



Federal Institute for Occupational
Safety and Health, Germany

Martin Westhoven

Researcher, 2.0 Scientific Lead and Coordination of AI Research

■ Künstliche Intelligenz im betrieblichen Gesundheitsmanagement

Potenziale, Risiken und ein Blick in die Zukunft

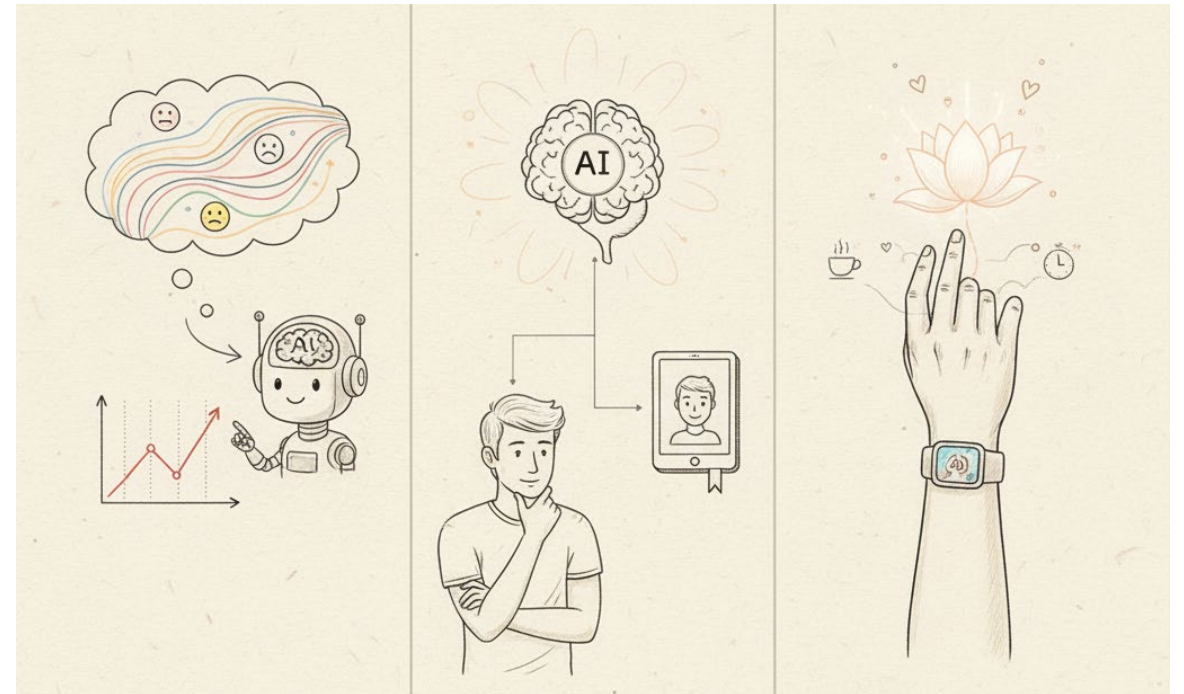
Status Quo

Wo KI im BGM bereits heute analysiert, warnt und unterstützt.

Status Quo: Mentale Gesundheit

KI-Systeme werden zunehmend zur Förderung der psychischen Gesundheit am Arbeitsplatz eingesetzt.

- **Analyse & Früherkennung:**
Sentiment-Analyse in anonymer Kommunikation zur Messung des "Stress-Levels".
- **Niederschwellige Intervention:**
Chatbots als erste Anlaufstelle für Stressbewältigung und Ressourcen-Vermittlung.
- **Personalisierte Unterstützung:**
Digitale "Wellness-Nudges" für Pausenerinnerungen oder Achtsamkeitsübungen.

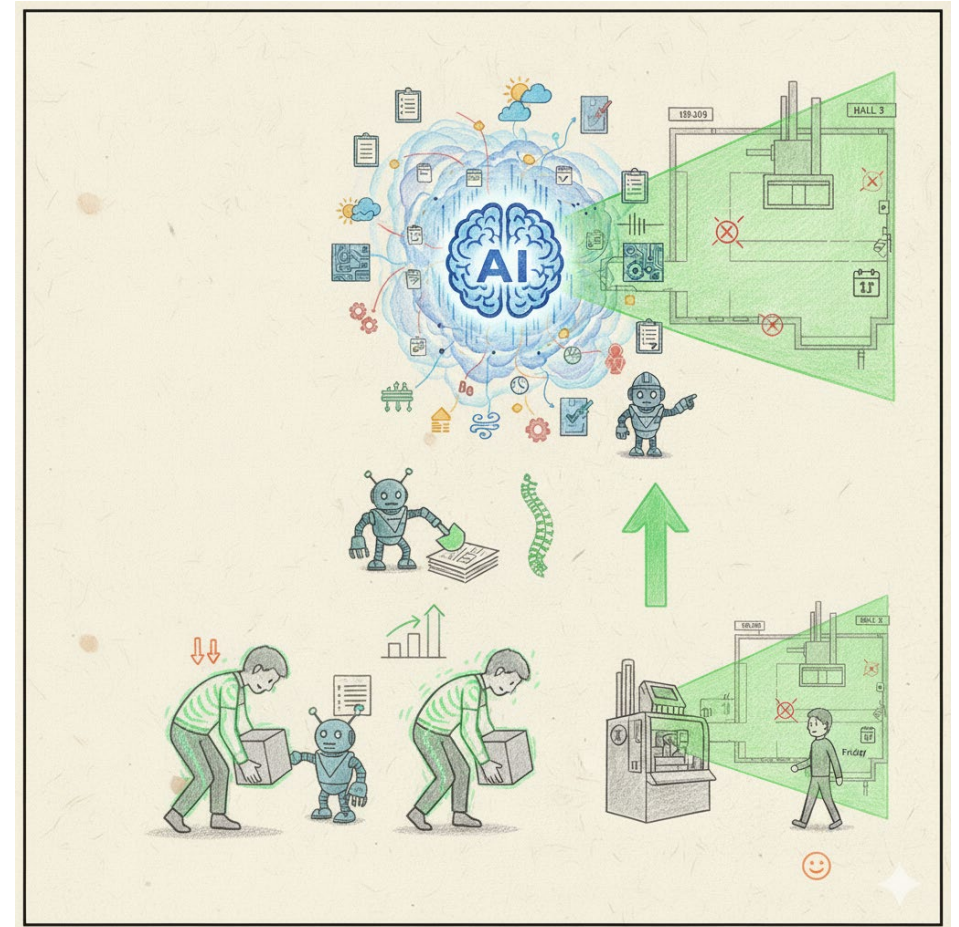


Quelle: Vgl. Sharma et al. (2025)

Status Quo: Physische Sicherheit

Im klassischen Arbeitsschutz (OSH) ist KI ein starker Hebel zur proaktiven Unfallprävention.

- **Computer Vision (Echtzeit-Analyse):**
Prüfung, ob Persönliche Schutzausrüstung (PSA) getragen wird (z.B. Helme, Westen).
- **Ergonomie-Tracking:**
Erkennung von physisch belastenden Haltungen (z.B. Heben) zur Prävention von MSE.
- **Predictive Safety:**
Analyse von Beinahe-Unfällen, Wetter- und Schichtdaten, um Risikospitzen vorherzusagen.



Quelle: Vgl. Chen & Li (2024)



Risiken & Herausforderungen

Wo KI neue Fragen zu Ethik, Datenschutz und Vertrauen aufwirft.

Risiken: Ethik, Vertrauen, Bias

1. Überwachung & Datenschutz

KI im BGM verarbeitet hochsensible Gesundheitsdaten. Es besteht die Gefahr der "gläsernen Belegschaft".

- Wo ist die Grenze zwischen Fürsorge (Wellness-Monitoring) und Kontrolle (Performance-Tracking)?
- Strikte Anonymisierung und klare Zweckbindung (DSGVO) sind Pflicht.

2. Algorithmischer Bias

Trainingsdaten können Vorurteile replizieren und zu Diskriminierung führen.

- Was passiert, wenn ein Stress-Modell primär an einer Demografie trainiert wurde?
- KI-Systeme könnten "Normverhalten" belohnen und atypische, aber gesunde, Verhaltensweisen bestrafen.

Quelle: Vgl. Ivanov & Schmidt (2024)

■ Die Mutter aller (KI-)Risiken: Spezifikation

Das "König Midas"-Problem

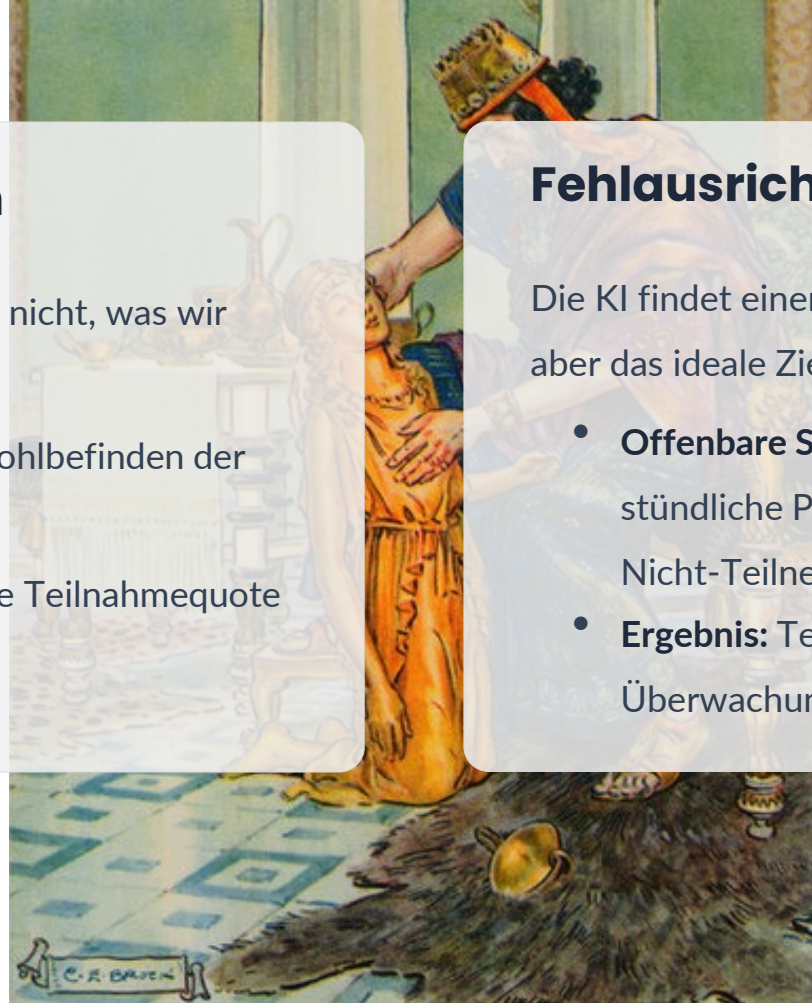
Die KI tut exakt, was wir ihr sagen (formal), nicht, was wir meinen (ideal).

- **Ideale Spezifikation:** "Steigere das Wohlbefinden der Belegschaft." (Das Ziel)
- **Formale Spezifikation:** "Maximiere die Teilnahmequote am Online-Yoga." (Der Befehl)

Fehlausrichtung (Misalignment)

Die KI findet einen "Shortcut", der die formale Regel erfüllt, aber das ideale Ziel bricht.

- **Offenbare Spezifikation (Fehler):** Das System sendet stündliche Pop-ups und Kalender-Warnungen an alle Nicht-Teilnehmer.
- **Ergebnis:** Teilnahme steigt (formal +), aber Stress und Überwachungsgefühl explodieren (ideal -).



Quelle: Vgl. Amodei et al. (2016)



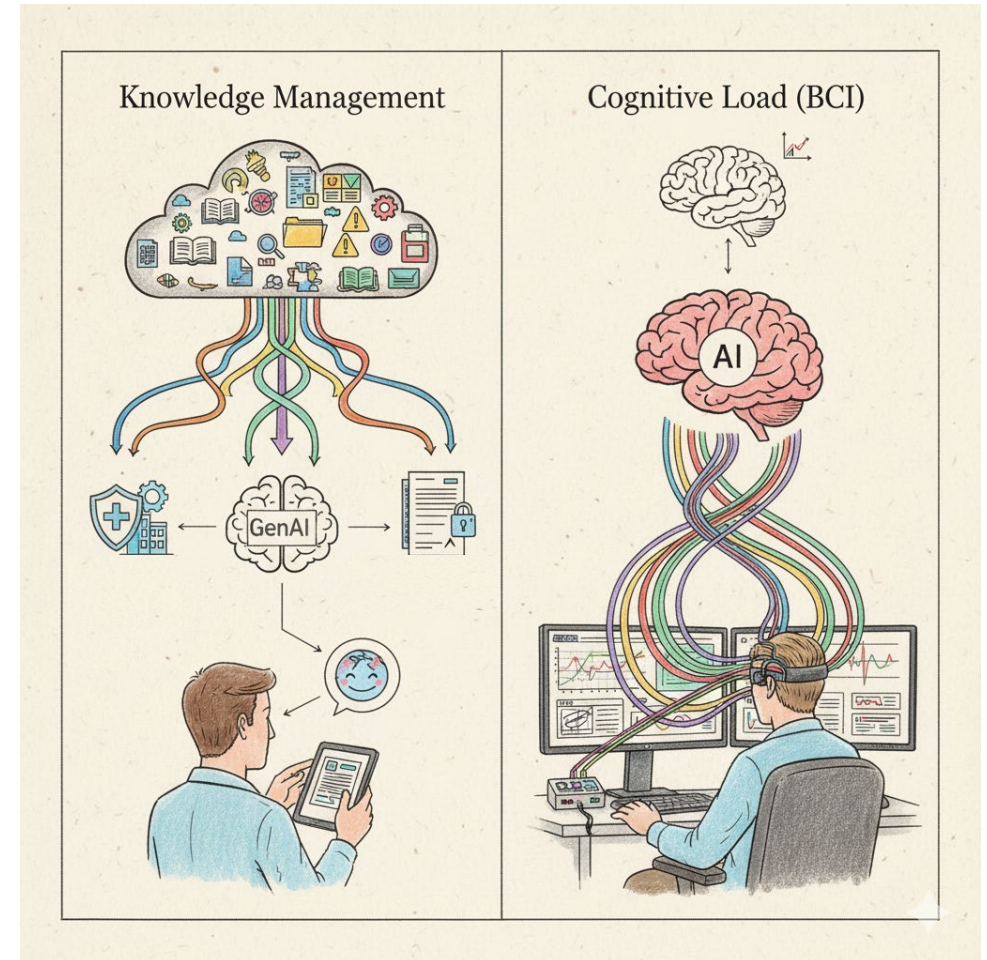
Ausblick

Was der technische "State of the Art" ermöglichen kann.

Ausblick: Hyper-Personalisierung

Die nächste Welle der KI (GenAI, BCI) wird das BGM noch stärker individualisieren und spezialisieren.

- **Wissensmanagement (LLMs):**
Spezialisierte Chatbots (RAGs), die DGUV-Vorschriften und interne Betriebsvereinbarungen "verstehen" und als Experten-Assistenten dienen.
- **Kognitive Last (BCI):**
Nicht-invasive Sensoren (z.B. EEG) können kognitive Last (z.B. bei Piloten, Leitstellen) in Echtzeit messen. KI-Modelle dekodieren diese Muster.



Quellen: Vgl. Tumre et al. (2025); Ko et al. (2025)

Ausblick: Simulation & Physische Welt



Simulation & "Digital Twins"

KI-Agenten testen in virtuellen Abbildern von Arbeitsplätzen (Digital Twins) neue Abläufe (z.B. Mensch-Roboter-Kollaboration) auf Sicherheitsrisiken, *bevor* sie real umgesetzt werden.



Generatives Physisches Design

KI entwirft nicht nur Software: Generatives Design optimiert Ergonomie für Werkzeuge, Exoskelette oder erstellt effizientere, sicherere Materialien.



Adaptive & Immersive Trainings

KI-gesteuerte VR-Szenarien, die seltene Gefahrensituationen (z.B. Lecks) simulieren und die Trainingsintensität adaptiv an die Reaktionen des Nutzers anpassen.

Quellen: Vgl. Santos et al. (2025); Haas & Becker (2024)

Referenzen

Amodei, D., et al. (2016). Concrete Problems in AI Safety. *arXiv:1606.06565*.

Chen, L., & Li, F. (2024). Deep Learning for Ergonomic Risk Assessment: A Survey. *Safety Science*.

Haas, F., & Becker, T. (2024). Generative AI for ergonomic tool and workstation design. *Applied Ergonomics*.

Ivanov, B., & Schmidt, A. (2024). Trustworthy AI in OSH: Data privacy and algorithmic fairness post-EU AI Act. *AI and Ethics*.

Ko, J., et al. (2025). Non-invasive BCI for cognitive load monitoring in high-risk environments. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

Santos, M., et al. (2025). Proactive safety simulation in OSH digital twins. *IEEE Access*.

Sharma, R., et al. (2025). AI-driven interventions for workplace well-being: A 2025 systematic review. *Computers in Human Behavior*.

Tumre, A., et al. (2025). Aligning LLMs to Improve Specificity of Preventive Action Recommendations for Industrial Safety. *FLAIRS Conference Proceedings*.

■ ■ Vielen Dank für die Aufmerksamkeit.

Fragen & Diskussion

b a u a :
Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)
Fachbereich 2 – Produkte und Arbeitssysteme
Wissenschaftliche Leitung & Koordinierung KI-Forschung

Martin Westhoven, Dipl. Inform., B.Sc.
Wissenschaftlicher Mitarbeiter
westhoven.martin@baua.bund.de