder Handlungen auszuführen, die typischerweise als Ausdruck menschlicher Intelligenz verstanden werden (High-Level Expert Group on AI, 2019; OECD, 2021). Diese Definition beschreibt vorrangig die Leistungsfähigkeit von KI-Systemen. Sie wird im vorliegenden Projekt durch eine soziotechnische Perspektive erweitert, die KI als eingebettet in soziale Praktiken, Diskurse und Machtverhältnisse versteht (von Garrel et al., 2023; Miebach, 2023). Damit wird KI nicht nur als technisches Instrument, sondern als aktiver Mitgestalter organisationaler und gesellschaftlicher Ordnung begriffen. Der Fokus liegt somit gleichermaßen auf der Funktionsweise wie auf den sozialen Implikationen von KI, insbesondere im Kontext von Führung, Verantwortung und psychischer Gesundheit. Bezogen auf den Technologieeinsatz greift das Projekt zudem auf die Systematik von Russell und Norvig (2020) zurück, die vier verschiedene Ansätze zur Klassifikation von KI unterscheiden: Systeme, die wie Menschen denken (zum Beispiel durch Nachbildung kognitiver Prozesse), Systeme, die rational denken (etwa durch logische Schlussfolgerungen), Systeme, die wie Menschen handeln (zum Beispiel in dialogischen Interaktionen) und Systeme, die rational handeln (durch zielgerichtete Entscheidungen auf Basis von Umweltinformationen). Diese Differenzierung ermöglicht eine fundierte Analyse der unterschiedlichen Auswirkungen von KI auf Arbeitsprozesse, Rollenverständnisse und psychosoziale Dynamiken im Unternehmen. Die wissenschaftliche Relevanz des Projekts ergibt sich unter anderem aus einer systematischen Übersichtsarbeit von Soulami et al. (2024), die erstmals die psychosozialen Auswirkungen von KI am Arbeitsplatz umfassend analysiert. Dabei werden zentrale Belastungsdimensionen wie kognitive Erschöpfung, Kontrollverlust, Rollenunsicherheit und soziale Isolation identifiziert. Besonders betont wird die bislang unzureichende empirische Evidenz für den Dienstleistungssektor, obwohl dort KI-basierte Systeme in hohem Maße mit menschzentrierter Interaktion und emotionaler Arbeit verknüpft sind. Das Projekt KIM greift diese Forschungslücke gezielt auf und liefert auf Basis der qualitativen Analyse in der Versicherungsbranche erste empirische Einblicke in die spezifischen psychosozialen Dynamiken und Belastungen in einer wissens- und kommunikationsintensiven Dienstleistungsumgebung. Die Versicherungsbranche ist hierbei ein besonders relevantes Untersuchungsfeld, da sie einerseits durch hohe Wissensintensität, andererseits durch strukturelle Trägheit geprägt ist: beides Faktoren, die im Zusammenspiel mit KI erhebliche psychosoziale Auswirkungen entfalten können.

Methode

Im Rahmen eines Mixed-Methods-Designs wurde eine qualitative Erhebung als erste Projektphase realisiert in Kooperation mit dem Bundesverband der Assekuranzführungskräfte (VGA) zwischen Februar und April 2025. Zur Erhebung der qualitativen Daten wurden 17 halbstrukturierte Expert:inneninterviews mit Führungskräften aus der Versicherungsbranche geführt (Gläser & Laudel, 2010). Das Projekt KIM verfolgt das Ziel, zentrale psychosoziale Herausforderungen und Gestaltungsmöglichkeiten beim Einsatz Künstlicher Intelligenz (KI) in der Führungsarbeit zu identifizieren. Im Mittelpunkt stehen dabei folgende Forschungsfragen:

- Welche psychischen Belastungen beobachten Führungskräfte bei ihren Mitarbeitenden im Umgang mit KI?
- Welche spezifische Unterstützung benötigen Führungskräfte zur Förderung mentaler Gesundheit in der digitalen Transformation?
- Welche Chancen und Risiken sehen sie im Einsatz von KI für die Mitarbeiter:innengesundheit?
- Wie verändert KI die Rolle von Führungskräften im Gesundheitsmanagement?
- Wie stehen Führungskräfte zum Einsatz von Coaching-Chatbots und welche Nutzungsmöglichkeiten sehen sie?

Die Datenerhebung orientierte sich an einem theoriebasierten Leitfaden, der die oben genannten Dimensionen systematisch erfasste. Die qualitative Analyse erfolgte auf Grundlage der zusammenfassenden Inhaltsanalyse nach Mayring (2015) unter Anwendung der Software MAXQDA (Kuckartz & Rädiker, 2019). Dabei wurde ein methodisch kombiniertes Vorgehen gewählt: In einem ersten Schritt wurde das Material induktiv ausgewertet, um neue, aus dem Datenmaterial emergierende Kategorien zu identifizieren. Diese wurden anschließend mit deduktiv aus Theorie und Forschungsfragen abgeleiteten Kategorien abgeglichen, erweitert und systematisch integriert. Die Kodierung erfolgte zunächst durch eine Forscherin. In einem zweiten Schritt überarbeitete und ergänzte eine weitere Forscherin das Kategoriensystem. Zur Sicherung der inhaltlichen Konsistenz, intersubjektiven Nachvollziehbarkeit und reflexiven Validierung wurde ein „reflektierendes Team“ einbezogen (Bohnsack, 2021). Die Zwischenergebnisse wurden im Forschungsteam diskutiert und kritisch gespiegelt. Dabei wurde besonderer Wert auf die Offenlegung interpretativer Entscheidungen und die Reflexion eigener Vorannahmen gelegt. Die Analyse strebte eine theoretische Sättigung an, die nach etwa 14 Interviews erreicht wurde; drei weitere Interviews wurden ergänzend zur Validierung und zur Erfassung seltener Perspektiven herangezogen (Fusch & Ness, 2015). Die angewendeten Methoden wurden dokumentiert und entlang der Kriterien Transparenz, Objektivität, Validität und Intersubjektivität reflektiert. Zur Absicherung der Bedeutungskontexte (Indexikalität) wurden zentrale Aussagen stets im thematischen Zusammenhang betrachtet und nicht losgelöst interpretiert (Flick, 2018). Die interpretativen Prozesse und Entscheidungen wurden fortlaufend dokumentiert, um die Nachvollziehbarkeit des Forschungsprozesses zu gewährleisten. Auf Basis der qualitativen Ergebnisse wird im August/September 2025 eine bundesweite Onlinebefragung zur Überprüfung der abgeleiteten Hypothesen durchgeführt. Die Zielgruppe umfasst Führungskräfte der Versicherungs- und Finanzbranche. Geplant ist eine Stichprobengröße von ca.300 – 400 Teilnehmenden, rekrutiert über den Bundesverband der Assekuranzführungskräfte (VGA) und branchenspezifische Netzwerke. Im Sinne eines konvergenten Mixed-Methods-Designs werden qualitative und quantitative Daten wie folgt kombiniert: Die qualitative Analyse liefert die inhaltliche Grundlage für die Entwicklung des Fragebogens. Die Ergebnisse beider Phasen werden im Anschluss systematisch verglichen, um Muster zu validieren und Unterschiede sichtbar zu machen (Kuckartz, 2014). Die Interpretation erfolgt integriert, um praxisrelevante Handlungsempfehlungen zur zukünftigen Arbeitsgestaltung abzuleiten. Ziel ist es, individuelle Perspektiven mit generalisierbaren Befunden zu verknüpfen und so evidenzbasierte Impulse für gesundheitsförderliche Führungsstrategien im digitalen Wandel zu entwickeln.

Ergebnisse

Die Auswertung zeigt ein komplexes Belastungsbild, das sich in folgenden Hypothesen verdichtet:

- Technologische Überforderung: Das hohe Tempo und die Vielfalt an KI-Anwendungen führen zu kognitiver Überlastung.
- Strategische Orientierung: Fehlende strategische Orientierung führt zu Unsicherheit und destabilisiert das Führungsverhalten.
- Intransparenz und Kontrollverlust: Intransparente Entscheidungslogiken sowie fehlende Partizipation und Kommunikation im Umgang mit KI verstärken Unsicherheiten und mindern das Vertrauen.
- Erosion von Selbstwirksamkeit: Algorithmische Steuerung und automatisiertes Feedback untergraben subjektive Handlungskontrolle.
- Rollenverunsicherung in der Führung: KI verändert Führungsrollen in Richtung technischer Moderation statt personaler Verantwortung.
- Ökonomisch-ethische Zielkonflikte: Zwischen Effizienz- und Fürsorgezielen entstehen Spannungen, die nicht systematisch reflektiert werden.
- Soziale Fragmentierung: Digitale Prozesse beeinträchtigen Zugehörigkeitsgefühl und informelle Kommunikation.

Diese Hypothesen lassen sich mit bestehenden empirischen Studien zu Technostress (Barton & Pöppelbuß, 2022), KI-Ethik (Heesen et al., 2020) und psychosozialen Belastungen am Arbeitsplatz (RobelsKI et al., 2022) in Verbindung bringen und erweitern diese um branchenspezifische Perspektiven aus der Versicherungswirtschaft.

Diskussion

Die Ergebnisse der qualitativen Analyse zeigen deutlich, dass die Integration von Künstlicher Intelligenz (KI) in die Arbeitswelt mehr ist als eine technologische Innovation: Sie verändert grundlegend die psychologischen Bedingungen von Arbeit und Führung. Dabei treten insbesondere Belastungen in Erscheinung, die auf Intransparenz, Kontrollverlust, Rollenverunsicherung sowie eine Erosion von Selbstwirksamkeit zurückzuführen sind. Diese Muster stehen in engem Zusammenhang mit den Wahrnehmungen der Führungskräfte hinsichtlich der Dynamik von KI-Einführung, organisationaler Veränderungsbereitschaft und individueller Handlungsspielräume. Die gesetzlich vorgeschriebene Gefährdungsbeurteilung psychischer Belastungen nach § 5 ArbSchG sollte um KI-spezifische Indikatoren erweitert werden. Dazu zählen unter anderem algorithmische Fremdsteuerung, kognitive Fragmentierung und digitale Reizüberflutung – aber auch psychosoziale Belastungen wie Rollenunsicherheit, Kontrollverlust, soziale Isolation und die Erosion von Sinn- und Wirksamkeitserleben.

Diese Belastungsdimensionen wurden sowohl in der qualitativen Analyse des Projekts KIM als auch in systematischen Übersichtsarbeiten (Soulami et al., 2024) als zentrale Herausforderungen im KI-gestützten Arbeitskontext identifiziert. Darüber hinaus bedarf die Rolle der Führungskräfte einer Neuausrichtung. Neben betriebswirtschaftlichen und personalbezogenen Kompetenzen gewinnen ethische Reflexionsfähigkeit und technologische Sensibilität zunehmend an Bedeutung. Dies entspricht auch den Empfehlungen der WHO (2022), die Führung ausdrücklich als gesundheitsprägende Ressource definiert. Als systemischer Vorschlag erscheint daher die Einführung einer neuen internen Funktion sinnvoll: der Fachkraft für KI-gestützte Arbeitsgestaltung. Diese soll an der Schnittstelle zwischen Technologie, Gesundheit und Führung wirken und eine kontinuierliche ethisch-psychologische Begleitung der digitalen Transformation ermöglichen. Zugleich weisen die Ergebnisse auf die Bedeutung salutogenetischer Prinzipien im Umgang mit KI hin. Der Sense of Coherence (SOC) – also das Kohärenzgefühl, bestehend aus Verstehbarkeit, Handhabbarkeit und Sinnhaftigkeit – stellt einen zentralen Schutzfaktor gegen technikbedingte Stressoren dar.

Führungskräfte berichten insbesondere dann von Überforderung und Kontrollverlust, wenn technologische Entwicklungen als unverständlich, nicht beeinflussbar oder sinnentleert erlebt werden. Umgekehrt deuten einzelne Aussagen darauf hin, dass KI-Prozesse dann als entlastend, hilfreich und akzeptabel empfunden werden, wenn sie nachvollziehbar eingeführt, beteiligungsorientiert begleitet und in einen sinnstiftenden organisationalen Rahmen eingebettet sind. Daraus ergibt sich die Empfehlung, sowohl das individuelle Kohärenzgefühl von Führungskräften gezielt zu stärken, als auch organisationale Bedingungen zu schaffen, die ein kollektives Sense of Coherence fördern. Auf individueller Ebene können etwa Schulungen, ethische Reflexionsräume oder begleitende Coaching-Angebote wirksam sein. Auf organisationaler Ebene braucht es transparente Entscheidungsprozesse, dialogische Einführungsformate sowie eine klare strategische Verankerung der KI-Nutzung im Sinne von Sinnorientierung und Mitarbeitendenschutz. Die Ergebnisse dieser qualitativen Phase sind jedoch im Kontext der untersuchten Zielgruppe zu interpretieren. Die befragten Personen waren ausschließlich Führungskräfte aus der Versicherungsbranche – einer wissensintensiven, zugleich stark regulierten und traditionell organisierten Dienstleistungsbranche. Die identifizierten Belastungsmuster spiegeln somit spezifische branchentypische Rahmenbedingungen wider. Für andere Berufsgruppen, etwa im Gesundheitswesen, der öffentlichen Verwaltung oder der Industrie, könnten sich abweichende Belastungsprofile und Handlungsbedarfe ergeben. Die im nächsten Schritt geplante quantitative Onlinebefragung wird daher so konzipiert, dass auch Vergleiche zwischen verschiedenen Subgruppen möglich sind, etwa in Bezug auf Führungslevel, Unternehmensgröße oder Digitalisierungserfahrung. Ein weiterer Aspekt betrifft die Vielfalt der eingesetzten KI-Systeme. Im Rahmen der qualitativen Interviews wurde nicht systematisch zwischen verschiedenen Systemtypen differenziert – etwa zwischen Entscheidungsunterstützungssystemen, generativen Sprachmodellen oder automatisierten Arbeitsprozessen. Die Sicht der Führungskräfte war stärker von allgemeinen Transformationsprozessen und Wirkungsannahmen als von konkreten Technologien geprägt. Dennoch zeigen sich bereits erste Hinweise darauf, dass unterschiedliche KI-Systeme mit jeweils spezifischen psychosozialen Auswirkungen einhergehen können – zum Beispiel in Bezug auf Vertrauen, Verantwortung oder soziale Interaktion. Diese Differenzierung soll in der quantitativen Phase gezielt aufgegriffen werden, um Belastungstypen technologiebezogen systematisch zu erfassen. Insgesamt verdeutlichen die Ergebnisse, dass arbeitsbezogene KI nicht isoliert betrachtet werden kann. Sie ist eingebettet in ein soziotechnisches Gefüge aus technologischer Infrastruktur, organisationalen Aushandlungsprozessen, kulturellen Leitbildern und individuellen Sinnzuschreibungen. Gesundheitsförderliche Gestaltung erfordert deshalb nicht nur technische Regulierung, sondern auch eine stärkere interdisziplinäre Zusammenarbeit zwischen Gesundheitsmanagement, Personalentwicklung, IT, Ethik und Organisationspsychologie. Das Projekt KIM leistet mit der Kombination qualitativer Tiefenanalyse und anschließender quantitativer Validierung einen Beitrag zur evidenzbasierten Weiterentwicklung menschenzentrierter KI-Gestaltung in der Arbeitswelt.

Referenzen

Antonovsky, A. (1997). Salutogenese: Zur Entmystifizierung der Gesundheit. DGVT-Verlag.

Bakker, A. B., & Demerouti, E. (2007). The Job Demands-Resources model: State of the art. Journal of Managerial Psychology, 22(3), 309–328.

Barton, M. C., & Pöppelbuß, J. (2022). Prinzipien für die ethische Nutzung künstlicher Intelligenz. HMD Praxis der Wirtschaftsinformatik, 59, 468–481.

Bohnsack, R. (2021). Rekonstruktive Sozialforschung: Einführung in qualitative Methoden (10. Aufl.). UTB.

Eriksson, M., & Lindström, B. (2006). Antonovsky's sense of coherence scale and the relation with health: A systematic review. Journal of Epidemiology & Community Health, 60(5), 376–381.

Flick, U. (2018). Qualitative Sozialforschung: Eine Einführung (8. Aufl.). Rowohlt.

Fusch, P. I., & Ness, L. R. (2015). Are We There Yet? Data Saturation in Qualitative Research. The Qualitative Report, 20(9), 1408–1416.

Gläser, J., & Laudel, G. (2010). Experteninterviews und qualitative Inhaltsanalyse (4. Aufl.). VS Verlag für Sozialwissenschaften.

Heesen, J., Schmid, U., & Wrobel, S. (2020). Ethik-Briefing. Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen. Plattform Lernende Systeme. *https://www.plattform-lernende-systeme.de/publikationen.html*

High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. European Commission. *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419*

Kuckartz, U. (2014). Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren. Springer VS.

Kuckartz, U., & Rädiker, S. (2019). Analyzing qualitative data with MAXQDA. Springer.

Mayring, P. (2015). Qualitative Inhaltsanalyse: Grundlagen und Techniken (12. Aufl.). Beltz.

Miebach, C. (2023). Künstliche Intelligenz als soziales Phänomen: Eine diskurstheoretische Perspektive. Springer VS.

OECD. (2021). OECD Framework for the Classification of AI Systems. *www.oecd.org/publications/oecd-framework-for-the-classification-of-AI-systems-cb6d9eca-en.htm*

RobelsKI, S., Vock, S., Richter, M., KIttelmann, M., Westhoven, M., Gabriel, S., Voß, S., Sommer, S. (2022). Technischer und organisatorischer Arbeitsschutz in der digitalisierten Arbeitswelt. In A. Tisch & S. WischniewsKI (Hrsg.), Sicherheit und Gesundheit in der digitalisierten Arbeitswelt (S. 273–311). Nomos.

Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson Education.

Soulami, A., Fragkos, K. C., & Fragkos, K. (2024). Exploring how AI adoption in the workplace affects employees: A bibliometric and systematic review. Technological Forecasting and Social Change, 198, 122943.

Tenner, T., & Diaz, V. A. (2024). Einsatz von KI in Banken – Chancen und Herausforderungen. Bankenverband. *https://bankenverband.de/kuenstliche-intelligenz/einsatz-von-KI-banken-chancen-herausforderungen*

von Garrel, J., Jung, J., & Werens, M. (2023). KI in der Arbeitswelt: Soziotechnische Perspektiven auf neue Steuerungslogiken. Beltz Juventa.

Werner, M. H. (2021). Einführung in die Ethik (2. Aufl.). J. B. Metzler.

World Health Organization. (2021). Mental health: Strengthening our response.
www.who.int/news-room/fact-sheets/detAIl/mental-health-strengthening-our-response

World Health Organization. (2022). WHO guidelines on mental health at work.
www.who.int/publications/i/item/9789240053052

Artificial Intelligence in Workplace Health Promotion: How Technology Enhances Participation, Engagement, and SustAInable Impact

Frederik Blohm, Milad Geravand
DeepCare

Despite the growing number of corporate health promotion programs, long-term participation and measurable behavior change remAIn limited. Systematic reviews report participation rates in workplace health programs are generally very low (Bensa & Širok, 2023; Robroek et al., 2009).

While digital initiatives such as smartphone apps offering exercises, meditation, or educational content benefit from high accessibility and low entry barriers, their actual effectiveness in changing dAIly behavior is often weak. Studies show that while these apps see high initial download and usage rates, they suffer from steep declines in engagement within weeks. For example, median user retention after 15 days is just 3.9 %, dropping to 3.3 % after 30 days (Boucher et al., 2024). Similarly, AmagAI et al. (2022) report that up to 71 % of users disengage from health apps within 90 days. Even when participation is high, systematic reviews find little sustAIned impact on actual health behaviors (Milne-Ives et al., 2020).

Beyond Motivation: The Case for a Behavioral Science Approach

Research consistently shows that motivation alone is not enough to establish sustAInable health routines at work or in different contexts. As Schulze et al. (2024) argue, motivation without supporting factors such as sKIlls or conducive environments rarely leads to behavior change. This principle is embodied in the COM-B model, which identifies Capability, Opportunity, and Motivation as equally necessary conditions for behavior (Michie et al., 2011). Yet most workplace health initiatives still focus primarily on rAIsing awareness or offering one-off activities. Common examples include health days, ergonomic consultations, or mandatory trAInings – often within the scope of occupational safety.

This contribution presents a novel approach in which artificial intelligence (AI) is not merely used to monitor or analyze behavior, but actively fosters engagement, lowers behavioral barriers, and enables lasting habit formation.

Introducing Isa: An AI-Based Assistant for Healthier Work Behavior

The centerpiece of this approach is Isa, the Intelligent Seating Assistant – a privacy-preserving, sensor-based AI system that promotes healthier behavior in sedentary work environments (cf. Figure 1). Isa operates entirely offline and provides real-time, context-sensitive feedback to help users improve posture, increase movement, and stay hydrated. Its interventions are subtle, personalized, and seamlessly embedded into dAIly workrotines.

Isa consists of a compact, USB-powered, AI-enabled hardware unit that can be flexibly placed on the desk and has merly the size of a smartphone. The unit contAIns a 3D depth sensor and environmental sensor array (CO_2, VOC, sound, light, temperature) and provides visual feedback through a small toch display (screen diagonal is 14.3 cm / 5.6 inches) used for visual cues (cf. Figure 2). Unlike fitness apps, which often suffer from high attrition rates, Isa integrates passively into the dAIly work routine without requiring users to unlock a screen, input data, or log in. There's no need to take out a smartphone – Isa is always present on the desk, ready to interact, and never competes with distracting notifications or unrelated apps. Its plug-and-play design and fully offline processing ensure minimal friction and support long-term engagement. The device requires no technical expertise and can be set up in just a few minutes.

Fig. 1 Isa on the desk with visualization of 3D ToF sensor, Deep Care GmbH

Isa's intelligence is powered by a suite of deep learning models trAIned on over one million real-world sensor recordings and over ten million simulated datasets. These convolutional neural networks (CNNs) process data from depth and ambient sensors to identify ergonomic patterns, postural deviations, and micro-movements with sub-millimeter precision. The models were developed and optimized using a hybrid dataset composed of real sensor data and simulation-based augmentations. The simulation model was built using a baseline of real user data, from which synthetic depth patterns were generated to reflect diverse conditions (e. g., desk types, lighting, body sizes, movement angles). This process increased data diversity by a factor of 10 and improved model generalization to unseen ergonomic configurations.

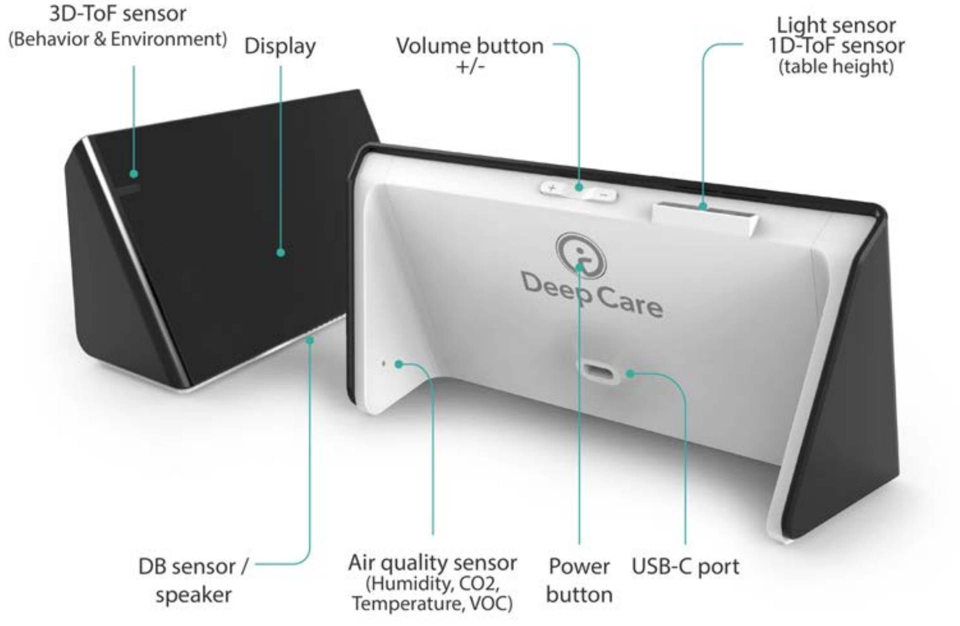

Fig. 2 Overview of Isa device components, Deep Care GmbH

In addition to postural analysis, Isa continuously evaluates environmental conditions including lighting, AIr quality, and ambient noise. For these data streams, a combination of rule-based systems and lightweight machine learning classifiers is used. While CNNs handle spatial depth information (e. g. neck tilt, distance from monitor), decision trees or threshold-based AI heuristics process temporal trends in noise levels or CO_2 concentrations. Isa uses a mix of signal processing and lightweight classifiers (e. g., decision trees) for interpreting environmental data. These include moving average filters for CO_2 and dB(A) noise trends, threshold-based triggers for ventilation prompts, and learned rules for detecting low-light and poor acoustics. These systems are calibrated to recognize health-relevant environmental patterns while mAIntAIning interpretability. All analyses are performed locally on the device, ensuring full compliance with strict data protection standards. No internet connection or cloud storage is required.

Isa's AI engine also powers a self-guided, legally compliant workstation assessment tool for both office and remote settings, developed in collaboration with our partners at B·A·D GmbH. It provides personalized ergonomic evaluations and long-term environmental monitoring, including posture, lighting, AIr quality, and noise. A privacy-compliant PDF report can be generated to document initial risks, corrective actions, and residual risks. This supports both individual health awareness and organizational OSH documentation duties.

Adapting to the User: Avoiding Overstimulation and Building Habits

To foster sustainable routines without causing overload, Isa uses adaptive feedback cycles. Based on user response history (e. g. reaction to previous nudges, frequency of changes), the system adjusts the intensity, frequency, and type of prompts.

For instance, the AI tracks compliance trends and uses learning principles to personalize interaction strategies Isa continuously evaluates interaction patterns, including posture correction latency, responsiveness to nudges, and session-level compliance. If the user consistently ignores posture prompts, Isa reduces their frequency or changes feedback modality (e. g., from visual to subtle movement-based cues). Reinforcement-based scoring adjusts nudge timing, AIming to reinforce successful behavior change without triggering alert fatigue.

Grounding AI in Behavioral Science: COM-B, BCW and SDT

Isa's design is grounded in established behavioral science models such as COM-B (Capability, Opportunity, Motivation – Behavior), the Behaviour Change Wheel, and Self-Determination Theory (SDT). It promotes not only behavioral capability and opportunity but also supports intrinsic motivation by reinforcing users' sense of autonomy and competence. These models have been shown to significantly improve outcomes in digital and workplace health interventions (Cox et al., 2023; Ross & Barnes, 2018; Sheeran et al., 2020). An interaction concept based on Self-Determination Theory ensures sustAIned engagement and supports the development of lasting micro-habits by adapting to individual user behavior, offering a high degree of personalization, and reinforcing progress and competence growth – while simultaneously preventing overstimulation or habituation.

Efficacy in Practice: Results from Real-World Deployments

Isa has been deployed in over 200 organizations – including E.ON, Deutsche Bahn, and Deutsche Rentenversicherung – across Germany, the Netherlands, Austria, Switzerland, and Belgium – with consistently positive results (Deep Care, 2024). In addition, Isa is already being utilized and supported by several German health insurance providers – including Techniker Krankenkasse (TK) and AOK – as part of workplace health promotion initiatives.

A scientific study conducted by Heidelberg University Hospital confirmed Isa's effectiveness in reducing long periods of uninterrupted sitting at work and supporting healthier routines through unobtrusive digital interventions (Profit, 2024). In a 12-week randomized controlled study with 43 participants, the intervention group using Isa showed a significant increase in movement breaks ($p \leq 0.05$, Cohen's $d = 1.58$) and reported improved general health perception ($p \leq 0.05$, Cohen's $d = 0.74$), while the control group saw no comparable change.

In parallel, research at the Technical University of Munich demonstrated significant positive effects on musculoskeletal health and productivity among office workers (Liang et al., 2025). A longitudinal field study across 2,325 office workers from 50 companies demonstrated impressive outcomes: participants experienced 56 % fewer musculoskeletal related absences, 58 % higher productivity, 88 % increased ergonomic awareness, 97 % more frequent movement breaks, and 56 % higher fluid intake – all with large effect sizes (Cohen's $d > 0.8$).

Technology and Ethics: Privacy, Trust and Acceptance

From a technological perspective, Isa's AI architecture, embedded sensor systems, and privacy-by-design principles ensure that no personal data is processed. No cloud services are involved, and all data remAIns local. Instead of camera-based monitoring, Isa uses depth and ambient sensors to ensure anonymous, fine-grAIned behavior tracKIng.

Furthermore, the system has been designed to reduce risks such as techno-stress, perceived surveillance, or interface fatigue. These considerations are embedded in Isa's interaction design: visual cues are minimalistic, interruptions are context-sensitive, and users can manually pause or adjust interaction frequency. The goal is to create trust, ensure long-term acceptance, and integrate health promotion into users' everyday routines.

Developing Isa required solving key challenges in embedded AI. These included achieving real-time inference on low-power microcontroller hardware, ensuring robust posture recognition across varied user physiologies and lighting conditions, and minimizing false positives in depth-based classification. Additionally, personalization of feedback logic for different workplace contexts and individual interaction styles remAIns an open area for further AI research.

Conclusion: Toward a Proactive and Human-Centered Health Promotion Paradigm

In conclusion, the contribution advocates for a paradigm shift in health promotion – from static, punctual or app-based solutions toward interactive, context-aware AI systems that are proactive, effective, and human-centered. When grounded in behavioral science and ethical technology design, such systems can significantly enhance participation, perception, and measurable outcomes in workplace health promotion.

References

Amagai, S., Pila, S., Kaat, A. J., Rausch, J., & Thumm, E. (2022). Challenges in Participant Engagement and Retention Using Mobile Health Apps: Literature Review. Journal of Medical Internet Research, 24(4), e35120. *https://doi.org/10.2196/35120*

Bensa, K., & Širok, K. (2023). A Scoping Review of Participation Rates in Workplace Health Promotion Programs. International Journal of Environmental Research and Public Health, 20(3), 2757. *https://doi.org/10.3390/ijerph20032757*

Boucher, E. M., Pbert, L., Williams, J., & Bartels, S. J. (2024). Engagement and Retention in Digital Mental Health Interventions: A Narrative Review. BMC Digital Health, 1(1). *https://doi.org/10.1186/s44247-024-00105-9*

Cox, J. S., Burgess, N. A., & Garnett, N. (2023). Integrating COM-B and the Person-Based Approach for Adolescent Weight Management. BMC Health Services Research, 23, 1158. *https://doi.org/10.1186/s12913-023-09930-6*

Deep Care. (2024). Real-world effectiveness of Isa across 50+ organizations in industry, services, finance, research, and healthcare. Internal evaluation report (Q4 2023–Q4 2024).

Liang, Q., Yang, J., Nasseri, M. A., & Zapp, D. (2025). The Impact of Deep Care Isa on Reducing Musculoskeletal Disorders and Enhancing Productivity Among Office Employees: A Comprehensive Study [Manuscript submitted for publication]. Klinikum rechts der Isar, Technical University of Munich (TUM), Germany.

Michie, S., van Stralen, M. M., & West, R. (2011). The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions. Implementation Science, 6, 42. *https://doi.org/10.1186/1748-5908-6-42*

Milne-Ives, M., Lam, C., De Cock, C., Van Velthoven, M. H., & Meinert, E. (2020). Mobile Apps for Health Behavior Change: Systematic Review. JMIR mHealth and uHealth, 8(3), e17046. *https://doi.org/10.2196/17046*

Profit, F. (2024). Evaluation of the Effectiveness of the Intelligent Seating Assistant (Isa) in Reducing Long Periods of Sitting at Work – A Quantitative Survey [Poster, published by Heidelberg University Hospital for Deep Care]. Corporate Health Management Unit, Heidelberg University Hospital.

Robroek, S. J. W., van Lenthe, F. J., van Empelen, P., & Burdorf, A. (2009). Determinants of Participation in Worksite Health Promotion Programmes: A Systematic Review. International Journal of Behavioral Nutrition and Physical Activity, 6, 26. *https://doi.org/10.1186/1479-5868-6-26*

Ross, B. M., & Barnes, D. M. (2018). Self-Determination Theory With Application to Employee Health Settings. Workplace Health & Safety, 66(8), pp. 367–372. *https://doi.org/10.1177/2165079917749863*

Schulze, M., Kopainsky, B., Klöckner, C. A., & Egermann, M. (2024). How to Move the Transition to SustAInable Food Consumption Towards a Societal Tipping Point. Technological Forecasting and Social Change, 189, 122336. *https://doi.org/10.1016/j.techfore.2024.123329*

Sheeran, P., Klein, W. M. P., & Rothman, A. J. (2020). Self-Determination Theory Interventions for Health Behavior Change: Meta-Analysis. Journal of Consulting and Clinical Psychology, 88(8), pp. 726–737. *https://doi.org/10.1037/ccp0000501*

Artificial Intelligence Tools in Algorithmic Management: Towards Understanding Digital Sovereignty of Employees

Martin Brenzke
Artificial Intelligence in Work Systems, Federal Institute for Occupational Safety and Health

Introduction

With the increasing availability of process and personnel related data in the world of work, practices of algorithmic management (AM) have become and still are becoming more and more widespread. A broad definition of AM refers to all instances in which algorithms take on managerial functions such as allocation of work tasks or monitoring and evaluating workers [1]. According to [2], AM is typically characterized by, among others, surveillance of workers and large-scale data collection. Simultaneously, the proliferation of approaches from the field of artificial intelligence (AI) in the world of work also facilitates an increase in AI-based AM systems and affects employees, e. g., by the collection of private and sensitive data or increased performance pressure through monitoring and evaluation of workers [3].

According to Annex III of the EU AI Act such AM systems incorporating AI are categorized as potentially high-risk AI systems [4] and are as such subject to increased scrutiny. Handling sensitive employee data in such systems in a responsible manner is of utmost importance to guarantee employee safety, health and digital sovereignty.

As stated in [3], even the safeguards provided by the General Data Protection Regulation (GDPR) might not suffice to protect workers' privacy as the GDPR does not focus on workers' privacy rights and employees might be coerced into consenting to data collection out of fear of losing their jobs.

However, it is currently unclear what AI approaches and underlying machine learning (ML) models are in use for what use cases of AM and to what extent employee data is used in these cases and, thus, how these AM systems affect employees and their privacy rights. Therefore, more detAIled insights into these topics are needed.

Background and Related Work

A recent study conducted by the OECD [5] Aimed at increasing the data basis on AM via a survey among 6000 mid-level managers from Germany, France, SpAIn, Italy, Japan and the United States. To this end, the study divided AM tools into three categories, "Instruction", "Monitoring" and "Evaluation" and took a very broad approach to what is considered AM (e. g., software collecting worKIng times).

The study found that in Europe the prevalence of AM tools is significant but they are mostly used for the purpose of instructing employees or in a basic monitoring way (i. e., tracking work times). Hence, these systems might not require sensitive employee data.

On the other hand, the study reports that the majority of participants from Europe reported that the AM tools used by them would not require any data on individuals at all which is in stark contrast to how monitoring systems commonly work. According to the study, this might be caused by the participants not knowing any better or trying to give an answer compliant with the GDPR. In any case this obfuscates the actual use of employees' data in AM and necessitates a more detAIled study on the use of employees' data.

Furthermore, in recent studies ([3] and [5]) it remains unclear to what extent AI approaches are used for which AM use cases, further hindering an evaluation of how employee data are used and processed. Insights into what ML models and approaches are in use for which AM applications are needed to understand where potential technical risks to employees' privacy lie and how to mitigate them.

On top of that, a lack of knowledge on how AM systems work, how they process employees' data and how to asses such systems can also create a further power imbalance between employees and employers [2]. Thus, a more detAIled view on the models being used and how the employees' data is being processed is needed to evaluate if workers' privacy rights are at risk and how to mitigate such potential risks when designing and using AM systems incorporating AI approaches.

The OECD study [5] also found that workers often times cannot opt out of the collection and processing of their data for AM uses, as was already indicated by [3]. Moreover, the study reports that accessing their data is sometimes not possible for workers and that requesting corrections of that data is also not possible in some cases. It is hypothesized these findings might be due to the fact that the managers who participated in the study simply were not aware of employees' options to opting out of data collection and to access or correct their data. Nevertheless, improving transparency of AM processes and the underlying technical aspects and granting workers an extended right of co-determination would be desirable as well in order to mitigate negative effects of AM on employees. This is especially relevant for AI-based AM systems as these tend to be black-box models. To this end, again, it is paramount to identify what AI/ML models and techniques are in use in the field of AM.

Finally, [5] reports that guidelines on the use of AM are the most commonly implemented method of governance in companies using AM. Still, less than 50 % of the studies' participants from German companies with AM is in use reported that guidelines on the use of AM were in place in their organizations. Thus, it would be beneficial to adopt a more standardized approach to these guidelines to facilitate a wider adoption and set clearer rules on AM and the use of employees' data in that context.

Study

To tackle missing information on the aforementioned details on the use of AM, the processing of employees' data and the use of AI in this context, we plan on carrying out an explorative study on the German market. In several one-on-one interviews with relevant stakeholders, such as HR- and IT-departments or data protection officials from mid-sized to large companies in Germany, we want to shed some light on the more detailed questions of:

1. In what use cases of AM is the use of AI relevant?
2. What data on employees is used in AI-based AM?
3. Which of these data categories (might) contAIn sensitive information on individuals?
4. What corporate guidelines do already exist on the use of AM and AI?
5. What ML models are used in AM, e. g., decision trees, neural networks, etc.?

The first four research questions can possibly be elaborated with representatives from organizations where AM is in use. However, to answer research question five, the inclusion of providers of AM tools is necessary to gAIn some insight into what classes of ML algorithms are being used for the AM use cases identified. From the results of these interviews, we plan on further determining to what extent employees' privacy rights might be at risk due to AI-based AM and give recommendations on how to ascertain employees' privacy.

The overall structure of the study and future goals are depicted in figure 1. The gathered information on use cases of AI-based AM, the underlying ML models and employees' data used will help asses and mitigate risks to employees' privacy rights in the future.

By assessing existing corporate guidelines on the use of AM and AI, we hope to be able to deduce overarching topics and suggestions that might be applicable to a broad range of AM scenarios such that adoption of AM guidelines by further organizations is encouraged to set clearer rules and boundaries on the use of AI-based AM.

The types of ML models in use are highly relevant to further determine pathways to improve the security of employees' privacy in the context of AM and currently pose blind spots when it comes to investigating AM of workers. The needs for and means to secure employees' privacy strongly depend on what data is being used. If no (sensitive) data on individuals is used for AM purposes, the employees' privacy might already be sufficiently secure. However, if data on individuals is used, even if it is not obviously sensitive, employees' privacy has to be protected. Especially so, since allegedly non-sensitive data could still be correlated with sensitive information. Through this avenue employees' privacy rights could be violated indirectly. Moreover, if the outputs of an AI-based AM system are based on such data, this might lead to discrimination based on sensitive information on employees by proxy information.

The types of ML models used in AM are relevant in order to determine how the data is processed and to design and evaluate technical approaches to securing employees' privacy rights. For example, safeguarding a decision tree-based AM system is tentatively easier than safeguarding highly complex neural network architectures inferring higher-order dependencies from the data fed to the model.

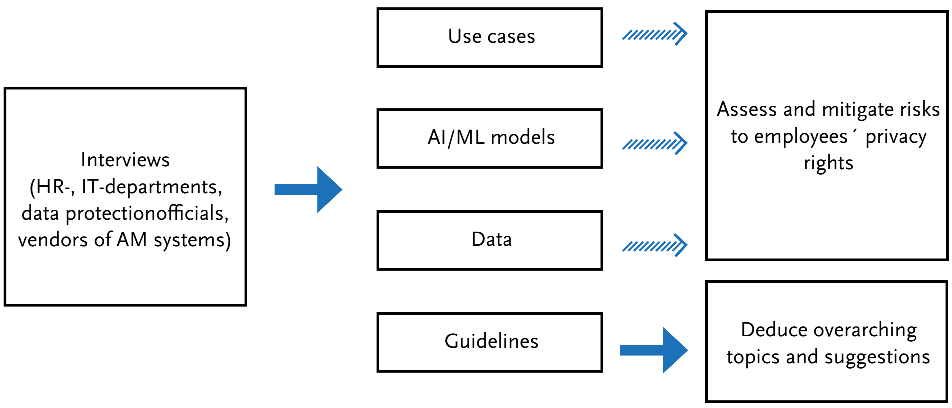

Fig. 1 Overall structure of the described study and future goals. Information from interviews with relevant stakeholders on AI-based AM use cases, underlying ML models and employees' data in use will help mitigating risks to employees' privacy rights in the future. Extraction of overarching topics and suggestions from existing guidelines on AI and AM will foster adoption of such guidelines. Dashed arrows indicate future goals.

Conclusion and Outlook

As discussed above, more detailed information on the use of AI and employees' data in AM is needed to assess risks to and ascertAIn employees' privacy rights. The outlined study is designed to garner such detailed insights. Depending on the results of this study, potential further research could branch into algorithmic approaches to securing employees' digital sovereignty. One potential avenue could be to investigate possibilities of machine unlearning, i. e., removal of certain data or information from AI models without fully retraining the models (see e. g., [6]). This could be especially relevant with regard to concerns mentioned in [5] that correction of data used for AM might not always be available or possible. This problem is exacerbated in AI-based AM tools trained on employee data, even more so if a full retraining of the underlying ML models is not feasible. In such cases employee data remains baked into the AM system without straightforward ways of correction.

For this, investigating what influence certain data has on the models underlying AM applications is crucial which in turn, again, requires the knowledge about what ML model types are used in AM.

References

[1] Lee, M. K., Kusbit, D., Metsky, E., Dabbish, L. (2015). Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers. CHI '15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 10.1145/2702123.2702548.

[2] Mateescu, A., Nguyen, A. (2019). Explainer: Algorithmic management in the orkplace. Data & Society Research Institute.

[3] Reinhold, K., Järvis, M., Christenko, A., JankauskAItė, V., Paliokaitė, A., Riedmann, A. (2022). Artificial intelligence for worker management: implications for occupational safety and health. European Agency for Safety and Health at Work.

[4] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

[5] Milanez, A., Lemmens, A., Ruggiu, C. (2025) Algorithmic management in the workplace: New evidence from an OECD employer survey. OECD Artificial Intelligence Papers (31).

[6] Xu, J., Wu, Z., Wang, C. (2024). Machine Unlearning: Solutions and Challenges. IEEE Transactions on Emerging Topics in Computational Intelligence. pp. 1-19. 10.1109/TETCI.2024.3379240.

Requirements and Demands for a Sensor Network for AI-Assisted Risk Assessment in Occupational Safety and Health

Arn Dietz
Artificial Intelligence in Work Systems, Federal Institute for Occupational Safety and Health

Introduction

Modern technology enables us to measure environmental parameters with relatively high quality at low cost (low-cost high-volume devices). In combination with state-of-the-art hardware, huge streams of sensor data can be processed in real time. This opens up the opportunity to use continuously generated sensor data, like the data from smart buildings, and process it using algorithms from the fields of data science and machine learning to distill hints that Aid in the assessment of risks at workplaces. The long-term Aim of our project is to implement a prototypical system that gathers sensor data and automatically computes hints that Aid in risk assessment. In this work we present our research on the requirements for a stationary sensor network to capture environmental parameters for said system. We also show use cases for methods from data science and machine learning in the domain of occupational safety and health (OSH).

Assessing Requirements and Demands

To assess the requirements and demands the sensors and the software in the project have to meet, we conducted semi structured interviews (Mayring, 2019). Experts within the federal institute for occupational safety and health (BAuA) who are either professionals in risk assessment, technicians with high expertise in capturing environmental parameters for OSH, or researchers who are investigating the impact of different environmental parameters in OSH have been interviewed.

We designed an interview guide, featuring a brief description of the Aim of our project and 21 questions regarding requirements and demands. Before conducting the interviews, the guide was tested in interview dry runs and refined and updated as interviews continued. In the beginning, we interviewed experts we knew in the institute. Later on we could broaden the scope to experts that were recommended to us in earlier interviews. Altogether, we conducted 18 interviews.

In the following sections we present our main findings from the interviews. We give an overview which environmental parameters are relevant and where potential for machine learning lies to extend or compliment classical measurements and methodology used in risk assessment.

Environmental Parameters

Figure 1 shows a taxonomy of all the environmental parameters that emerged from the interviews. The graphic also color codes the parameters for which low-cost high-volume sensors are available (green) and for which devices that are capable of measuring them in an automated, convenient fashion exist (yellow). We give a short list of all the environmental parameters, their special role for OSH and specialties involved in their measurement.

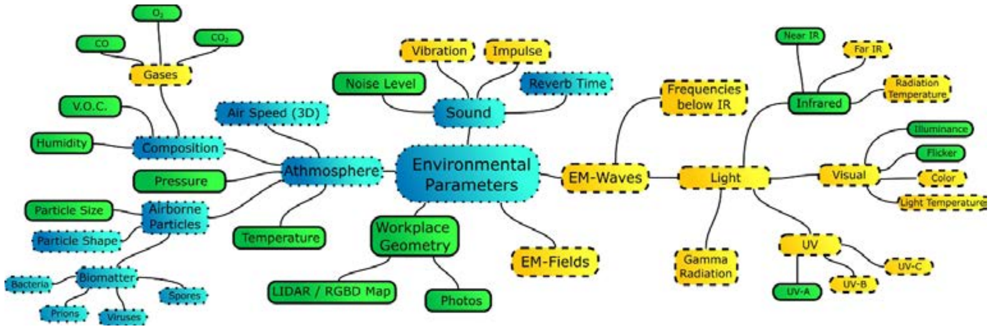

Fig. 1 A Taxonomy of Environmental parameters. Green items (solid outline) indicate parameters where low-cost high-volume sensors are available, yellow items (dashed outline) indicate parameters that can be measured automatically but require specially calibrated or expensive equipment. Blue items (dotted outline) indicate parameters that are only measurable with considerable human effort, either while gathering the data or during the post processing of raw data because the parameters are not directly measurable.

Overview of Environmental Parameters:

Temperature, Airspeed and Humidity: These Parameters are often evaluated together. Whether a workplace feels too cold or too warm depends on all three. The Airspeed is also an interesting parameter for the modeling of the spread of Airborne particles. This modeling is highly computationally intensive and would require vast amounts of data on the geometry of the space and measurements of the Airspeed in multiple spots. Therefore, it is not realistically obtainable outside laboratory conditions.

Gases: All gases need specialized detectors, one per gas. Low-cost high-volume detectors are available for the gases that are commonly found in the Air like CO_2 or O_2. Detecting the concentration of gases is not only relevant for harm reduction but also to assess workplace quality. CO_2 concentrations for example can be used to indicate the "freshness" of Air.

Volatile Organic Compounds V.O.C.: Chemical organic substances that become Airborne at room temperature e. g. acetone. This is a vast group of chemical substances that is often loosely defined. There are V.O.C. sensors at various price points. It is hard to evaluate how suited they are for a given task, since the class of V.O.C. is so vast. V.O.C. are interesting to look at because they can hint towards malicious Airborne substances.

Airborne Particles: The size of Airborne particles is the parameter that has the most influence on where they will sediment in the human body. For small particles that end up in the lung, their stiffness and shape dictate if the human body is capable of removing them on its own without harm. Hence, to assess if Air born dust is harmful to the human body, its chemical composition and shape needs to be known. The state-of-the-art method to examine these properties is to apply particles on a filter and observe its using a scanning electron microscope. This way, malicious particles can be classified manually. The only particle sensors that are available as low-cost high-volume sensors are optical particle counters. These can only measure the particle count in specific size ranges. PM10 e. g. denotes the particles that are smaller than, or equal to 10 µm in diameter.

AIrborne Biomatter: These are viruses, bacteria, spores and prions. Prions are proteins that can harm human cells by e. g. dissolving their cell membrane due to their specific geometry. To identify either of these, extensive measurements need to be conducted.

Sound: Noise level, the impulse a soundwave carries, and vibration can all harm humans if they are exceeding certain values. While there exist low-cost high-volume sensors to measure noise level, the impulse of a soundwave and the measurement of vibration requires specialized equipment.

In contrast to the before mentioned parameters, the reverb time is a more abstract acoustic property of a room.
It is a measure of how long sound reverberates within a room until the sound is dampened below a certain threshold. For workplaces, this value should not be too high, so that sound is not accumulating too much, resulting in a high noise level. Additionally, sound has an impact on cognition and focus. Human speech for example distracts us.
To assess how distracting the sound scape of a place is could be (if it is even possible) an interesting task for machine learning and data science research as well as psychology.

Workspace Geometry: There are multiple rules for the design of a save workspace that involve the spatial layout of the place. Pathways need to be wide enough to evacuate the building in an emergency, cables at workplaces should be managed to minimize trip hazards and heavy objects should not be stored on top of furniture above a certain height to minimize fall hazard. 3D-Scanning devices like LIDAR or RGBD cameras in combination with point cloud object recognition could automatically check for those hazards.

EM-Fields: Electric and magnetic fields only introduce hazards if they exceed specific values. The areas of concern are usually localized, and workplaces where they occur have standard operating procedures and specialized measurement equipment to deal with them.

EM-Waves: EM-waves include the whole spectrum from low frequent fields, radio, gigahertz and mm-wavelength, far and near infrared, visual light, ultra violet to A, B and C, and gamma radiation. How susceptible a detector is to EM-waves of a certain wavelength depends on its construction. In most areas of the spectrum, the dependence is so sensitive that the susceptibility varies considerably among detectors of the same type, even with modern manufacturing and quality control practices. Thus, for equipment to be accurate, it needs to be calibrated. All low-cost high-volume devices measure the illumination integrated over certAIn ranges of the spectrum.
To quantitatively compare them to one another, a calibration is needed. The difficulty of measuring EM-waves varies along the spectrum because different detection principles cover different spectral ranges, with areas in between where no principle is really suited.

Different Frequency Ranges:

- Low frequent EM-fields up until far infrared are only harmful if they exceed certAIn field strengths, e. g. in the vicinity of large antennas. Measuring these fields requires specialized equipment.
- Infrared radiation can harm the unprotected eye because it gets absorbed and thus transfers heat. The point until which the radiation travels is frequency dependent. Near infrared radiation can be detected with low-cost high-volume semiconductor detectors. Far infrared radiation requires special, sometimes cryogenic detectors.
- There are multiple ways in which visual light plays an important role in OSH. First, workplaces need to be illuminated. Depending on the task that should be accomplished at the workplace, different demands for illumination are present. In addition, the spectral composition of the light (e. g. how much blue light is present) has effects on fatigue.
- Ultra violet light can harm the human body. Whereas infrared light harms the eyes by depositing heat, ultra violet light can deposit enough energy immediately at absorption to alter weak chemical bonds, burn the skin and harm the human DNA. On warm sunny days UV light is a thread for people that work outside. There are low-cost high-volume semiconductor detectors for UV-A radiation. The higher the frequency of the UV radiation, the harder it is to detect with conventional semiconductor detectors.

Parameters we decided to detect:

From the present parameters, we chose those, that are measurable with low-cost, high-volume sensors. We furthermore excluded parameters which are highly variable with the location of the measurement. The value of e. g. a temperature measurement can be used as a proxy for the value of the temperature in the vicinity of the measurement. This approximation is doable for most values except the once from the “EM-Waves”-Section.
We thus excluded those.

General Limitations Regarding Stationary Low-Cost High-Volume Sensors

In Interviews, the experts have pointed out some general limitations to the approach of monitoring environmental parameters by using a stationary sensor network consisting of low-cost high-volume devices.

Price VS Performance: The measurements taken with low-cost high-volume sensors tend to lack in precision and accuracy compared to their industrial or scientific counterparts. This trend does not affect all environmental parameters in the same way. There are parameters like temperature and Air pressure where the low-cost high-volume sensor offers remarkable precision and accuracy (f. e. the Bosch BMP280 Sensor (Bosch BMP280 – Data sheet)). On the other hand, there are parameters like the shape of AIrborne particles where there is no automatic inexpensive way of measuring them.

Stationary VS Workplace Centered: Most of the rules and research towards OSH in Germany is focused on the exposure of workers to environmental parameters. Environmental parameters that are captured by a stationary sensor station in the vicinity of the workplace can only be used to approximate the workers exposure. Again, there are environmental parameters like e. g. temperature or noise where the approximation is rather exact and others like e. g. light, which has a high dependence on the location of measurement and the directionality of the sensor.

Problems and Challenges for Machine Learning and Data Processing

One of the maiIn challenges regarding measurements of environmental parameters at the workplace, is to conduct them without disturbing workers and workflow. People change their behavior if the measurement process is invasive (f. e. the Hawthorne effect). It remains an open question if stationary sensors in conjunction with sophisticated data processing can yield accurate approximations for a dosimetric measurement at the workplace, without disrupting the standard condition of the workplace.

Another directly related field of interest is the estimation of the impact of worKIng at a workplace on the long-term development of the workers’ health. “If I keep worKIng at this workplace for 30 years, how will this influence my chances of suffering lung cancer?”. There is no classical statistic approach to these questions since the outcomes to predict are typically influenced by a multitude of factors of which only a few can be measured or controlled for.

Finally, there is the summarization and categorization of events. Risk assessment professionals are usually responsible for several workplaces, so they don’t have much time to spend per workplace. To make the most of continuously measured data, preprocessing is needed. Software that automatically detects events, groups the associated features of the measured parameters and categorizes events could potentially save a lot of time and effort for professionals in the field. This is why our project puts its main focus of this area of interest.

Conclusion:
We have interviewed 18 Experts from the field of applied OSH and OSH research. From the interviews we collected a list of environmental parameters that play an important role in workplace safety and workplace quality. We gave brief information on the role of each of the parameters for OSH and highlighted potential research questions in the fields of machine learning and data science that are associated with them. Finally, we gave our opinion which parameters are the most interesting to measure with stationary low-cost high-volume sensors.

References

Bosch BMP280 – Data sheet. (kein Datum). Abgerufen am 02. Juli 2025 von *www.bosch-sensortec.com/media/boschsensortec/downloads/datasheets/bst-bmp280-ds001.pdf*

Mayring, P. (2019). Qualitative Inhaltsanalyse – Abgrenzungen, Spielarten, Weiterentwicklungen. Forum Qualitative Sozialforschung/Forum: Qualitative Social Research.

AI-Assisted Person Detection in Maritime Search and Rescue: Managing Workload Distribution of Helicopter Crew Operations

Marthe Gruner, Toni Schumacher, Clara Wilk, Angelina Grauerholz, Jan Erik Hannig, Christoph Schönicke, André Calero Valdez
Institute of Multimedia and Interactive Systems, University of Lübeck

Airborne search and rescue (SAR) missions impose considerable cognitive and attentional demands [1], [2]. Operations require members of the Search and Rescue Unit (SRU) to simultaneously manage, e. g., flight operations, communications, risk assessment, coordination of search patterns, mAIntAIning situational awareness, and visual searches for people in distress [1], [3]. As part of the SRU, the Tactical Coordinator (TACCO) acts as On-scene coordinator (OSC) in sole operations and is an important link to the Rescue Coordination Center (RCC [3], [4], [5]). The TACCO assures on-site situation assessment and mAIntAIns tactical communication. Concurrently, they operate and monitor advanced tactical systems with relevant sensors and navigation systems to generate a real-time situational picture, and plan and coordinate SAR strategies. Overall, this role ensures an effective and efficient use of operational resources to find the person in distress.

According to Multiple Resource Theory [6], human information processing capacity is distributed across different modalities (e. g., visual, auditory), processing levels (e. g., perceptual, cognitive, responsive), and processing codes (e. g., spatial, verbal/linguistic). How information is distributed shapes cognitive workload. Moreover, workload increases when multiple tasks are performed simultaneously, when a single task becomes more complex, when the bandwidth of a tracking task grows or when the number of targets to be monitored increases. Such demands are characteristic of maritime SAR operations. TACCOs must manage multiple tasks in parallel, process complex information streams, and remain vigilant when searching for people in challenging maritime environments. Visual search is constrained by these cognitive bottlenecks, which poses limitations on search effectiveness. These limitations can be mitigated by automatically identifying targets in complex environments by employing AI-based detection systems. For optimal system implementation, however, the resources most heavily stressed during mission-critical phases of AIrborne maritime SAR must be identified to characterize workload bottlenecks. Our research AIms to pinpoint these bottlenecks to develop a targeted system that strategically reallocates cognitive demands and ultimately improves both operational performance and occupational health.

Hierarchical Task Analysis to Capture Workload Demands

To assess the workload of Airborne maritime SAR missions during the search for people, we conducted a Hierarchical Task Analysis (HTA) based on Stanton [7]. In collaboration with maritime SAR experts, this approach allowed us to map current task flows, identify workload bottlenecks, and highlight cognitively demanding phases, particularly for the TACCO role. The HTA followed a structured, multi-stage process to ensure a comprehensive and operationally valid task model, including expert interviews and workshops, document analysis, and in-situ observation. The analysis insights inform the design of user-centered AI-support for object detection.

In sum, the analysis showed that operators manage and frequently switch between multiple competing tasks, with especially high demands on visual resources, creating potential performance bottlenecks. TACCOs must coordinate communications, monitor and prioritize sensor data, plan search patterns, maintain spatial awareness, direct the crew, track environmental conditions, log mission activity, and conduct visual searches, all under time pressure and uncertain conditions, causing high workload. This high workload often forces operators to prioritize flight safety and navigation tasks at the expense of effective visual search performance, potentially compromising SAR effectiveness.

Visual search itself is exacerbated by the 'low frequency problem' (i. e., critical targets are more likely to be overlooked due to low frequency) and rapid passing of potential targets through the operator's narrow field of view. Furthermore, maritime conditions such as glare from sunlight, wave crests (white caps), fluctuating light conditions, and swell variations add to the strain on perception and reduce the target visibility. Visual perception challenges are amplified by prolonged missions (i. e., up to eight hours) as vigilance declines and the likelihood of successful detection decreases over time.

Design of AI-Based Detection Support

To mitigate the aforementioned challenges, automated detection systems can support human performance and reduce cognitive load [8] by integrating visual cues. Research shows that visual cues in scanning tasks can enhance decision accuracy and reduce operator effort, especially when those cues are highly reliable [9]. Furthermore, automation has been shown to improve task performance by reducing interference among concurrent tasks, thereby enabling more efficient resource allocation [10].

However, the benefits of automation are highly dependent on its design. Poorly implemented systems may introduce new cognitive demands, such as the need to monitor automated functions or the risk of automation bias [11]. System failures in target detection can lead to mis- or disuse of the system [12], [13]. Additionally, the structure and clarity of the user interface (UI) itself play a crucial role in shaping visual search and decision-making strategies [14]. Our system requirements, derived from the HTA findings, include:

Adaptive User Interface: A responsive UI is required that modifies its functionalities and content presentation based on behavioral patterns and contextual factors, improving usability, accessibility, and overall user experience (UX) in all possible interaction scenarios.

Synchronous and Asynchronous Annotations: The system should support both real-time (synchronous) and time-delayed (asynchronous) annotations of detected areas to enable review of flagged areas.

Context-sensitive and minimally intrusive cue integration: AI-generated cues, visual overlays and warnings should be contextually adapted to support user judgment without obscuring essential mission data or straining perception.

Transparent Detection Feedback: AI detections should be explainable, offering justifications (e. g., detection confidence, sensor source) to support operator understanding and informed intervention when necessary.

User feedback mechanisms: Users should be able to provide feedback on accuracy and relevance of object detection results to enable iterative model refinement and re-training, improve detection performance over time, and foster system transparency and user confidence.

Compliance with tactical systems: To ensure system interoperability, real-time data exchange and compliance with key operating standards for maritime SAR missions, the embedded system should integrate seamlessly into the existing helicopter infrastructure.

Modular Architecture: Given rapid developments in Computer Vision, a flexible system architecture consisting of interchangeable components is required to support the independent development, testing and possible maintenance of individual software and hardware modules, thereby simplifying system upgrades.

Anticipated Occupational Health Benefits

The integration of AI-assisted visual search support in Airborne maritime SAR operations is expected to yield several occupational health benefits for SRUs: (1) By assisting with visually demanding tasks, the system helps decrease overall cognitive workload during operations, allowing operators to allocate attention more efficiently. (2) Continuous AI support can help counteract the natural decline in vigilance during prolonged operations, maintain detection performance over time, and potentially reduce frustration. (3) With reduced task load, operators can more effectively maintain a comprehensive understanding of their operational environment, leading to improved situational awareness (SA). (4) Lower cognitive demands and better task support contribute to reduced operational stress, promoting safer and more sustAInable worKIng conditions for mission personnel.

Methodological Approach

Training data and study stimuli: Acquiring real-life examples of people in distress at sea for AI training and experimental testing poses an ethical and practical challenge. Methodology: We will use open source real-live imagery, if possible, [15], and Unity-based simulation environments [16] to generate high-fidelity synthetic imagery under varying environmental conditions (e. g., sea state, lighting, weather) and target characteristics. These datasets will be used both for AI model (i. e., YOLO [17]) training and controlled experimental testing.

Workload: The introduced system may shift or add to cognitive workload, especially during system monitoring. Methodology: Both subjective measures (e. g., NASA-TLX [18]) and objective metrics (e. g., dual tasks, secondary tasks) are utilized to comprehensively assess cognitive load.

Situation Awareness: Maintaining SA is essential when introducing automated systems. Methodology: We will employ multiple validated objective and subjective assessment techniques (e. g., SAGAT [19]) to evaluate how the AI system impacts different levels of SA.

Trust Calibration and Automation Bias: Adequate trust is essential, as both over- and under-trust in AI can compromise effectiveness, leading to omission or commission errors [8].

Methodology: Simulations include various AI reliabilities across scenarios of diverse complexity, including instances with diagnostic errors (i. e., false alarms and failed attempts). We evaluate trust using subjective measures (e. g., FMV [20]) along with behavioral indicators (i. e., reliance and compliance [21]).

Field Validation: The AI-based system must ultimately prove its effectiveness in real-world environments where unpredictable variables, environmental conditions and team dynamics can significantly impact performance. Methodology: In field tests, the performance of the AI-based detection system and the impact on operational demands will be evaluated in exercises with experienced maritime SAR crews under real operational conditions.

Iterative Prototyping and Validation: All development phases follow the user-centered design methodology [22] incorporating iterative design, testing, and refinement cycles.

Conclusion

Building on our completed HTA and current prototype development, this research demonstrates how empirical workload analysis can inform the systematic design of AI-assistance systems for high-stakes operational environments. The complexity and workload demands of Airborne maritime SAR missions, especially for TACCOs, highlight the urgent need for intelligent support systems to reduce cognitive load, improve search performance, and ensure occupational health. Based on a HTA, we identified the main workload bottlenecks and can translate these findings into concrete requirements for the development of an AI-supported person detection system. The proposed system aims to assist the operator through adaptive, transparent, and context-aware automation while maintaining critical situational awareness and trust. Our multi-method evaluation framework ensures that the system is ethically tenable, operationally valid, and ergonomically robust.

Broader Impact and Transferability: The AI-Aided detection system and human factors methodology developed for maritime SAR operations offer strong potential for transfer to related domains such as land-based search and rescue, aerial surveillance and emergency coordination. While domain-specific adaptations will be required for target detection algorithms and interface design, the core principles of cognitive workload redistribution, transparent AI feedback and trust calibration provide a generalizable scope for human-AI collaboration in safety-critical applications. The HTA-based design methodology provides a transferable approach for identifying workload bottlenecks and developing targeted AI support systems in different operational contexts.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT for text revision, paraphrasing, rewording, grammar and spelling checks, as well as for generating stimuli. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Hamlet, O. E. D., Irwin, A., & McGregor, M. (2020). Is it all about the mission? Comparing non-technical skills across offshore transport and search and rescue helicopter pilots, The International Journal of Aerospace Psychology, Bd. 30, Nr. 3–4, S. 215–235, Okt. 2020, *doi: 10.1080/24721840.2020.1803746*.

[2] Meadley, B., Bowles, K.-A., Smith, K., Perraton, L., & Caldwell, J. (2021). Applied Ergonomics, Bd. 93, S. 103375, Mai 2021, *doi: 10.1016/j.apergo.2021.103375*.

[3] IMO und ICAO, Hrsg., International Aeronautical and Maritime Search and Rescue Manual: IAMSAR Manual. Volume III: Mobile Facilities, 12. Aufl. in IMO Publication. London: International Maritime Organization, 2022.

[4] IMO und ICAO, Hrsg., International Aeronautical and Maritime Search and Rescue Manual: IAMSAR Manual. Volume II: Mission Coordination, 9. Aufl. in IMO Publication. London: International Maritime Organization, 2022.

[5] IMO und ICAO, Hrsg., International Aeronautical and Maritime Search and Rescue Manual: IAMSAR Manual. Volume I: Organization and Management, 12. Aufl. in IMO Publication. London: International Maritime Organization, 2022.

[6] Wickens, C. D. (2008). Multiple Resources and Mental Workload, Hum Factors, Bd. 50, Nr. 3, S. 449–455, Juni 2008, *doi: 10.1518/001872008X288394*.

[7] Stanton, N. A. (2006). Hierarchical task analysis: Developments, applications, and extensions, Applied Ergonomics, Bd. 37, Nr. 1, S. 55–79, Jan. 2006, *doi: 10.1016/j.apergo.2005.06.003*.

[8] Mosier, K. L., Manzey, (2029). D. Humans and Automated Decision AIds: A Match Made in Heaven?, in Human Performance in Automated and Autonomous Systems, 1. Aufl., M. Mouloua, P. A. Hancock, und J. Ferraro, Hrsg., Boca Raton, FL.: CRC Press/Taylor & Francis Group, 2019.: CRC Press, 2019, S. 19–42. *doi: 10.1201/9780429458330-2*.

[9] Yeh, M., Wickens, C.D. (2001). Display Signaling in Augmented Reality: Effects of Cue Reliability and Image Realism on Attention Allocation and Trust Calibration, Hum Factors, Bd. 43, Nr. 3, S. 355–365, Sep. 2001, *doi: 10.1518/001872001775898269*.

[10] Dixon, S. R., Wickens, C. D., Chang, D. (2005). Mission Control of Multiple Unmanned Aerial Vehicles: A Workload Analysis, Hum Factors, Bd. 47, Nr. 3, S. 479–487, Sep. 2005, *doi: 10.1518/001872005774860005*.

[11] Parasurama, R., Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration, Hum Factors, Bd. 52, Nr. 3, S. 381–410, Juni 2010, *doi: 10.1177/0018720810376055*.

[12] Lee, J. D., See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance, Human Factors: The Journal of the Human Factors and Ergonomics Society, Bd. 46, Nr. 1, S. 50–80, 2004, *doi: 10.1518/hfes.46.1.50_30392*.

[13] Y. Lu, Y., Sarter, N. (2019). Eye TracKIng: A Process-Oriented Method for Inferring Trust in Automation as a Function of Priming and System Reliability, IEEE Trans. Human-Mach. Syst., Bd. 49, Nr. 6, S. 560–568, Dez. 2019, *doi: 10.1109/THMS.2019.2930980*.

[14] Starke S. D., Baber, C. (2018). The effect of four user interface concepts on visual scan pattern similarity and information foraging in a complex decision maKIng task, Applied Ergonomics, Bd. 70, S. 6–17, Juli 2018, *doi: 10.1016/j.apergo.2018.01.010*.

[15] Martinez-Esteso, J. P., Castellanos, F. J., Calvo-Zaragoza, J., & Gallego, A. J. (2025). Maritime search and rescue missions with aerial images: A survey, Computer Science Review, Bd. 57, S. 100736, Aug. 2025, *doi: 10.1016/j.cosrev.2025.100736*.

[16] Unity. (2022). Unity Technologies, San Francisco, USA. [Online]. Verfügbar unter: *https://unity.com/de*

[17] J. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, Juni 2016, S. 779–788. *doi: 10.1109/cvpr.2016.91*.

[18] Hart, S. G., Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, in Advances in Psychology, Bd. 52, Elsevier, 1988, S. 139–183. *doi: 10.1016/S0166-4115(08)62386-9*.

[19] Endsley, M. R. (2020). The Divergence of Objective and Subjective Situation Awareness: A Meta-Analysis, Journal of Cognitive Engineering and Decision MaKIng, Bd. 14, Nr. 1, S. 34–53, März 2020, *doi: 10.1177/1555343419874248*.

[20] Wiczorek, R. (2011). Entwicklung und Evaluation eines mehrdimensionalen Fragebogens zur Messung von Vertrauen in technische Systeme, in Reflexionen und Visionen der Mensch-Maschine-Interaktion – Aus der Vergangenheit lernen, Zukunft gestalten, S. Schmid, M. Elepfandt, J. Adenauer, und A. Lichtenstein, Hrsg., Berlin, 2011, S. 621–626.

[21] Meyer, J. (2001). Effects of Warning Validity and Proximity on Responses to Warnings, Hum Factors, Bd. 43, Nr. 4, S. 563–572, Dez. 2001, *doi: 10.1518/001872001775870395*.

[22] DIN EN ISO 9241-210:2020-03, Ergonomie der Mensch-System-Interaktion_- Teil_210: Menschzentrierte Gestaltung interaktiver Systeme (ISO_9241-210:2019); Deutsche Fassung EN_ISO_9241-210:2019, 2020. *doi: 10.31030/3104744*.

Behind the Click: How Personality, Experience, and Privacy shape AI-based Learning in the Workplace

Marina Klostermann, Thea Radüntz
Artificial Intelligence in Work Systems, Federal Institute for Occupational Safety and Health

Introduction

The digital transformation of the workplace, driven by factors such as the adoption of artificial intelligence (AI) and the impact of the pandemic, requires new approaches of workplace learning. AI-based learning recommender systems (AI-LRS) have emerged as an efficient alternative to traditional time- and cost-intensive face-to-face trAIning methods by offering personalized learning paths, recommending content, and identifying experts (Wang et al., 2018; Drachsler et al., 2015). By offering personalized learning opportunities, AI-LRS are data-driven and use advanced machine learning algorithms to adapt to both employee learning needs and organizational goals (Geng et al., 2020; Lex & Schedl, 2022). Their functioning depends on the collection and analysis of large volumes of personal data (e. g., online behavior, geolocation, biometrics, socio-demographics), which can trigger privacy concerns that may hinder user acceptance and implementation success (Kobsa et al., 2016). Despite the growing use of AI in workplace learning, little is known about how individual differences, such as demographics, personality, and experience, influence the willingness to use such systems, especially in light of privacy concerns (Bandara et al., 2020). This study addresses this gap by examining how personality trAIts, prior experiences with AI and privacy invasions, and demographics relate to privacy concerns and the intention to use AI-LRS.

While personality research has a long tradition, the connection between personality traits and privacy concerns has only recently gained scholarly attention. Studies have identified personality as a significant antecedent to privacy concerns, which in turn influence system usage (Buck et al., 2022; Stewart & Segars, 2002). Personality traits refer to consistent patterns in thinking, feeling, and behaving across situations (McCrae et al., 1986). However, not all traits affect privacy concerns equally. Studies have demonstrated that extraversion and neuroticism, two factors from the five-factor model which also includes conscientiousness, agreeableness, and openness, are especially pertinent. Extraversion, associated with sociability and positive emotions, has been linked to lower privacy concerns, although some studies found no significant correlation (Bansal et al., 2016; Korzaan & Boswell, 2008; Roberts et al., 2007). In contrast, neuroticism – linked to emotional instability and anxiety – is consistently associated with higher privacy concerns (Bansal et al., 2016; Korzaan & Boswell, 2008). Based on these findings, we hypothesize that extraverts will show lower privacy concerns (H1) and a higher intention to use an AI-based learning recommendation system (AI-LRS) (H2), while individuals high in neuroticism will exhibit greater privacy concerns (H3) and a lower likelihood of using such systems (H4).

In addition to personality, experience plays a key role. Privacy experience is defined as individuals' prior experiences with privacy violations (Smith et al., 1996). Those with such experiences tend to be more protective of their personal data and more cautious in disclosing information, leading to increased privacy concerns (Smith et al., 1996; Tao et al., 2024). Similarly, experience with AI technologies influences acceptance and usage. Individuals familiar with AI-based tools are more likely to understand how personal data is used to personalize services, which may reduce their concerns and increase their intention to use AI-LRS (Culnan & Armstrong, 1999; Acosta-Enriquez et al., 2024). Familiarity fosters trust and competence, promoting greater engagement with AI systems. Hence, we hypothesize that privacy experience leads to greater privacy concerns (H5) and that previous experience with AI-based technologies leads to greater intention to use the AI-LRS (H6).

Demographics, particularly age and gender, have shown mixed effects on privacy concerns and technology use. Some studies suggest that women and older individuals perceive higher privacy concerns (Youn & Hall, 2008; Hoy & Milne, 2010), while others show that these groups may still disclose more personal information or have higher usage intentions (Van Gool et al., 2015). Given these inconsistent findings, our study AIms to shed more light into the relationship between demographics and privacy concerns. We hypothesize that gender and age have an influence on privacy concerns (H7).

In summary, personality traits, experience with privacy and AI, and demographics are all considered relevant factors in shaping privacy concerns and the intention to use AI-LRS. Furthermore, privacy concerns are proposed as a mediating variable, linKIng individual differences to system usage behavior (H8).

Methods

From September to November 2024, we conducted an online experiment at the Federal Institute of Occupational Safety and Health to test our hypotheses and investigate employees' beliefs and perceptions regarding the implementation of an AI-LRS for workplace learning. A total of 256 employees participated in our study (78 female, 177 male, and 1 diverse). Participants were recruited via the online data collection platform Prolific. The mean age of the participants was 32.83 years (SD = 9.28), and most had a university degree (49.21 %) or a university entrance qualification (29.13 %). The majority of participants were engaged in knowledge-based tasks (53.3%) or person-based tasks (17.2 %). A total of 14.5 % of the other participants were engaged in object-based tasks, while 14.1 % were involved in leadership and management tasks.

Participants were randomly assigned to one of two groups. The experimental group was given autonomy. They could decide which data to share with the recommender system so that it can be analyzed and used to create personalized learning recommendations. In contrast, the control group was not given the autonomy to determine which data to disclose. Participants in this group were prompted to disclose all data required by the system. They were informed that the recommender system would determine which personal data was necessary for further analysis.

The hypothesis testing was based on multiple (H1-H7) and mediated (H8) linear regression models conducted with the open-source program JASP (Version 0.19.3). Initially, the direct effects of individual differences on privacy concerns and on intention to use the AI-LRS were examined. Next, the indirect effect of individual differences on intention to use the AI-LRS, mediated by privacy concern, was investigated through the implementation of mediated regression analyses. The experimental condition was incorporated into the analyses as a potential confounder, and confidence intervals were estimated using the bootstrap method with 5,000 iterations. The alpha level was set to .05.

The study is part of a larger project that was preregistered in the Open Science Framework (OSF) repository (*https://osf.io/hg5nz/?view_only=3fe2de47b4a84589944f064379da8c89*). Ethical approval was received from the review board at Ruhr University Bochum (No. 862), ensuring adherence to all guidelines for research involving human participants. Participants provided informed consent and received a financial incentive of €15. They could withdraw from the study at any time without facing any disadvantage.

Results

Of the personality traits examined, extraversion was found to directly influence the intention to use the AI-LRS ($\beta = 0.14$, $p < .05$, 95 % confidence interval [CI] = -0.01 – 0.28). At the same time, neuroticism was found to have a direct impact on privacy concerns ($\beta = 0.18$, $p < .01$, 95 % CI = 0.04 – 0.31). In consideration of the experience variable, the findings indicate a direct effect of privacy experience on privacy concerns (β = 0.19, $p < .01$, 95 %-CI = 0.06–0.31). Moreover, experience with AI has been shown to predict the intention to use the AI-LRS directly (β = 0.13, $p < .05$, 95 %-CI = 0.01–0.26). The degree to which individuals vary in terms of extraversion and their experience with AI was found to account for 16.3 % ($R^2 = .163$) of the variance in intention to use the AI-LRS. This result indicates that individual differences in extraversion and familiarity with AI have a significant impact on an individual's tendency to utilize the AI-LRS. The results of the study indicated that neuroticism and privacy experience explained 21.5 % of the variance in privacy concerns ($R^2 = .215$). The findings indicated that demographics were not associated with privacy concerns and that privacy concerns did not mediate the effect of individual differences on intention to use the AI-LRS. An overview of the outcomes associated with the hypotheses can be found in Table 1.

Tab. 1 Summary of hypotheses testing

Hypothesis	Confirmed
H1 Extraversion correlates negatively with privacy concerns.	No
H2 Extraversion correlates positively with intention to use the AI-LRS.	Yes
H3 Neuroticism correlates positively with privacy concerns.	Yes
H4 Neuroticism correlates negatively with intention to use the AI-LRS.	No
H5 Privacy experience correlates positively with privacy concerns. H6	Yes
H7 Experience with AI correlates positively with intention to use the AI-LRS.	Yes
H8 Privacy concerns mediate the effect of individual differences on intention	No
to use the AI-LRS.	No

Discussion

In response to global competition and ongoing challenges, companies and HR departments increasingly rely on AI-assisted systems – such as AI-based Learning Recommender Systems (AI-LRS) – to save time and resources. While personalized learning offers significant benefits in the workplace, it also raises concerns about privacy and the amount of personal data collected from employees. These concerns highlight the importance of understanding how such systems are perceived and should be designed in a human-centered way.

Building on existing literature on individual differences and privacy concerns, this study explored how personality traits and prior privacy experiences influence employees' willingness to use AI-LRS. Furthermore, this study examined in line with the conceptual framework from Buck et al. (2022) whether privacy concerns mediate the relationship between individual characteristics and the intention to use AI-LRS.

The research landscape on privacy concern antecedents remains fragmented and lacks consistent replication. However, existing studies suggest that privacy concerns are influenced by factors such as personality, demographics, and previous privacy-related experiences (Buck et al., 2022; Korzaan & Boswell, 2008; Smith et al., 1996; Youn & Hall, 2008). In our present study, individuals high in neuroticism are more likely to worry about data privacy. This can be explained by the fact that individuals high in neuroticism experience more negative emotions and anxiety (Roberts et al., 2007; McCrae & Costa, 2003). Similarly, those with negative privacy experiences tend to be more cautious about data use. This is consistent with the findings of Tao et al., 2024.

Interestingly, extraversion, age, and gender did not significantly impact privacy concerns. While some studies (e. g., Bansal et al., 2016) suggest extraverted individuals are less concerned about privacy, our study did not confirm such findings. However, extraverts were generally more open to using new technologies like AI-LRS, probably due to their social and exploratory nature. Moreover, consistent with Gerber et al. (2018), neither older nor female individuals showed significantly higher privacy concerns in this context. The role of experience with AI technologies cannot be denied: individuals familiar with AI reported a greater intention to use AI-LRS. This suggests that familiarity with AI may reduce apprehension and increase perceived usefulness as stated by Culnan & Armstrong (1999).

Our findings emphasize the need to consider individual differences– especially personality traits and prior experiences – when designing human-centered AI-LRS. For instance, systems designed with transparent data handling and compliance with data protection regulations could reduce the privacy concerns of neurotic users or those with past negative privacy experiences. Empowering users by giving them control over what data they share could also enhance their trust and increase usage intention.

To fully understand the dynamics among individual traits, privacy concerns, and technology acceptance, future research should explore additional personality traits (e. g., agreeableness, openness, conscientiousness) and levels of privacy awareness. This study provides valuable insights for developing human-centered AI-LRS that balance personalization with robust privacy protections, thereby taKIng individual differences into account.

References

Acosta-Enriquez, B. G., Farroñan, E. V. R., Zapata, L. I. V., Garcia, F. S. M., Rabanal-León, H. C., Angaspilco, J. E. M., & Bocanegra, J. C. S. (2024). Acceptance of artificial intelligence in university contexts: A conceptual analysis based on UTAUT2 theory. Heliyon.

Bandara, R., Fernando, M., & Akter, S. (2020). Privacy concerns in E-commerce: A taxonomy and a future research agenda. Electronic Markets, 30(3), 629–647. *https://doi.org/10.1007/s12525-019-00375-6*

Bansal, G., Zahedi, F. M., & Gefen, D. (2016). Do context and personality matter? Trust and privacy concerns in disclosing private information online. Information & Management, 53(1), 1–21. *https://doi.org/10.1016/j.im.2015.08.001*

Buck, C., Dinev, T., & Anaraky, R. G. (2022). Revisiting APCO. In B. P. Knijnenburg, X. Page, P. WisniewsKI, H. R. Lipford, N. Proferes, & J. Romano (Eds.), Modern Socio-Technical Perspectives on Privacy (pp. 43–60). Springer International Publishing.

Culnan, M. J., & Armstrong, P. K. (1999). Information Privacy Concerns, Procedural FAIrness, and Impersonal Trust: An Empirical Investigation. Organization Science, 10(1), 104–115. *https://doi.org/10.1287/orsc.10.1.104*

Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N. (2015). Panorama of Recommender Systems to Support Learning. In F. Ricci, L. Rokach, & B. Shapira (Eds.), Recommender Systems Handbook (second edition, pp. 421–451). Springer. *https://doi.org/10.1007/978-1-4899-7637-6_12*

Geng, S., Tan, L., Niu, B., Feng, Y., & Chen, L. (2020). Knowledge recommendation for workplace learning: a system design and evaluation perspective. Internet Research, 30(1), 243–261. *https://doi.org/10.1108/INTR-07-2018-0336*

Gerber, N., Gerber, P., & Volkamer, M. (2018). ExplAIning the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. Computers & Security, 77, 226–261. *https://doi.org/10.1016/j.cose.2018.04.002*

Hoy, M. G., & Milne, G. (2010). Gender differences in privacy-related measures for young adult Facebook users. Journal of interactive advertising, 10(2), 28-45.

Kobsa, A., Cho, H., & Knijnenburg, B. P. (2016). The effect of personalization provider characteristics on privacy attitudes and behaviors: An E laboration L ikelihood M odel approach. Journal of the Association for Information Science and Technology, 67(11), 2587–2606. *https://doi.org/10.1002/asi.23629*

Korzaan, M. L., & Boswell, K. T. (2008). The Influence of Personality Traits and Information Privacy Concerns on Behavioral Intentions. Journal of Computer Information Systems(48), Article 4, 15–24. *https://doi.org/10.1080/08874417.2008.11646031*

Lex, E., & Schedl, M. (2022). Psychology-informed Recommender Systems Tutorial. In J. Golbeck, F. M. Harper, V. Murdock, M. Ekstrand, B. Shapira, J. Basilico, K. Lundgaard, & E. Oldridge (Eds.), Proceedings of the 16th ACM Conference on Recommender Systems (pp. 714–717). ACM. *https://doi.org/10.1145/3523227.3547375*

McCrae, R. R., Costa Jr, P. T., & Busch, C. M. (1986). Evaluating comprehensiveness in personality systems: The California Q-Set and the five-factor model. Journal of personality, 54(2), 430-446.

McCrae, R. R., & Costa, P. T. (2003). Personality in adulthood: A five-factor theory perspective. Guilford press.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. Perspectives on Psychological science, 2(4), 313-345.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals' Concerns about Organizational Practices. MIS Quarterly, 20(2), 167. *https://doi.org/10.2307/249477*

Stewart, K. A., & Segars, A. H. (2002). An empirical examination of the concern for information privacy instrument. Information systems research, 13(1), 36-49.

Tao, S., Liu, Y., & Sun, C. (2024). Examining the inconsistent effect of privacy control on privacy concerns in e-commerce services: The moderating role of privacy experience and risk propensity. Computers & Security, 140, 103794. *https://doi.org/10.1016/j.cose.2024.103794*

Van Gool, E., Van Ouytsel, J., Ponnet, K., & Walrave, M. (2015). To share or not to share? Adolescents' self-disclosure about peer relationships on Facebook: An application of the Prototype Willingness Model. Computers in Human Behavior, 44, 230-239.

Wang, M. (2018). Emerging Technologies for Workplace Learning. In M. Wang (Ed.), E-Learning in the Workplace (pp. 29–39). Springer International Publishing. *https://doi.org/10.1007/978-3-319-64532-2_3*

Youn, S., & Hall, K. (2008). Gender and online privacy among teens: Risk perception, privacy concerns, and protection behaviors. Cyberpsychology & behavior, 11(6), 763-765.

Artificial intelligence (AI) in occupational medicine – how can we optimize the work of occupational physicians?

Susanne Voelter-Mahlknecht[1], *Felix Leitner*[1], *Leon Liman*[1,2], *Carolin Wienrich*[2]

[1] University Medical Center Göttingen, Institute of Occupational, Social und Preventive Medicine

[2] University of Würzburg, Psychology of Intelligent Interactive Systems

Introduction

The project BAKI (Occupational Health Action: Future-oriented, interdisciplinary and evidence-based with AI) Aims to address a demographic which is often overlooked in occupational health management, digital employees (RobelsKI and Sommer 2020). Digital work is defined as work where information- and communication technologies (ICTs) are integral to accomplish tasks and communicate at work (Czarniawska 2016). This often leads to a physical dispersion since employees no longer need to be present within the same building in order to efficiently collaborate with each other (Hossain and Wigand 2004). While this new type of work has a lot of benefits and brought forth new work arrangements like homeoffice and digital nomadism, it also comes with new challenges, both for employees and companies (Marx et al. 2023). Due to the fact that these employees are often not present in the office, company doctors often have no insight into the working conditions of these digital employees (Wütschert et al. 2022). This is further complicated by the fact that a lot of regulations relating to work and health do not encompass digital work (Bretschneider et al. 2020). This is the gap BAKI is trying to address. To accomplish this, an interdisciplinary approach will be used, which combines the fields of artificial intelligence and occupational medicine. Using interdisciplinary competencies from two universities (University Medical Center Göttingen and the University of Würzburg) two new tools will be developed: A learning algorithm called BAKI-AI, which will provide individualised resource and risk profiles, as well as suggestions for occupational health measures for employees. Secondly, a digital interaction platform called BAKI-social, which will give company doctors a new way to more effectively communicate with employees worKIng from remote locations. The project is funded by the BAuA (Federal Institute for Occupational Safety and Health) as part of the FoGA program (Funding for Research and Teaching on Health in the World of Work). The program focuses on strengthening research related to health and work as well as promoting and furthering the education of young researchers in the field.

Methodology

In order to support digital workers with a fast and efficient tool, a learning algorithm called BAKI-AI will be developed. This algorithm will be able to give its users individual personalised feedback on their worKIng conditions, personal resources, demands and possibly useful occupational health measures. To traiIn the model for BAKI-AI, a large sample of data will be gathered from digital workers in Germany. To make sure that every important facet of digital work is reflected in the data, first an extensive scoping review on the resources and demands which affect the health of employees worKIng in digital environments will be carried out. The review will ensure a literature-based selection of constructs for the data collection. Based on the results from the review, a multimodal framework for the data acquisition will be constructed. This will include questionnaires, smartwatches, room sensors and programs. These will capture multiple psychological and physiological variables like stress, digital competence, heart rate variability, Air quality, the frequency of mouse clicks and many more. Using a multimodal approach will only assure a high level of data quality. And since the data acquisition will be done mostly autonomously after the sensors are set up, this will also take pressure off the employees who are participating in the study. Using the data gathered, the model for BAKI-AI will be trAIned. In order to find the best algorithm on the data, multiple methods from the fields of machine learning and data science will be tested.

This includes more classic models like clustering, support vector machines or random forest models and also more modern approaches like CNN, GAN or AutoEncoder. When training new models' privacy and data security are a high concern. To ensure that there is no possibility of any breaches of privacy, a technique called federated learning will be used. This means that the data will be split up into multiple "nodes" on which local models will be trained. In the next step, these models will be aggregated into a single model. This ensures that the model never has access to the whole dataset. Everything will be hosted on internal servers to comply with German data privacy laws. The data will be anonymised and will not be given out to third parties or employers. Every participant has the option to opt out at any time or to have their data deleted even after the data collection is finished. The system will also be able to provide recommended occupational health measures. These will be gathered separately from interviews with experts in the field of occupational health. This includes different stakeholders like company doctors, employees from health insurance companies, occupational health experts and service providers. These interviews will not only give insight into the goals and effectiveness of these practices. They will also dive into the intended target audience for each measure so they can be linked to the data. After completing the data aggregation and development of the algorithm, it will be ready to be tested. After filling out a shortened version of the questionnAIre, BAKI-AI will provide its users with individualised resource and risk profiles and also recommended occupational health measures. Their effectiveness will be continuously evaluated by the users of BAKI-AI. Experts from the field will be involved in every step of the process to ensure a user-centered process. They will give input on use cases, user design and data security.

Following the development of BAKI-AI, the extended reality interaction platform BAKI-social will be developed. For the development of BAKI-social, the University of Würzburg has extensive capabilities with their XR HUB Würzburg, which offers technologies in the areas of augmented and virtual reality (Fiedler et al. 2025). This includes 3D tracking systems which enable users to control digital avatars in real time and collaborate with each other in digital spaces. These digital environments will be designed in order to maximise the benefits for their users and can range from cartoon-like abstractions to photorealistic environments. The conceptualisation of BAKI-social was left ambiguous by choice. A number of different prototypes will be developed and tested with occupational doctors and simulated patients in order to measure their effectiveness. This will be done using methods of user experience like interviews and task analyses. XR technologies have started gaining traction in medical research, with a number of studies showing their benefits in clinical settings (van der Kruk et al. 2022; Segal et al. 2011). In relation to doctor/patient relationships, XR can help with information visualisation (Li 2019) and spatial awareness (Ghaednia et al. 2021), while being easy to use (Shafi et al. 2020). This is associated with better learning (Meccawy 2022) and a deeper interaction between actors (Zhou et al. 2024). In this regard a social XR platform could help occupational doctors to visualise a healthy worKIng space or highlight problems in a 3D environment in order to better represent issues that people face while working from different locations. Since this interaction can be done from whichever workplace they prefer, it is also more cost-efficient and less disruptive than traditional methods (Esen et al. 2025). While the research about XR technologies is on the rise, there is still a lot of room for further discoveries, especially in the field of occupational health. BAKI-social Aims to address this gap and provide further evidence for the use of XR technologies in occupational health and medicine.

Additionally, a number of work packages will be worked on during the entire duration of the project. The findings of the project will be thoroughly shared with the public through publications, articles and appearances at different conferences. The University Medicine Center Göttigen and the University of Würzburg will also conceptualise new modules in order to bring interdisciplinary competences to their students. This will give the medical students of the Medicine Centre Göttigen an insight into artificial intelligence, while also giving the IT students of the University of Würzburg an insight into occupational medicine. This will hopefully encourage them to further seek out interdisciplinary competencies and spark their interests in new fields of research. An expert panel was put together at the beginning of the project, which includes occupational doctors, AI experts, people with a focus on the ethics of technology and many more. The panel will be consulted at every step of the project in order to bring expertise from different fields and stakeholders into every step of the development and research process. To ensure this, yearly meetings will be held with the full project team and members of the panel, while smaller meetings will be held every couple of months, which will be more catered towards specific topics like the evaluation of BAKI-AI. For individual steps like the data protection concept, specific experts will be contacted bilaterally.

Discussion

BAKI Aims to close gaps in the literature and further the fields of occupational medicine and artificial intelligence. Using innovative methodologies like multimodal data collection, artificial intelligence and virtual/augmented reality, new tools will be developed and tested. As a model project, the goal is to show a path to the future of occupational care and technology. If there is one word that describes the modern work environment, it would be "change". In order to support companies and employees in this culture of change, it is essential to always improve on existing methodologies and to further the development of new ones. This is what BAKI and the FoGA program are trying to address. To not only develop new and innovative solutions but to further the education and interests of a new generation of scientists in order to secure a better and healthier future for the working world.

References

Bretschneider, M., Drössler, S., Magister, S., Zeiser, M., Kämpf, D., & Seidler, A. (2020). Digitalisierung und Psyche – Rahmenbedingungen für eine gesunde Arbeitswelt. Ergebnisse des Projektes GAP. Zeitschrift für Arbeitswissenschaft, 74(2), 63–75. *doi: 10.1007/s41449-020-00206-x*.

Czarniawska, B. (Ed.) (2016). A Research Agenda for Management and Organization Studies: Edward Elgar Publishing.

Esen, Y. E., Kaan Ç., Berk; Y., Kübra; G. G., Hüseyin; Y., U. (2025). XR-Driven Robotic System Training for Occupational Health, Safety, and Maintenance. In IEEE Access 13, pp. 61708–61727. *doi: 10.1109/ACCESS.2025.3556699*.

Fiedler, M. L., Bürger, A., Mittermeier, S., Botsch, M., Latoschik, M. E., Wienrich, C. (2025). Evaluating VR and AR Mirror Exposure for Anorexia Nervosa Therapy in Adolescents: A Method Proposal for Understanding Stakeholder Perspectives. In : 2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). 2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Saint Malo, France, 08/03/2025 - 12/03/2025: IEEE, pp. 965–970.

Ghaednia, H., Fourman, M. S., Lans, A., Detels, K., Dijkstra, H., Lloyd, S. et al. (2021). Augmented and virtual reality in spine surgery, current applications and future potentials. In The spine journal : official journal of the North American Spine Society 21 (10), pp. 1617–1625. *doi: 10.1016/j.spinee.2021.03.018*.

Hossain, L., Wigand, R, T. (2004). ICT Enabled Virtual Collaboration through Trust. In Journal of Computer-Mediated Communication 10 (1), p. 0.
doi: 10.1111/j.1083-6101.2004.tb00233.x.

Li, D. (2019). 5G and intelligence medicine-how the next generation of wireless technology will reconstruct healthcare? In Precision clinical medicine 2 (4), pp. 205–208.
doi: 10.1093/pcmedi/pbz020.

Marx, J., Stieglitz, S., Brünker, F., Mirbabaile, M. (2023). Home (Office) is where your Heart is. In Bus Inf Syst Eng 65 (3), pp. 293–308.
doi: *10.1007/s12599-023-00807-w*.

Meccawy, M. (2022). Creating an Immersive XR Learning Experience: A Roadmap for Educators. In Electronics 11 (21), 3547. *doi: 10.3390/electronics11213547*.

Robelski, S., Sommer, S. (2020). ICT-Enabled Mobile Work: Challenges and Opportunities for Occupational Health and Safety Systems. In International journal of environmental research and public health 17 (20). *doi: 10.3390/ijerph17207498*.

Segal, R., Bhatia, M., Drapeau, M. (2011). Therapists' perception of benefits and costs of using virtual reality treatments. In Cyberpsychology, behavior and social networKIng 14 (1-2), pp. 29–34. *doi: 10.1089/cyber.2009.0398*.

Shafi, K., Lovecchio, F., Forston, K., Wyss, J., Casey, E., Press, J. et al. (2020). The Efficacy of Telehealth for the Treatment of Spinal Disorders: Patient-Reported Experiences During the COVID-19 Pandemic. In HSS journal : the musculoskeletal journal of Hospital for Special Surgery 16 (Suppl 1), pp. 17–23. *doi: 10.1007/s11420-020-09808-x*.

van der Kruk, S. R., Zielinski, R., MacDougall, H., Hughes-Barton, D., Gunn, K. M. (2022). Virtual reality as a patient education tool in healthcare: A scoping review. In Patient education and counseling 105 (7), pp. 1928–1942.
doi: 10.1016/j.pec.2022.02.005.

Wütschert, M. S., Romano-Pereira, D., Suter, L., Schulze, H., Elfering, A. (2022). A systematic review of worKIng conditions and occupational health in home office. In Work (Reading, Mass.) 72 (3), pp. 839–852. *doi: 10.3233/WOR-205239*.

Zhou, D., Deng, Y., Globa, A,, Davies, R., Hu, Y., Wu, R. X. et al. (2024). Social xr: Designing an extended reality application for interaction in long-distance relationships. In Journal of Extended Reality Research 20 (YY), pp. 1–20.

Hyperparameter Optimization for Improved CNN-Based Image Regression of Dermal Exposure in Occupational Settings

Deniz Weißbrodt, Michael Roitzsch, Urs Schlüter, Thea Radüntz
Artificial Intelligence in Work Systems, Federal Institute for Occupational Safety and Health

Abstract

This study investigates the impact of systematic hyperparameter optimization, implemented via the Optuna framework, on the performance of convolutional neural networks (CNNs) for image-based dermal exposure assessment. Using fluorescence images as input data, the study demonstrates how carefully tuned hyperparameters can significantly improve the predictive accuracy and robustness of CNN models in this application domain. A dataset of 192 labeled images was analyzed in both original and preprocessed form.

Across 500 optimization trials with 5-fold cross-validation, the best model trained on preprocessed data achieved a coefficient of determination (R^2) of 0.750, whereas the best model using the original input data reached only $R^2 = 0.217$. These results highlight the importance of both data preprocessing and structured hyperparameter search, particularly when working with small and heterogeneous datasets. The findings further underscore the potential of AI-driven approaches to enhance exposure assessment in occupational health.

Introduction

As in other domains, the world of work is being impacted by rapidly evolving technological advances. Artificial intelligence (AI) is one of the rapidly evolving technologies driving major change across various domAIns. One such area that could benefit from AI is hazardous substance identification. The assessment of dermal exposure to hazardous substances is a subset of this hazard identification. Chemical hazards in the workplace still pose a significant risk from an occupational safety perspective [1] regardless of numerous measures already taken by experts [2]. A major challenge that severely limits the development and validation of models is the limited availability of robust data. This is compounded by increased complexity due to different exposure situations and varying data collection methods.

Studies conducted to date on dermal exposure to chemicals have focused on comparing measurement methods using conventional sampling techniques [3]. Three distinct sampling principles are distinguished when determining dermal exposure: sampling methods, removal methods, and in situ methods [4]. However, these have the disadvantage of being time-consuming and requiring chemical analysis.

An image-based in situ assessment method developed by Franken et al. used fluorescence images taken before and after various activities involving chemicals [5]. Next, these images were evaluated by analyzing the brightness of individual pixels [5]. Building on this approach, AI could serve as a promising enhancement to existing image-based evaluation methods for occupational safety. By processing pixel intensity data, AI algorithms could potentially determine the extent of chemical exposure with greater efficiency and precision.

To ensure that such AI models can be reliably applied, careful training and optimization are required. In particular, hyperparameter tuning is critical for improving model robustness and generalizability, especially when datasets are small or heterogeneous, as is often the case in occupational field studies. This paper focuses on the application of convolutional neural networks (CNNs) for image-based regression of dermal exposure and investigates how systematic hyperparameter optimization can enhance model performance. By examining both original and preprocessed data, the study Aims to identify optimal training conditions and model configurations that can support future occupational health assessments through AI.

Methods

Dataset and Image Preprocessing

For this study, we used a subset of fluorescence images from a research project by the Federal Institute for Occupational Safety and Health, which Aimed to optimize and standardize methods for dermal exposure assessment using in situ imaging under UV light [4]. Our dataset includes tasks such as rolling and dipping with liquid substances, where image-based evaluation methods based on pixel-wise brightness value correlations demonstrated limitations in accurately interpreting dermal exposure [5]. From the original 320 images, we selected a curated subset of 192 images based on consistent exposure distribution and their relevance to the modeled scenarios. Each image was linked to a chemically determined exposure value.

Image preprocessing was performed in MATLAB using a morphological pipeline to segment test subjects from the background and extract fluorescence patterns. Steps included thresholding, morphological opening and closing, grayscale conversion, and application of binary masks to isolate relevant regions. This resulted in a preprocessed dataset emphasizing chemically exposed areas for improving model focus during training.

Model Architecture and Computational Environment

A convolutional neural network (CNN) was implemented in Python using Keras, featuring three convolutional layers with ReLU activation and max pooling, followed by global average pooling and a dense ReLU layer. A linear output neuron performed regression. The model used the Adam optimizer and mean squared error cost function. The chosen measure for model quality evaluation is the coefficient of determination (R^2) [6].

Hyperparameter optimization and model training were conducted on a high-performance compute cluster with two NVIDIA RTX 6000 Ada GPUs (48 GB VRAM each), 1 TB RAM, and a 128-core CPU. This setup enabled efficient execution of numerous trials without memory limitations and significantly reduced training and evaluation times compared to previous implementations. All experiments ran on Ubuntu 24.04.2 LTS.

Hyperparameter Optimization

To systematically improve the model architecture and training configuration, a hyperparameter optimization was performed using the Optuna framework integrated in Keras, a widely used open-source library for automated hyperparameter search based on Bayesian optimization and pruning techniques [7]. The goal of the optimization was to identify optimal combinations of architectural and training parameters for convolutional neural networks (CNNs) applied to the task of image-based dermal exposure assessment. In total, 500 optimization trials were carried out, each involving a complete training cycle over 200 epochs.

In each trial, a model was build using hyperparameters sampled from predefined search spaces. These included the number of filters in the first convolutional layer, ranging from 1 to 20, in the second layer from 1 to 128, and in the third layer from 1 to 100. The dense layer size was varied between 1 and 500 units The learning rate for the Adam optimizer was sampled as a continuous float on a logarithmic scale between 1e-7 and 1e-2, allowing exploration across a wide range of step sizes for gradient descent. The batch size, representing the number of images processed together during training, was treated as an integer parameter ranging from 1 to 50.

Each sampled configuration was evaluated using fivefold cross-validation on the training and validation subset, which comprised 70 % of the dataset. For every fold, the model was newly initialized and trained. The remaining 30 % of the dataset was held out as a constant test set and was used for the evaluation.

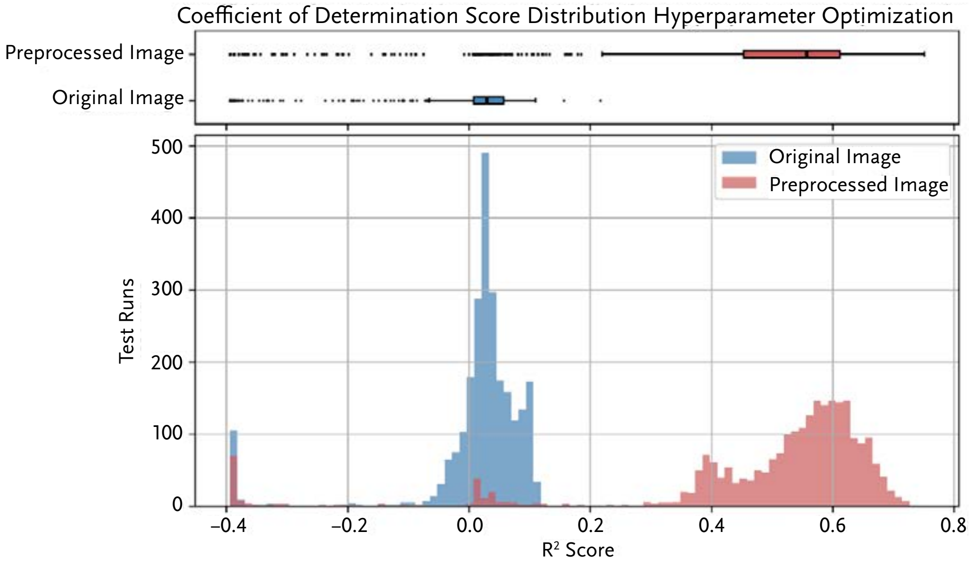

Fig.1 Distribution of the coefficient of determination (R^2 score) for 2500 test runs with different hyper-parameters for each original (red) and preprocessed (blue) image datasets.

Results

The histogram in Figure 1 illustrates the distribution of R^2 scores across all 500 × 5 Optuna trials, reflecting the performance range resulting from the hyperparameter optimization. The spread extends from negative values up to substantially high positive scores, highlighting the strong influence of hyperparameter configurations. Negative R^2 values indicate that some model configurations performed worse than a simple mean prediction, underscoring the importance of targeted optimization. The optimization process successfully identified high-performing model configurations, with the best model achieving an R^2 score of 0.750 when trained on preprocessed images.

This demonstrates that the search space and objective function used in the optimization were effective in guiding the training process toward performant solutions. Models trained on the original, unprocessed images achieved substantially lower results. The highest R^2 score in this group reached only 0.217, and the average performance remained close to 0.012, with many configurations resulting in negative R^2 values. This indicates that even with extensive tuning, the original input data lacked the informative structure needed for robust generalization. Preprocessing clearly played a critical role in enabling the model to extract meaningful features. The broader distribution of R^2 values in the preprocessed group further suggests that the optimized configurations were more sensitive to architectural and training hyperparameters, allowing for more effective exploration of the search space. In contrast, the narrow dispersion of low R^2 scores observed in the original group suggests that the potential for improvement is inherently limited, irrespective of the hyperparameter configuration.

Discussion and Future Research

This study demonstrates how systematic hyperparameter optimization enhanced the performance of CNNs for predicting dermal exposure from fluorescence-based images. Results also indicate that preprocessing plays a crucial role in enhancing model performance.

While this study demonstrates the benefits of systematic hyperparameter optimization for CNN-based exposure modeling, several factors may have limited the overall performance. The relatively small dataset and the fixed training configuration likely restricted generalizability despite extensive tuning. Increasing the number of cross-validation folds or trials, as well as refining the training dynamics, could further enhance model robustness. Additionally, the impact of image preprocessing on model outcomes suggests that data quality and representation remain critical elements alongside architectural optimization [8].

Despite the improvements achieved, the overall performance still leaves room for enhancement. Future studies could explore ensemble approaches, pretrained networks of classification (transfer learning), or hybrid models that integrate domain-specific knowledge. Moreover, expanding the dataset to include more subjects, exposure conditions, or imaging modalities would strengthen the robustness and applicability of the models in real-world occupational health scenarios.

In summary, this study provides initial evidence that automated hyperparameter optimization, when pAIred with meaningful preprocessing strategies, can lead to substantial performance gains in CNN-based dermal exposure modeling. However, further work is needed to address dataset size limitations, refine the optimization framework, and explore more diverse modeling approaches to unlock the full potential of image-based exposure assessment.

References

[1] Rim, K.-T. & Lim, C.-H. (2015). Chronic inhalation/carcinogenicity studies and their applications to prevent the occupational cancers: A review of recent reports. Toxicology and Environmental Health Sciences, 7(3), 171–183.

[2] Levy, L. (2004). Chemical hazards in the workplace: An overview. Occupational Medicine, 54(1), 67–68.

[3] Gorman Ng, M., Milon, A., Vernez, D., & Lavoué, J. (2015). A web-based tool to aid the identification of chemicals potentially posing a health risk through percutaneous exposure. The Annals of Occupational Hygiene, 60(3), 276–289..

[4] Kasiotis, K. M., Spaan, S., TsaKIraKIs, A. N., Franken, R., Chartzala, I., Anastasiadou, P., Machera, K., Rother, D., Roitzsch, M., Poppek, U., Lucadei, G., Baumgärtel, A., Schlüter, U., & Gerritsen-Ebben, R. M. (2019). Comparison of measurement methods for dermal exposure to hazardous chemicals at the workplace: The SysDEA project. Annals of Work Exposures and Health, 64(1), 55–70.

[5] Franken, R., Turkenburg, J., Kasiotis, K. M., Shandilya, N., Baan, J., TsaKIraKIs, A. N., Chartzala, I., Anastasiadou, P., Machera, K., Rother, D., Roitzsch, M., Poppek, U., Meyer, J., Schlüter, U., Gerritsen-Ebben, R. M., & Spaan, S. (2021). Prediction of dermal exposure to chemical substances using a fluorescence method within the SysDEA project. Annals of Work Exposures and Health, 65(5), 668–681.

[6] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623.

[7] Shekhar, S., Bansode, A., & Salim, A. (2021). A comparative study of hyper-parameter optimization tools. In 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE).

[8] Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2025). The effects of data quality on machine learning performance on tabular data. Information Systems, 132, 102549.

Suggested citation
André, Stefanie; Blohm, Frederik; Brenzke, Martin; Dietz, Arn; Geravand, Milad; Gruner, Marthe et al., Klostermann, Marina; Radüntz, Thea; Voelter-Mahlknecht, Susanne et al., Weißbrodt, Deniz et al.: 2025. Smarter Workplaces: The Role of AI in Promoting Occupational Safety and Health. Federal Institute for Occupational Safety and Health (BAuA). (baua: Focus). DOI 10.21934/baua:focus20250729 | July 2025.

Gender-neutral language is used in this publication. Where this is not possible or would detract from the readability of the text, terms used to refer to persons include all genders.

Imprint | Publisher: Federal Institute for Occupational Safety and Health (BAuA),
Friedrich-Henkel-Weg 1-25, 44149 Dortmund | Telephone: 0231 9071-2071, mail to: *info-zentrum@baua.bund.de*,
Internet: *www.baua.de*, | Design: R. Grahl, E. N. Simon | DOI: 10.21934/baua:focus20250729 | July 2025

baua:
Federal Institute for Occupational Safety and Health