

## Potenziale Künstlicher Intelligenz zur Risikoanalyse im betrieblichen Arbeitsschutz

Martin WESTHOVEN

*Bundesanstalt für Arbeitsschutz und Arbeitsmedizin,  
Friedrich-Henkel-Weg 1-25, D-44149 Dortmund*

**Abstract:** Die Dissertation beleuchtet Potenziale von KI-Methoden zur Unterstützung von Risikoanalysen im betrieblichen Arbeitsschutz. Dies bringt auf Datenebene verschiedene Probleme mit sich, die mit passenden Methoden und Algorithmen angegangen werden müssen. Zu nennen sind hierbei heterogene und unvollständige Daten sowie ein Mangel an gelabelten Daten für überwachtes Lernen. Darüber hinaus ist insbesondere auch die Interaktionsgestaltung zwischen Mensch und KI von hoher Bedeutung, da das Ergebnis einer Risikoanalyse im betrieblichen Arbeitsschutz unmittelbar sicherheitsrelevant ist und Fehler so weit wie möglich ausgeschlossen werden müssen.

**Keywords:** Risikoanalyse, Gefährdungsbeurteilung, Künstliche Intelligenz, Interaktionsdesign

### 1. Problem- und Zielstellung

In diesem Vorhaben sollen die Möglichkeiten der Anwendung Künstlicher Intelligenz als Werkzeug des betrieblichen Arbeitsschutzes untersucht werden. Als Anwendungsfeld soll das Risikomanagement im betrieblichen Arbeitsschutz in den Fokus genommen werden. Hier entstehen heterogene und komplexe Daten, die sich nicht einfach hinsichtlich ihrer Zusammenhänge analysieren lassen. Künstliche Intelligenz liefert Ansätze, um diese Datenproblematik anzugehen und Zusammenhänge zu analysieren, zu modellieren und Risiken zu identifizieren.

Aufgrund der Vielfalt unterschiedlicher Risiken im Betrieb wird auf ein prototypisches System fokussiert, welches mit im Betrieb vorliegenden oder einfach zu erhebenden Daten abbildbar ist. Im Arbeitsschutz ergeben sich solche Daten z. B. aus den typischen Faktoren von Gefährdungsbeurteilungen, können aber auch abhängig vom Kontext Branchen- bzw. Betriebseigenschaften sowie die immer häufiger vernetzt vorliegenden Maschinendaten umfassen. Hinsichtlich betrachteter Gefährdungsfaktoren werden physikalische Gefährdungen als Hauptaugenmerk bei der Vermeidung schwerer Unfälle in den Fokus genommen.

Die sehr erfolgreiche KI-Methode des Deep Learning kann das Vorhandensein einer Fülle von gelabelten Daten benötigen (siehe z. B. Shams (2014)), während im Risikomanagement aufgrund der Komplexität der Problemstellung nur spärliche objektive Daten sowie entsprechende Label zu erwarten sind (Clemen & Winkler 1999). Durch Einbindung von a priori-Wissen können jedoch sogenannte Few-Shot Learning-Ansätze zum Einsatz gebracht werden (siehe z. B. Wang et. al (2020)), welche darauf zielen, mit nur wenigen oder sogar nur einem Trainingsbeispiel zu arbeiten. Um das Problem

fehlender Daten anzugehen, sei weiter das sogenannte Deep Latent Variable Modeling (Kingma & Welling 2013) erwähnt, welches mittels eines generativen Ansatzes fehlende Daten imputiert und zuletzt in vielen Bereichen erfolgreich eingesetzt wurde.

Aufgrund der Komplexität der Aufgabe und der menschlichen Entscheidungshoheit soll der Ansatz hybrider Intelligenz (Dellermann et. al 2019) Beachtung finden. Eine Voraussetzung dafür ist Erklärbarkeit (Explainable AI, siehe Doran et. al 2017)). Hybride Intelligenz erfordert weiter auch ein sorgfältig gestaltetes User Interface, ganz besonders, wenn der Systemoutput sicherheitskritisch ist. Dazu stellen bspw. Schmidt und Herrmann (2017) eine Anpassung von Shneiderman's Golden Rules (Shneiderman et al. 2016) zur Diskussion.

Um Anforderungen zu ermitteln, soll für Experteninterviews mit einem größeren Unternehmen kooperiert werden. Der Gesprächsleitfaden greift auf das Konzept der Grounded Theory (Strauss & Corbin 1998) zurück und beinhaltet Frageblöcke zur Erfahrung, zum Ist-Zustand, zu einem möglichen Wunschzustand des Prozesses sowie zu ergänzenden Anmerkungen. Neben Fragen zum Prozess selbst wurde der Leitfaden um Aspekte der soziotechnischen Systemgestaltung nach Herrmann et al. (2021) ergänzt.

Zusammenfassend bleibt festzuhalten, dass für einen Einsatz in einem realitätsnahen bis realistischen Umfeld folgende Herausforderungen zu bewältigen sind:

- Heterogene und hochdimensionale Eingabedaten
- Wenige bewertete / gelabelte Eingabedaten
- Fehlende Daten
- Adäquate UI-Gestaltung inklusive Explainable AI

Für sich betrachtet existieren zu jeder der Herausforderungen Lösungsansätze. Der Kern der Dissertation ist daher, die unterschiedlichen Ansätze an die Anwendungsdomäne anzupassen, zu integrieren, sie mit einer adäquaten Benutzungsschnittstelle auszustatten, und im Anwendungsumfeld hinsichtlich Systemleistung und Usability zu erproben.

## **2. Bisherige Ergebnisse**

Auch wenn die Arbeit sich am ehesten zu Linie 1 zuordnen ließ, sind im Laufe der Arbeit schon einige Ergebnisse zusammengekommen. Diese sollen hier kurz dargestellt werden.

Zunächst wurde in Westhoven und Herrmann (In Press) das grundlegende Konzept der Arbeit ausgearbeitet und die Anknüpfungspunkte in den verschiedenen Problem-bereichen Daten, KI-Ansätze und Interaktionsdesign beleuchtet. Als Fundament für den Interaktionsdesignaspekt wurden in Westhoven (2022) die Ergebnisse einer ersten Interviewreihe (halb-strukturiert) mit Arbeitsschutzpraktikern der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin berichtet und bezüglich der Anforderungen an ein mögliches unterstützendes KI-System interpretiert. In Westhoven und Jadid (In Press) wird schließlich ein erstes implementiertes KI-System zur Unterstützung des betrieblichen Arbeitsschutzes beschrieben. Dieses System wird mittels Transferlernen von wissenschaftlichen Abstracts und Nachrichten-Artikeln auf Unfallbeschreibungen tödlicher Arbeitsunfälle neu trainiert und soll damit dann auch ermöglichen, aus ähnlichen Texten, insbesondere aus Checklisten für Gefährdungsbeurteilungen, Schlüsselwörter zu extrahieren. Denn anhand von Schlüsselwörtern fällt anschließend ein

Matching zwischen beispielsweise unstrukturiertem Text und dazu passenden Checklistenitems relativ einfach.

Weiterhin wird in Westhoven und Herrmann (In Press) das Interaktionsdesign hinsichtlich spezieller Anforderungen resultierend aus dem KI-Einsatz beleuchtet.

### 3. Methodische Überlegungen zum weiteren Verlauf

Gegenwärtig laufen mehrere Stränge parallel. Prioritär zu behandeln ist die Kooperation mit einem Kooperationsunternehmen in dem Arbeitsschutzdaten gesammelt werden können, um deren Handhabung in KI-Systemen zur Unterstützung der Risikoanalyse zu untersuchen. Erste Gespräche haben hier stattgefunden und es läuft derzeit eine Prüfung, welche Bereiche des Unternehmens (Energieversorger) infrage kommen und welche Daten überhaupt zur Verfügung gestellt werden können. Hier wurde bereits begonnen weitere Experteninterviews zu führen, um die Perspektive von Arbeitsschutzpraktikern eines Unternehmens ergänzen zu können.

Ebenfalls hoch priorisiert ist die Entwicklung sinnvoller und lauffähiger KI-Systeme zur Unterstützung von Risikoanalyseprozessen im betrieblichen Arbeitsschutz. Gegenwärtig konzentriert sich dies auf den oben beschriebenen Ansatz, aus textuellen Beschreibungen passende Checklistenitems abzuleiten und damit eine schnelle, dynamische Erstellung von angepassten Checklisten zu ermöglichen.

Diese implementierten Systeme, die natürlich auch wenigstens zum Teil auf den zur Verfügung stehenden Daten des Kooperationsunternehmens funktionieren müssen, sollen im Verlauf des Projektes auch mit dem Unternehmen vor Ort evaluiert werden, um weitere Einblicke in Potenziale, aber auch Fallstricke von KI in diesem Bereich zu gewinnen.

Niedriger priorisiert, aber in jedem Fall auch noch anzugehen, ist der Kontakt zu Arbeitsschutzexperten auf KMU-Ebene für Experteninterviews. Ziel ist es hier einen Abgleich machen zu können, ob Methoden auch auf in KMUs realistischerweise vorzufindenden/zu erhebenden Daten und auf der vorhandenen Infrastruktur lauffähig sind und falls nein, ob man sich algorithmisch oder auch betrieblich entsprechend anpassen kann.

### 4. Literatur

- Clemen RT, Winkler RL (1999). Combining probability distributions from experts in risk analysis. *Risk analysis*, 19 (2), 187-203.
- Dellermann D, Ebel P, Söllner M, Leimeister JM (2019). Hybrid intelligence. *Business & Information Systems Engineering*, 61 (5), 637–643.
- Doran D, Schulz S, Besold TR (2017). What does explainable AI really mean? A new conceptualization of perspectives. *arXiv preprint arXiv: 1710.00794*.
- Herrmann T, Jahnke I, Nolte A (2021). A problem-based approach to the advancement of heuristics for socio-technical evaluation. *Behaviour & Information Technology*, 1–23.
- Kingma DP, Welling M (2013). Auto-encoding variational bayes. *arXiv preprint arXiv: 1312.6114*.
- Schmidt A, Herrmann T (2017). Intervention user interfaces: a new interaction paradigm for automated systems. *Interactions*, 24 (5), 40–45.
- Shams R (2014). Semi-supervised classification for natural language processing. *arXiv preprint arXiv: 1409.7612*.
- Shneiderman B, Plaisant C, Cohen MS, Jacobs S, Elmqvist N, Diakopoulos N (2016). *Designing the user interface: strategies for effective human-computer interaction*: Pearson.

Strauss A, Corbin J (1998). Basics of qualitative research techniques.

Wang Y, Yao Q, Kwok JT, Ni LM (2020). Generalizing from a few examples: A survey on few-shot learning. ACM computing surveys (csur), 53 (3), 1-34.

Westhoven M (2022). Requirements for AI Support in Occupational Safety Risk Analysis. In Proceedings of Mensch und Computer 2022 (pp. 561–565).

Westhoven M, Herrmann T (In Press). Interaction Design for Hybrid Intelligence: The Case of Work Place Risk Assessment.

Westhoven M, Jadid A (In Press). Supporting Work Place Risk Assessments by Means of Natural Language Processing. Paper presented at the 69th GfA Frühjahrskongress, Hannover, Germany.



Gesellschaft für Arbeitswissenschaft e.V.

## Nachhaltig Arbeiten und Lernen

**Analyse und Gestaltung lernförderlicher  
und nachhaltiger Arbeitssysteme  
und Arbeits- und Lernprozesse**

69. Kongress der  
Gesellschaft für Arbeitswissenschaft e.V.

Gottfried Wilhelm Leibniz Universität Hannover

01. – 03. März 2023

---

## GfA-Press

---

**Bericht zum 69. Arbeitswissenschaftlichen Kongress vom 01. – 03. März 2023**

**Fakultät Maschinenbau, Institut für Berufswissenschaften der Metalltechnik (IBM) und  
Institut für Fabrikanlagen und Logistik (IFA), Leibniz Universität Hannover**

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.  
Sankt Augustin: GfA-Press, 2023  
ISBN 978-3-936804-32-4

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle (s. u.) erhältlich.

Alle Rechte vorbehalten.

© GfA-Press, Sankt Augustin

**Schriftleitung: Prof. Dr. Rolf Ellegast**

im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Ohne ausdrückliche Genehmigung der Gesellschaft für Arbeitswissenschaft e.V. ist es nicht gestattet:

- den Kongressband oder Teile daraus in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) zu vervielfältigen,
- den Kongressband oder Teile daraus in Print- und/oder Nonprint-Medien (Webseiten, Blog, Social Media) zu verbreiten.

Die Verantwortung für die Inhalte der Beiträge tragen alleine die jeweiligen Verfasser; die GfA haftet nicht für die weitere Verwendung der darin enthaltenen Angaben.

### **Geschäftsstelle der GfA**

Simone John, Tel.: +49 (0)30 1300-13003

Alte Heerstraße 111, D-53757 Sankt Augustin

[info@gesellschaft-fuer-arbeitswissenschaft.de](mailto:info@gesellschaft-fuer-arbeitswissenschaft.de) · [www.gesellschaft-fuer-arbeitswissenschaft.de](http://www.gesellschaft-fuer-arbeitswissenschaft.de)

### **Screen design und Umsetzung**

© 2023 fröse multimedia, Frank Fröse

[office@internetkundenservice.de](mailto:office@internetkundenservice.de) · [www.internetkundenservice.de](http://www.internetkundenservice.de)