

Robelski, Swantje; Dietz, Arn; Jelenko, Marie; Fischer-Schwarz, Robert

Article — Published Version

Chatbots für den Arbeitsschutz: Nutzungsszenarien, Risiken und Empfehlungen

Zeitschrift für Arbeitswissenschaft

Provided in Cooperation with:

Springer Nature

Suggested Citation: Robelski, Swantje; Dietz, Arn; Jelenko, Marie; Fischer-Schwarz, Robert (2025) : Chatbots für den Arbeitsschutz: Nutzungsszenarien, Risiken und Empfehlungen, Zeitschrift für Arbeitswissenschaft, ISSN 2366-4681, Springer, Berlin, Heidelberg, Vol. 79, Iss. 3, pp. 460-471, <https://doi.org/10.1007/s41449-025-00483-4>

This Version is available at:

<https://hdl.handle.net/10419/330878>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/deed.de>



Chatbots für den Arbeitsschutz: Nutzungsszenarien, Risiken und Empfehlungen

Swantje Robelski¹ · Arn Dietz¹ · Marie Jelenko² · Robert Fischer-Schwarz²

Angenommen: 4. September 2025 / Online publiziert: 30. September 2025
© The Author(s) 2025

Zusammenfassung

Chatbots, die auf großen Sprachmodellen (*large language model*, LLM) basieren, wie z. B. ChatGPT, stellen eine besondere Form der Künstlichen Intelligenz (KI) dar. Diese Sprachmodelle wurden mit Hilfe großer Mengen an Textdaten darauf trainiert, natürliche Sprache zu generieren. Auch im Arbeitsschutz können solche Chatbots Anwendung finden, beispielsweise durch niedrigschwellige Informationsbereitstellung. Sie können dabei ein Werkzeug sein, das gleichermaßen von Beschäftigten und Verantwortlichen im Arbeitsschutz eingesetzt wird. In einem Kooperationsprojekt zwischen der österreichischen Allgemeinen Versicherungsanstalt (AUVA) und der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) in Deutschland werden daher Nutzungsszenarien, Risiken und Empfehlungen beim Einsatz frei zugänglicher Chatbots im Kontext von Sicherheit und Gesundheit bei der Arbeit beleuchtet.

Basierend auf der Literatur und einer nutzerzentrierten Analyse werden Anwendungsfälle im Bereich des Arbeits- und Gesundheitsschutzes entwickelt, in Fallbeispielen untersucht und von Expert:innen evaluiert. Darauf aufbauend werden Empfehlungen abgeleitet, die Nutzer:innen bei der Anwendung und Interpretation von Ergebnissen unterstützen.

Es zeigt sich, dass die Nutzung von Chatbots im Bereich Sicherheit und Gesundheit bei der Arbeit Unterstützungspotential und Zugang zu Informationen bietet. Die Prüfung der Angaben und/oder der Rückgriff auf Fachexpertise bleiben aufgrund des Risikos fehlender oder fehlerhafter Informationen jedoch unerlässlich.

Nutzen für die Praxis: Die Projektergebnisse bieten praxisnahe Empfehlungen für den Einsatz von generativen KI-Chatbots im Bereich der Arbeitssicherheit und des Gesundheitsschutzes. Die entwickelten Hinweise und Leitplanken können betriebliche Praktiker:innen in der Entscheidung unterstützen, ob und inwieweit Chatbots für sie geeignet sind und eine erste Annäherung an das Thema erleichtern.

Die praxisorientierten Fallbeispiele verdeutlichen das breite Anwendungsspektrum von Chatbots, wie etwa bei Frage-Antwort-Formaten oder der Erstellung von Informationsmaterial. Zudem hilft das entwickelte Modell zur Klassifikation von Anwendungsfällen, die geeigneten Einsatzgebiete von KI-Chatbots zu identifizieren und eine zielgerichtete Nutzung zu fördern. Die Empfehlungen zur sicheren Handhabung und Interpretation der Chatbot-Antworten tragen dazu bei, Missverständnisse und Fehlinterpretationen zu vermeiden.

Schlüsselwörter Künstliche Intelligenz · Große Sprachmodelle · Chatbots · Arbeitsschutz · Gesundheit

✉ Swantje Robelski
Robelski.Swantje@baua.bund.de

¹ Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA),
Berlin/Dortmund, Nölderstraße 40–42, 10317 Berlin,
Deutschland

² Allgemeine Unfallversicherungsanstalt (AUVA), Wien,
Österreich

Chatbots for Occupational Health and Safety: Use Cases, Risks, and Recommendations

Abstract

Chatbots based on large language models (LLMs), such as ChatGPT, represent a specific form of artificial intelligence (AI). These language models are trained on large volumes of textual data to generate natural language. In the field of occupational safety and health, such chatbots can also find application—for example, by providing low-threshold access to information. They can serve as tools used both by employees and by those responsible for occupational safety.

In a collaborative project between Austrian Workers' Compensation Board (AUVA) and the German Federal Institute for Occupational Safety and Health (BAuA), usage scenarios, risks, and recommendations related to the use of publicly available chatbots in the context of occupational safety and health are being examined. Based on the literature and a user-centered analysis, exemplary use cases in the field of occupational health and safety are developed and evaluated by experts. The results serve as a basis for deriving recommendations that help users apply and interpret chatbot-generated answers appropriately.

The findings suggest that chatbots have the potential to support access to information in the field of occupational safety and health. However, due to the risk of incomplete or inaccurate information, verification and/or consultation with subject matter experts remains essential.

Practical Relevance: The results of the project provide practice-oriented recommendations for the use of generative AI chatbots in occupational safety and health. The developed guidelines and orientation frameworks aim to support workplace practitioners in deciding whether and to what extent chatbots may be suitable for their needs, and to facilitate an initial engagement with the topic.

The user stories illustrate the wide range of possible applications for chatbots—for example, in question-and-answer formats or the creation of informational materials. In addition, the developed model for classifying use cases helps identify appropriate areas for deploying AI chatbots and promotes targeted use. The recommendations for safe handling and interpretation of chatbot responses contribute to preventing misunderstandings and misinterpretations.

Keywords Artificial intelligence · Large language models · Chatbots · Occupational safety and health

1 Einleitung

Im November 2022 bringt eine US-amerikanische Firma einen Chatbot auf den Markt, der basierend auf einem generativen Sprachmodell in der Lage ist, mit seinen Nutzer:innen in den Dialog zu treten. Rezepte anfragen, Ideen zur Freizeitgestaltung in einer unbekannten Stadt sammeln, ein kreativer Titel für einen Vortrag oder das Verfassen von Glückwünschen erscheinen im Austausch mit dem Chatbot leicht zugänglich und das vermeintliche Wissen, das in dem Modell steckt, ist groß. ChatGPT – so der Name des Chatbots – hat seitdem nicht nur eine enorme mediale Aufmerksamkeit erregt, sondern ist mit 800 Mio. Nutzer:innen¹ ein breit genutztes Werkzeug.

Seitdem haben zahlreiche weitere Firmen Chatbots, die auf generativen Sprachmodellen basieren, veröffentlicht. Die Anwendungen sind auf eine lockere Konversation ausgelegt, aber können auch auf spezifische Themen und Arten von Dialogen zugeschnitten werden. Die große Dynamik in diesem Bereich macht es nahezu unmöglich, einen Überblick über das Geschehen zu behalten. Obwohl sich Chatbots nach wie vor großer Beliebtheit erfreuen,

treten zunehmend Bedenken auf, die sich sowohl auf den enormen Energieverbrauch als auch auf gesellschaftlich-soziale Effekte, z. B. Diskriminierung oder Fehlinformation beziehen (Haque und Li 2024). Erst kürzlich war das Verhalten eines großen Chatbots in die Kritik geraten, das nach einem Update zu „unterwürfig“ war (OpenAI 2025). Ein mittelmäßiger Freizeittipp oder ein umständlich formulierter Glückwunsch sind zwar ärgerlich, doch was passiert, wenn Chatbots im Kontext der Arbeitstätigkeit eingesetzt werden? Aktuelle Daten einer großen Befragungsstudie lassen vermuten, dass Chatbots eine KI-Technologie darstellen, die mittlerweile von vielen Beschäftigten eingesetzt wird – häufig auch ohne betriebliche Absprache (Arntz et al. 2025). Überwiegen im Kontext von Sicherheit und Gesundheit bei der Arbeit die Chancen, wie niedrigschwelliger Zugang zu Informationen? Oder sind die Risiken wie Falschinformationen und fehlende Informationen zu groß? Denn die Einhaltung der Arbeitsschutzbestimmung wird insbesondere von kleinen und mittleren Unternehmen nach wie vor als herausfordernd erlebt (Salguero-Caparrós et al. 2020; EU-OSHA 2025). Die Informationsgewinnung und Reduktion von Komplexität mithilfe von Chatbots könnte Erleichterung bringen. Gleichzeitig stellen Falschinformationen oder fehlende Informationen bei Fragen des Arbeitsschutzes große Risiken dar.

¹ <https://www.forbes.com/sites/martineparis/2025/04/12/chatgpt-hits-1-billion-users-openai-ceo-says-doubled-in-weeks/>.

Aus diesem Grund möchte das Kooperationsprojekt der deutschen Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) und der österreichischen Allgemeinen Unfallversicherungsanstalt (AUVA) dazu beitragen, Unternehmen und im Arbeitsschutz tätige Personen bei der reflektierten Nutzung von Chatbots zu unterstützen. Besonders im Fokus stehen dabei auch Zielgruppen ohne tiefgehende IT- oder Chatbot-Expertise, denen ein praxisnaher und verständlicher Zugang ermöglicht werden soll. Entsprechend wurde unter dem Kurztitel „Chatbotguide“ ein Leitfaden entwickelt, der die positiven Möglichkeiten der Nutzung von sprachbasierten Chatbots betrachtet, ohne problematische Aspekte aus den Augen zu verlieren. Das Projekt stellt eine praxisorientierte Erprobung dar und verfolgt in diesem Sinne keinen wissenschaftlichen Anspruch. Gleichwohl erfolgte die Leitfadententwicklung unter Berücksichtigung von aktueller wissenschaftlicher Literatur zur Chatbot-Nutzung im Bereich von Arbeitssicherheit und Gesundheitsschutz. Nach explorativen Erhebungen bei Fachkräften der Prävention wurden typische Anwendungsfälle entwickelt und mit experimentellen Fallbeispielen veranschaulicht. Diese werden im Leitfaden kritisch reflektiert und mit praxisorientierten Empfehlungen ergänzt. Zu Projektende wurden die Beispiele und Empfehlungen im Rahmen eines Workshops von Fachleuten bewertet und der Leitfaden dementsprechend angepasst.²

1.1 Was sind Chatbots?

Ein Chatbot ist ein Computerprogramm, mit dem der/die Nutzer:in in Form eines Chatverlaufes wie mit einem menschlichen Gegenüber interagieren kann. Im Allgemeinen können Chatbots sowohl mit Methoden des maschinellen Lernens als auch als klassisches Computerprogramm gestaltet werden.

Die leistungsfähigsten Chatbots, um die es im Weiteren gehen soll, basieren auf einem großen Sprachmodell (eng. Large Language Model, LLM). Ein LLM ist ein Computerprogramm, welches einen Text als Eingabe erhält und dann ein Wort an den Text anfügen soll, das den Text sinnvoll weiterführt (Wolfram 2023)³.

Die Aufgabe, den Text weiterzuführen, ist komplex und erfordert nicht nur ein Verständnis für die Grammatik einer Sprache, sondern auch, dass der Inhalt des Vorhergegangenen berücksichtigt wird. Eine sinnvolle Weiterführung des Satzes „Das Huhn überquerte die ...“ ist das Wort „Straße“.

Die Worte „Käfigtür“, „Legebatterie“ oder „Kuh“ haben im weiteren Sinne alle etwas mit Hühnern auf einem Bauernhof zu tun und sind syntaktisch gesehen eine korrekte Weiterführung des Satzes. Semantisch ergeben sie allerdings entweder keinen Sinn oder fühlen sich für menschliche Leser:innen weniger „wahrscheinlich“ an als „Straße“. Um den Satz weiterführen zu können, muss in dem LLM also eine präzise semantische Verknüpfung oder Abgrenzung verschiedener Begriffe hinterlegt sein (Dong et al. 2022).

Technisch handelt es sich bei einem LLM (Naveed et al. 2023) um ein großes, künstliches neuronales Netz (KNN): einen Algorithmus also, der einem biologischen neuronalen Netz nachempfunden ist ((Kriesel 2006) bietet einen guten Überblick über KNN Grundlagen). Durch einen Trainingsprozess wird das LLM in die Lage versetzt, sinnvoll Texte zu vervollständigen. Dabei werden in einer ersten Trainingsphase Teile bekannter Texte als Eingabe verwendet. Wenn die Ausgabe nicht der korrekten, bekannten Weiterführung des Textes entspricht, werden die künstlichen Nervenbahnen des LLM (in diesem Kontext als Gewichte bezeichnet) verändert. In diesem Vorgang liegt ein großer Teil der Unsicherheit begründet, die im Umgang mit KI herrscht. Denn während die Funktionsweise der einzelnen Bestandteile von künstlichen neuronalen Netzen gut erforscht ist, ist das Zusammenspiel der einzelnen Bestandteile beim Lösen von Aufgaben so komplex, dass sich bisher im Detail nicht sagen lässt, wie LLMs ihren Output generieren. Ähnlich zum menschlichen Nervensystem, wo gut verstanden ist, wie einzelne Nervenzellen funktionieren, aber bisher unverstanden ist, wie z. B. das Bewusstsein zustande kommt. Derzeit finden sich verschiedene Bemühungen zu verstehen, was LLMs tatsächlich verstehen und lernen (Kanaka 2025; Gao et al. 2024).

Eine Grundregel der Mathematik, die auch bei künstlichen neuronalen Netzen Anwendung findet, ist, dass zur Bestimmung jedes Freiheitsgrades eines Systems eine Information gebraucht wird, die von den anderen Informationen unabhängig ist. Da das Netzwerk sehr groß ist, wird eine Vielzahl von Texten zum Training gebraucht. ChatGPT 3 z. B. hat mehrere Milliarden dieser Freiheitsgrade⁴ und es muss davon ausgegangen werden, dass der komplette, im Internet offen verfügbare Text zum Training verwendet wurde.

Beim Einsatz von LLMs als Grundlage für Chatbots erfolgt die Textgenerierung durch mehrfaches Anwenden des LLMs auf der Nutzer:inneneingabe und dem Antworttext, den das LLM bereits generiert hat. Das LLM fügt die Antwort auf die Nutzereingabe also „Wort für Wort“ zusammen.

Die Funktion, einen Text semantisch und grammatikalisch sinnvoll zu ergänzen, erzeugt aber noch kein für

² Der Leitfaden wird unter dem Titel „Chatbotguide“ im Herbst 2025 veröffentlicht und ab dann über die Website auva.at/publikationen als kostenloser Download zur Verfügung stehen.

³ Für die geneigte Leser:in sei diese Quelle zur weiteren Vertiefung empfohlen, da sie einen gut verständlichen Überblick über die Funktionsweise von LLM gibt, der auch für Leser:innen ohne Hintergrund in der Informatik verständlich sein sollte.

⁴ <https://bbbycroft.net/llm>.

den Menschen als angenehm wahrgenommenes Gespräch. Zur Erreichung eines als natürlich wahrgenommenen Gesprächsverlaufs ist deshalb eine weitere Trainingsphase notwendig. In diesem Schritt wird Nutzer:innenfeedback im Training des Netzes berücksichtigt. Aufgrund der Menge an benötigten Ein- und Ausgaben bewerten Nutzende nicht jede Ausgabe des Netzes. Stattdessen wird ein weiteres, kleineres künstliches neuronales Netz trainiert, welches dazu geschaffen wurde, die Nutzerbewertung zu imitieren (Naveed et al. 2023).

1.2 Daten, Datenschutz und Co

Wie Abschn. 1.1. erläutert, sind moderne LLM-Chatbots aufgrund der Verarbeitung enormer Datenmengen funktionsfähig. Dies wirft insbesondere im Zusammenhang mit privatwirtschaftlichen Anbietern Fragen zum Datenschutz und zur Datensicherheit auf – etwa aufgrund teils unklarer Nutzungsbedingungen und offener Fragen zur Datenverarbeitung. Zusätzlich bestehen urheberrechtliche Unsicherheiten im Hinblick auf das verwendete Trainingsmaterial, das möglicherweise geschützte Werke umfasst, sowie in Bezug auf die rechtliche Einordnung der von Chatbots generierten Inhalte. Im Folgenden wird besonders der datenschutzrechtliche Aspekt vor dem Hintergrund der betrieblichen Anwendung von Chatbots beleuchtet.

Sowohl in der technischen Infrastruktur als auch in der Intransparenz der Funktionsweise der Modelle sind Risiken für die Nutzenden im Umgang mit Daten zu verorten. Privatwirtschaftliche Unternehmen, die Chatbots anbieten, führen diese in der Regel auf ihrer eigenen Serverinfrastruktur aus.⁵ Nutzer:innen greifen lediglich über Websites auf das Programm zu. Dementsprechend rühren Bedenken bezüglich der Datennutzung daher, dass Anbieter:innen von Chatbots die Nutzer:inneneingaben und die Ausgaben des Chatbots auf ihren privaten Servern dauerhaft speichern können. Die so generierten Daten können dann zum Training der nächsten Version des Chatbots genutzt werden, ohne dass Nutzende wissen, welche ihrer Inputs in das Training einfließen und welche der möglicherweise persönlichen oder betriebseigenen Daten in den nächsten Chatbot-Versionen ausgegeben werden. Dies kann z. B. passieren, indem die Chatverläufe einfach als zusätzlicher Text benutzt werden. Es gibt Fälle, in denen LLMs ganze Texte aus dem Training „auswendig gelernt“ haben. Diese Texte ließen sich dann im Dialog mit dem spezifischen Chatbot wiederherstellen (Carlini et al. 2022). Darüber hinaus ist denkbar, dass die Chatverläufe im Hinblick auf die Zufriedenheit der Nutzenden bewertet werden, um sie in den zweiten Trainingsschritt einfließen zu lassen und das

„Bewertungsnetz“ zu trainieren. Da nicht klar ist, wie im LLM Informationen verarbeitet und gespeichert werden, ist auch nicht klar, ob und welche der Trainingsdaten und damit auch Nutzereingaben sich später wieder herstellen lassen.

Daraus wird ersichtlich, dass die Nutzung von Chatbots im betrieblichen Kontext eine komplexe Abwägungsfrage darstellt, wie auch im Folgenden erläutert wird.

1.3 Stand der Forschung

Der Stand der Forschung zum Einsatz von Chatbots im Bereich von Sicherheit und Gesundheit bei der Arbeit ist derzeit noch begrenzt. Wie der nachfolgende Überblick zeigt, wurden in den vergangenen Jahren Chatbots in unterschiedlichen Anwendungsbereichen erprobt und die Ergebnisse veröffentlicht. Dabei reicht das Spektrum von eigenständig entwickelten oder angepassten Chatbot-Lösungen bis hin zur Nutzung bestehender, kommerziell verfügbarer LLM-Systeme. Beispielsweise wurden komplexe arbeitsmedizinische Fragestellungen sowohl von einem Chatbot als auch Arbeitsmediziner:innen beantwortet und die Antworten von Expert:innen bewertet (Padovan et al. 2024, 2025). Weitere Anwendungsfälle finden sich in Berufsfeldern wie dem Gesundheitswesen, dem Personalwesen und dem Bildungsbereich (Ooi et al. 2023).

Chatbots zeigen zudem Potenzial in der Aus- und Weiterbildung, da sie einen niederschweligen und abwechslungsreichen Zugang zu Themen wie der Gefahrenerkennung im Bausektor ermöglichen (Uddin et al. 2023). Sie können darüber hinaus bei der Bearbeitung von Routineaufgaben unterstützen. Dies zeigt eine Studie, in der die LLM-gestützte Auswertung von Testfragen, die im Rahmen von Arbeitsschutzunterweisungen gestellt worden sind, zu einer Zeitersparnis gegenüber manuellen Verfahren führte (Arici et al. 2023).

Auch im Bereich des Risikomanagements im Baugewerbe konnte ein Chatbot relevante Strategien zur Risikoreduktion identifizieren. Allerdings zeigten sich Schwächen hinsichtlich der Konsistenz der Antworten und der Priorisierung der Risiken (Al-Mhdawi et al. 2023). Das breite Anwendungsspektrum von Chatbots wurde weiterhin bei der Auswertung von Unfalldatenbanken und der Entwicklung darauf basierender Präventionskonzepte zur Identifikation von Unfallschwerpunkten deutlich (Smetana et al. 2024). Ebenfalls aus dem Bausektor finden sich Ansätze, bei denen mittels LLM das komplexe Regelwerk „sprechender“ gemacht werden soll. Die dialogbasierte Interaktion, bei der ein Chatbot sozusagen einen Teil des Regelwerks befragt, könnte die Informationssuche und auch die Zugänglichkeit von Fachtexten erhöhen (Tran et al. 2023).

Dennoch zeigt sich, dass die Qualität der Ergebnisse maßgeblich von der Qualität des zugrundeliegenden Trainingswissens sowie den bereitgestellten Kontextinformatio-

⁵ Chatbots können aber grundsätzlich, wie andere Computerprogramme auch, offline und lokal auf Computern ausgeführt werden.

nen der Chatbots abhängt (Padovan et al. 2024). Diese Faktoren beeinflussen die Vollständigkeit und Genauigkeit der Antworten. In bestimmten Einsatzfeldern – etwa der Cybersicherheit – waren die Ergebnisse entsprechend unzureichend (Kereopa-Yorke 2024). Andere Autor:innen weisen zudem auf die „Übervereinfachung“ komplexer Sachverhalte hin, sowie darauf, dass bei sicherheitsrelevanten Fragen (nicht nur auf den Arbeitskontext bezogen) häufig eine Überbetonung individueller Verantwortung zu beobachten ist (Oviedo-Trespacios et al. 2023).

Zudem ist die Einführung und Nutzung von Chatbots stets im Zusammenspiel von Mensch, Technik und Organisation zu betrachten. So weisen Santos und Victoria (2024) beispielsweise darauf hin, dass eine zunehmende Technologisierung im arbeitsmedizinischen Kontext potenziell die Beziehung zwischen Patient:innen und Fachpersonal beeinträchtigen kann.

Insgesamt lässt sich festhalten, dass es erste Ansätze zur Nutzung von Chatbots im Bereich von Sicherheit und Gesundheit bei der Arbeit gibt, die Arbeitsschutzakteure unterstützen können. Die inhaltlichen Schwerpunkte, aber auch die genutzten Sprachmodelle variieren dabei stark und spiegeln vor allem erste Erprobungen wider. Zudem wird deutlich, dass komplexe Funktionen und Anwendungsfälle meist mit speziellen Anwendungen umgesetzt werden, die weit über die Fähigkeiten leicht zugänglicher Basisversionen von Chatbots hinausgehen. Insgesamt besteht weiterhin erheblicher Forschungsbedarf, insbesondere in einem so sensiblen Feld wie dem Arbeits- und Gesundheitsschutz.

2 Chatbots für den Arbeitsschutz nutzen? Methodischer Ansatz und erste Erfahrungen

Vor dem Hintergrund der skizzierten technischen Herausforderungen sowie der auf Grundlage des aktuellen Forschungsstands identifizierten Potenziale und Risiken verfolgt das Projekt eine explorative, praxisorientierte Analyse von Chatbots mit zwei zentralen Zielsetzungen: 1) Es sollen Aussagen darüber getroffen werden, ob und in welchem Umfang Chatbots im Kontext von Sicherheit und Gesundheit bei der Arbeit genutzt werden können. 2) Es sollen Empfehlungen abgeleitet werden, um betriebliche Anwender:innen für mögliche Grenzen und Risiken zu sensibilisieren.

Das Vorgehen gestaltete sich dabei wie folgt: Zunächst wurde ausgehend von der dargestellten Studienlage und explorativen Erhebungen bei Fachkräften für Arbeitssicherheit ein Modell entwickelt, das die Anwendungsfälle von Chatbots im Arbeitsschutz nach drei Themenbereichen strukturiert. Im Anschluss daran wurden Fallbeispiele für typische Fragestellungen des Arbeitsschutzes in Form von

Szenarien entwickelt, getestet und analysiert.⁶ Es folgte eine Evaluation im Rahmen eines Expert:innen-Workshops. Ausgehend von den Erfahrungen wurden Empfehlungen abgeleitet.

In diesem Zusammenhang sei darauf hingewiesen, dass es aufgrund des dynamischen Entwicklungsgeschehens nicht möglich ist, alle Chatbot-Funktionen ausführlich und vollumfänglich darzustellen. Vielmehr geht es darum, auf einer hohen Abstraktionsebene Aussagen zu Möglichkeiten und Grenzen zu treffen und damit praktische Leitplanken zu beschreiben. Es handelt sich nicht um Produktempfehlungen. Wie die vorangegangenen Ausführungen zeigen, gibt es zudem verschiedene Arten, um Sprachmodelle zu trainieren, die wiederum mit verschiedenen Einschränkungen und Möglichkeiten im Hinblick auf den Funktionsumfang einhergehen. Aus Gründen der Zugänglichkeit und um das Potenzial frei zugänglicher Basismodelle auszuloten, wurden die im Folgenden beschriebenen Fallbeispiele anhand von kostenfreien Chatbots (ChatGPT, Copilot, Gemini und Perplexity) erstellt. Dabei ist zu berücksichtigen, dass ebendiese in der Regel lediglich eine Anmeldung mit einer gültigen E-Mail-Adresse sowie der Zustimmung zu den Nutzungsbedingungen bedürfen. Letztere kann auch beinhalten, dass die Sprachmodelle an den Dialogen trainiert werden dürfen und diese somit als Trainingsdaten für Folgeversionen dienen. Darüber hinaus können die Verarbeitungsgeschwindigkeit, die Menge zu verarbeitender Dokumente oder weitere Zusatzfunktionen eingeschränkt sein. Kostenpflichtige Zugänge enthalten häufig andere Regelungen und bieten weitere Möglichkeiten, z. B. in Bezug auf das Training der Modelle mit eigenen Daten. Ebenfalls nicht berücksichtigt wird das Generieren von Bildern mit Hilfe von Sprachmodellen.

2.1 Strukturierung der Anwendungsfälle

Die vorliegende Klassifikation der Anwendungsfälle basiert auf einer Kombination aus Literaturrecherche, Analyse technischer Fertigkeiten generativer Sprachmodelle (v. a. Verarbeitung von Text) sowie praktischen Erfahrungen aus der Arbeitsschutzpraxis. Es handelt sich um einen explorativen Ansatz, der nicht auf einem formalisierten Studiendesign beruht. Darauf aufbauend konnten drei übergeordnete Anwendungsfälle für die Chatbot-Nutzung identifiziert werden (siehe auch Abb. 1), die im Folgenden kurz erläutert werden.

⁶ Die Konstruktion bzw. Auswahl der Beispiele erfolgte auf Basis von Erfahrungen der Projektteammitglieder und in Rücksprache mit Fachleuten bzw. Berater:innen aus dem Bereich von Sicherheit und Gesundheit bei der Arbeit in der BAuA und AUVA.

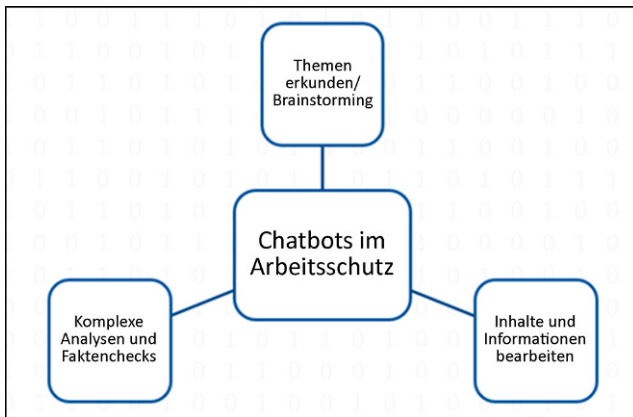


Abb. 1 Überblick über Nutzungsmöglichkeiten von Chatbots im Arbeitsschutz (eigene Darstellung)

Fig. 1 Overview of potential uses of chatbots in occupational safety (own illustration)

Brainstorming und Themen erkunden Hierbei handelt es sich um Fragestellungen, bei denen das intuitive Dialogformat Nutzer:innen den Zugang zu einem Thema erleichtert. Die Dialoge können dazu dienen, Themenfelder zu erkunden und Informationen zu sammeln – ohne Sorge vor Bewertung. Perspektivwechsel werden erleichtert und ein erster Überblick über ein Thema kann gewonnen werden. Darüber hinaus können im Dialog mit dem Chatbot auch Ideen weiterentwickelt werden oder Argumente in Form eines Brainstormings durchgespielt werden.

Informationen und Inhalte bearbeiten Eine Vielzahl von Anwendungsfällen lässt sich unter die Bearbeitung von Informationen und Inhalten fassen. Einerseits können eigene Texte bearbeitet werden: die Ausformulierung von Stichpunkten, das Verändern des Stils oder weitere Umformungen sind mit gezielter Dialoggestaltung möglich. Doch auch andere Texte wie Artikel oder Regelwerke können mit Chatbots be- und verarbeitet werden, beispielsweise beim Anfragen von Zusammenfassungen, der Generierung von Leitfragen oder der Überführung in leicht verständliche Sprache.

Ein weiterer Aspekt, der Chatbots zu einem interessanten Werkzeug im Bereich des Arbeits- und Gesundheitsschutzes macht, ist zudem die Mehrsprachigkeit. Inhalte können in verschiedene Zielsprachen übersetzt werden, wobei Tonalität und Kontext berücksichtigt werden.

Komplexe Analysen und Faktenchecks Die Bearbeitung komplexer Fragestellungen ist ebenfalls mit der Hilfe von Chatbots möglich. Die Identifikation und Analyse von Mustern in großen Datenmengen, Datenaggregationen und Konzept- bzw. Strategieentwicklung bieten großes Potenzial. Die Voraussetzung dafür ist in der Regel jedoch ein speziell trainiertes Modell. Auch die Datenqualität, zuvor

angegebene Regeln und Kontextinformationen beeinflussen, inwieweit Chatbots für komplexe Analysen genutzt werden können.

2.2 Entwicklung von Fallbeispielen

Aufbauend auf den drei Anwendungsfällen wurden ausgehend von der Literaturrecherche (Abschn. 1.3.) und dem Vorwissen der Autor:innen zu jedem Anwendungsfall verschiedene Fallbeispiele ausgearbeitet.

Bei den Fallbeispielen wurde eine Situation geschildert, in der Arbeitsschutzfachleute (z.B. Sicherheitsfachkraft, Arbeitsmediziner:in, Aufsichtsperson) oder andere für den Arbeitsschutz im Betrieb verantwortliche Personen einen Chatbot in einem der drei o. g. Bereiche verwenden. Es wurde darauf geachtet, schlaglichtartig einen diversen Kanon an Situationen abzudecken, in denen Arbeitsschutzfragen relevant werden können. Demnach behandeln die Szenarien unterschiedliche Situationen und Aufgabestellungen, die im Arbeitsalltag anfallen, z.B. die Gefährdungsbeurteilung und die Erstellung von Schulungsmaterial. Es sei darauf hingewiesen, dass die Fallbeispiele nicht vollständig trennscharf sind und Bestandteile der einzelnen Dialoge verschiedenen Anwendungsfällen zuordenbar sind. Je nach Situation wurden Anfragen an den Chatbot auch ohne expliziten Arbeitsschutzkontext gestellt. Bei dem Szenario zur Beschaffung einer Standbohrmaschine wurde beispielsweise allgemein danach gefragt, was bei der Beschaffung einer Standbohrmaschine zu beachten sei. Tab. 1 gibt einen Überblick über die Fallbeispiele.

Zu allen Fallbeispielen wurde ein Dialog mit einem Chatbot erstellt. Beim Design der Eingaben wurde darauf geachtet, im Stil von potentiellen Nutzer:innen zu formulieren. Die Dialoge wurden „in der Rolle“ weitergeführt, bis das Anliegen in dem jeweiligen Szenario bearbeitet worden war oder hinreichend Informationen gesammelt wurden.

Als Chatbots wurden ChatGPT, Copilot, Gemini und Perplexity verwendet. Tab. 2 zeigt einen exemplarischen Chatverlauf aus dem Szenario „Hautschutz auf der Baustelle“. Der Chatbot führt eine Reihe relevanter Argumente an, die aus Sicht von Arbeitgeber:innen bedeutsam sein können wie beispielsweise finanzielle Vorteile oder auch Unternehmensimage. Insbesondere unter „Gesundheit und Wohlbefinden der Mitarbeiter“ fehlen aber zentrale Punkte wie beispielsweise Hautkrebsgefahr durch vermehrte Sonneneinstrahlung.

Das Projektteam begutachtete die Verläufe und arbeitete erste Risiken und Chancen des Einsatzes von Chatbots im Arbeitsschutz heraus. Dazu gehört beispielsweise das Auftreten von Halluzinationen (in einem Fallbeispiel wurde eine Verordnung, die nicht existiert, herangezogen), aber auch das Auftreten von Fehl- und Falschinformationen. Dies ist insbesondere möglich, wenn vielschichtige Abhängigkeiten

Tab. 1 Überblick über Anwendungsfälle und Fallbeispiele (Auswahl); die Inhaltsbeschreibung enthält die jeweils gewählte Personenbeschreibung
Table 1 Overview of selected use cases and examples; content descriptions use gendered forms according to the chosen example (own illustration)

Anwendungsfall	Fallbeispiel	Inhalt
Themen erkunden und Brainstorming	Anschaffung einer Standbohrmaschine	Protagonist des Szenarios ist ein Mitarbeiter einer Werkstatt, der eine neue Standbohrmaschine kaufen möchte. Er informiert sich erst allgemein über die Aspekte, die beim Kauf der Maschine zu beachten sind, um dann später Nachfragen zu, vom Chatbot angebrachten, Arbeitsschutzthemen zu stellen
	Hautschutz auf der Baustelle	Ein Arbeitsmediziner sucht Argumente, um den Vorsitzenden einer großen mittelständischen Baufirma davon zu überzeugen, Hautschutzprodukte für die Beschäftigten bereitzustellen
	Gewerbeaufsicht im Handel	Eine Gewerbeaufsichtsbeamtin, die vorher primär für das Gastgewerbe verantwortlich war, soll einen Betrieb im Handel besichtigen und informiert sich über Arbeitsbedingungen, welche die Sicherheit und Gesundheit von Beschäftigten im Handel beeinträchtigen können.
Inhalte und Informationen bearbeiten	Zusammenfassung von Gesetzestexten	Beschäftigten sollen die Grenzwerte aus dem Arbeitsschutzgesetz nähergebracht werden. Eine Sicherheitsfachkraft bittet daher einen Chatbot den Gesetzestext in der Anfrage zusammenzufassen.
	Zusammenfassung von Normen zur Elektrotechnik	In diesem Beispiel möchte ein ungelernter Mitarbeitender in Abwesenheit des Fachangestellten die Prüfung einer Elektroinstallation selbst durchführen. Deshalb möchte er sich die Elektrikernormen von einem Chatbot zusammenfassen lassen.
	Sifa im Ausflugslokal	Eine Sicherheitsfachkraft arbeitet in einem großen Ausflugslokal, in dessen Küche es vermehrt zu Unfällen mit herunterfallenden Gegenständen kam. Die Sicherheitsfachkraft schlägt daher Arbeitsschuhe in der Küche vor und erstellt mit Hilfe des Chatbots eine Unterweisung sowie ein Quiz zu Schulungszwecken. (Auch wenn dieses Szenario auch zum 1. Themenbereich passen könnte, wird es hier geführt, da der Fokus im Dialog mit dem Chatbot stark auf der Erstellung des Quiz lag.)
Komplexe Analysen und Faktenchecks	Beschaffen einer Fräse	Hier wird eine neue Fräse in einer Werkstatt aufgestellt. In dem Szenario fragt eine Sicherheitsfachkraft, worauf bei der Gefährdungsbeurteilung der Fräse geachtet werden muss, da die Werkstatt besonders eng ist.
	Winterreifen in Tirol	Ein Mitarbeitender einer österreichischen Beratungsfirma nutzt einen Chatbot, um die vorgeschriebene Profiltiefe von Winterreifen zu erfragen.

Tab. 2 Beispielhafter Chatverlauf im Fallbeispiel „Hautschutz auf der Baustelle“, Dialog geführt mit ChatGPT am 24.06.2024

Table 2 Example chat in case study “skin protection on construction sites” (with ChatGPT, 24 June 2024)

Dialogteilnehmer:in Prompt/Antwort	
Nutzer:in	Ich bin Arbeitsmediziner in einer Baufirma. Mit welchen Argumenten kann ich meinen Arbeitgeber von Maßnahmen zum Hautschutz der Bauarbeiter überzeugen?
Chatbot	<i>Als Arbeitsmediziner können Sie den Arbeitgeber von Maßnahmen zum Hautschutz der Bauarbeiter mit den folgenden Argumenten überzeugen:</i> 1. Gesundheit und Wohlbefinden der Mitarbeiter Hautschutzmaßnahmen verhindern Hautkrankheiten wie Dermatitis, die durch den Kontakt mit Zement, Chemikalien und anderen Reizstoffen auf Baustellen entstehen können. Gesunde Haut trägt zur allgemeinen Gesundheit und zum Wohlbefinden der Arbeiter bei, was ihre Zufriedenheit und Moral steigert. 2. Produktivität und Effizienz [...] 3. Rechtliche Anforderungen und Haftung [...] 4. Finanzielle Vorteile [...] 5. Betriebsklima und Mitarbeiterbindung [...] 6. Verbesserung des Unternehmensimages [...] <i>Konkret können Sie dem Arbeitgeber vorschlagen:</i> - Bereitstellung von Schutzcremes und Handreinigungsmitteln - Schulungen zur richtigen Anwendung von Hautschutzmitteln - Regelmäßige Hautuntersuchungen durch medizinisches Fachpersonal - Verbesserung der Arbeitskleidung und Schutzausrüstung - Einrichtung von Hygienestationen auf Baustellen <i>Diese Maßnahmen tragen dazu bei, die Hautgesundheit der Bauarbeiter zu schützen und gleichzeitig die Effizienz und das Image des Unternehmens zu stärken.</i> [...]

Tab. 3 Die Ergebnisse des Workshops: aufbereitet in aller Kürze**Table 3** Workshop results: summarized briefly

<i>Halluzinationen/ Falschaussagen</i>	Einige Behauptungen des Chatbots sind irreführend und falsch (z. B. der Rat Handschuhe bei der Bedienung eine Fräse zu verwenden). Diese Falschinformationen sind besonders tückisch, da sie sich nahtlos in eine Reihe von Informationen einreihen, die ansonsten korrekt scheinen.
<i>(Ungewollte) Sug- gestivfragen</i>	Wenn Nutzende eine Suggestion in ihrer Anfrage haben (wie z. B. im Beispiel der Arbeitsschuhe in der Küche) nehmen Chatbots diese umgehend auf und produzieren vermehrt Antworten, die mit der Suggestion zu tun haben. Wenn diese Suggestion unbewusst passiert, kann ein falsches Bild von Sachverhalten vermittelt werden.
<i>Basis Prompt</i>	Dem Chatbot fehlen Informationen, um die Anfragen einzuordnen. Über die Entwicklung eines Basisprompts, der Informationen über den Arbeitsschutz enthält, ließe sich die Antwortqualität des Chatbots wahrscheinlich verbessern.

und Ausnahmen, wie sie beispielsweise im Rechts- und Regelwerk zu finden sind, thematisiert werden. Der hohe Abstraktionsgrad oder Übergeneralisierungen erschweren den Dialog zusätzlich. In Abhängigkeit von der Dialogführung kann es zudem zu einer Verengung der Ergebnisräume kommen, wenn beispielsweise unbeabsichtigt suggestive Fragen oder erste Lösungsansätze der Nutzer:innen selbst eingegeben werden. Einzelne Aspekte nehmen sodann einen starken Hinweisscharakter ein und werden durch den Chatbot überbetont. Die geordnete Darstellung von Ergebnissen durch Chatbots (bspw. durch Aufzählungen und Hervorhebungen) verbessert den Überblick und die Lesbarkeit, während sie gleichzeitig einen Anschein von Vollständigkeit erweckt.

Im Anschluss wurde ein Expertenworkshop durchgeführt, um die Szenarien und die Ausgaben des Chatbots weiter zu bewerten.

2.3 Post-hoc Validierung durch Expertenworkshop

Zum Zweck der Beurteilung der im vorherigen Abschnitt eingeführten Beispiele sowie der Bewertung der Antworten der Chatbots wurde an der BAuA im Frühjahr 2025 ein dreistündiger Workshop veranstaltet. Bei dem Workshop wurden mit einer Gruppe von verschiedenen Experten⁷ (zwei Experten zur Entwicklung von KI, davon einer spezialisiert auf LLM, vier Arbeitsschutzexperten (1x Brandschutz, 2x Sifa, 1x Sifa und Brandschutz)) die Fallbeispiele sowie die Antworten der Chatbots diskutiert. Die Experten wurden aus der Institution (BAuA) basierend auf vorangegangenen Projekterfahrungen rekrutiert.

Der Workshop wurde nach einem kurzen Kennenlernen in zwei Arbeitsphasen abgehalten. In einem ersten Schritt wurde die Zielsetzung des Workshops durch einen Impulsvortrag über die Funktionsweise von Chatbots und die Motivation hinter dem Kooperationsprojekt verdeutlicht. Nach der Vorstellung der Fallbeispiele setzten sich die Teilnehmer jeweils mit einem dieser in Eigenarbeit detailliert auseinander. Das zur Verfügung gestellte Material für diese Arbeitsphase umfasste einen einführenden Text sowie die

zuvor generierten Chatverläufe. Im zweiten Arbeitsschritt wurden die verschiedenen Fallbeispiele⁸ der Gruppe vorgestellt und anschließend diskutiert. Hierbei lag der Fokus auf der Realitätsnähe der Fallbeispiele an sich. Darüber hinaus wurden die Antworten des Chatbots auf ihre Tauglichkeit für den Kontext der Arbeitssicherheit evaluiert.

Bei dem Workshop handelte es sich nicht um eine systematische qualitative Erhebung im engeren Sinne. Vielmehr wurden zentrale Argumentationslinien und Einschätzungen dokumentiert und im Nachgang thematisch strukturiert aufbereitet, wobei ein besonderer Fokus auf die Anwendbarkeit und Übertragbarkeit der Fallbeispiele im realen Arbeitsalltag gelegt wurde. Dementsprechend stellt die nachfolgende Darstellung eine explorative, qualitativ-deskriptive Zusammenfassung der Diskussionsinhalte dar und bietet erste Einblicke in die Wahrnehmung und Bewertung von Chatbot-Anwendungen im Kontext des Arbeitsschutzes. Die Aufbereitung der Ergebnisse orientiert sich am Protokoll der Diskussionen während des Workshops. Die Analyse ergab verschiedene induktiv abgeleitete Punkte, die nachfolgend zusammengefasst werden. Darüber hinaus sind die Ergebnisse in aller Kürze in Tab. 3 aufbereitet.

Basisprompt Empirische Forschung zu Chatbots ist aufgrund der Dynamik technischer Entwicklungen sowie des Marktes herausfordernd. Mit Blick auf die Praxis wurde es als wichtig erachtet, Akteur:innen des Arbeitsschutzes den Umgang mit Chatbots zu erleichtern. Eine Möglichkeit, dies zu erreichen, besteht darin, einen Basisprompt zu entwickeln. Das ist ein Zusatz zu der Eingabe an den Chatbot. Sie enthält Informationen darüber, dass es sich bei der Eingabe um etwas handelt, was im Arbeitsschutzkontext bearbeitet werden soll. Dem Chatbot werden durch den Basisprompt außerdem zusätzliche Verhaltensweisen auferlegt. Z. B. „Bitte begründe deine Schlussfolgerungen.“ Oder „Wenn du dir unsicher bei einer Antwort bist, antworte bitte nicht, sondern kommuniziere die Unsicherheit“.

⁷ An dem Workshop waren ausschließlich männliche Teilnehmende beteiligt, weshalb an dieser Stelle das Maskulinum verwendet wird.

⁸ Folgende Szenarien wurden im Workshop bearbeitet: Sifa im Ausflugslokal, Anschaffung einer Standbohrmaschine, Einrichtung eines Schweißarbeitsplatzes, Elektrikernormen zusammenfassen, Hautschutz auf der Baustelle, Startup Räumlichkeiten, Gewerbeaufsicht im Handel.

Analyse der Fallbeispiele Die explorative Analyse der Fallbeispiele lässt sich in drei zentrale Bereiche gliedern: fachliche Korrektheit, Kontextverständnis und Interaktionsempfehlungen.

1. Fachliche Korrektheit: Sicherheitsrisiken durch falsche Empfehlungen Die Experten stellten in mehreren Fallbeispielen fest, dass Chatbots inhaltlich fehlerhafte oder sicherheitskritische Empfehlungen geben können.

Dies zeigte sich beispielsweise am Fallbeispiel „Anschaffung einer Standbohrmaschine“, bei dem der Chatbot sowohl das Tragen von Schutzhandschuhen als auch das Reinigen mit Druckluft empfahl – beides Maßnahmen, die bei rotierenden Maschinen eine erhebliche Gefährdung darstellen, und somit im Widerspruch zu grundlegenden Arbeitsschutzprinzipien stehen. Zudem war die Ausgabe des Chatbots relativ lang und richtete stichwortartig Schlaglichter auf einzelne Aspekte des beschriebenen Falls. Dies vermittelte einen Eindruck von Vollständigkeit, obwohl weitere relevante Hinweise (z. B. CE-Kennzeichnung bei der Anschaffung eines Elektrogerätes) fehlten. Noch gravierender wurde das Fallbeispiel bewertet, in dem der Chatbot um eine Zusammenfassung elektrotechnischer Sicherheitsnormen gebeten wurde. Da in dem Fallbeispiel ein mangelndes Grundverständnis skizziert wurde, wäre ein eindeutiges Abraten bzw. der Ausschluss von dieser Arbeit zu empfehlen gewesen. Die Experten schätzten das potenzielle Risiko als „lebensgefährlich“ ein, insbesondere da der Chatbot keine Sicherheitsvorkehrungen wie Unterweisung oder Fachkenntnis explizit einforderte.

2. Kontextsensitivität: Fehlende Einbettung in betriebliche Realität In mehreren Fällen war laut der Experten zu beobachten, dass dem Chatbot der nötige situative Kontext fehlte, um valide Empfehlungen geben zu können. Im Fallbeispiel „Einrichten eines Schweißarbeitsplatzes“ fehlte beispielsweise die Betrachtung der Wechselwirkungen mit anderen Tätigkeiten in der Kfz-Werkstatt vollständig. Ein Experte kommentierte dies treffend als die Beschreibung eines Arbeitsplatzes im Vakuum. Eine mangelnde Kontextsensitivität der Antwort des Chatbots konnte auch im Fallbeispiel „Zusammenfassen von Elektronormen der Sicherheitstechnik“ beobachtet werden. Der Chatbot behandelte die Anfrage isoliert, ohne den Arbeitskontext oder die Qualifikation der anfragenden Person zu hinterfragen. So entstand der Eindruck fachlicher Kompetenz, obwohl sicherheitsrelevante Voraussetzungen nicht erfüllt waren.

3. Kommunikation und Interaktion: Stärken und Fallstricke Neben fachlich-inhaltlichen Aspekten wurde auch die Art und Weise der Interaktion zwischen Chatbot und Nutzer:in thematisiert. Beim Fallbeispiel „Sicherheitsfachkraft im Ausflugslokal“ wurde deutlich, wie stark die Formulierung

der Frage die Antwort des Chatbots beeinflusst. In diesem Fall reagierte der Chatbot direkt auf die implizite Annahme, dass Sicherheitsschuhe notwendig seien, ohne deren Notwendigkeit infrage zu stellen oder Alternativen vorzuschlagen. Die Reflexion darüber mündete in die Erkenntnis, dass LLM-basierte Systeme stark durch die Art der Frage gelenkt werden – ein Umstand, der bei sicherheitskritischen Themen besondere Aufmerksamkeit erfordert. Als positiv hervorgehoben wurde jedoch auch, dass im Fallbeispiel „Bürraumplanung für ein Startup“ der Chatbot von sich aus dazu riet, eine Expertenmeinung einzuholen. Dies wurde als konstruktiver Umgang mit Unsicherheit gewertet und als Beitrag zur Prävention potenzieller Fehlentscheidungen anerkannt.

Insgesamt verdeutlicht die Diskussion des Expertenworkshops, dass Chatbots auch im Bereich des Arbeitsschutzes durchaus Potenzial besitzen, jedoch auch klare Grenzen aufweisen. Insbesondere die fehlende Kontextsensitivität sowie das Risiko fachlich fehlerhafter Empfehlungen machen eine kritische Einordnung und fachliche Kontrolle durch qualifizierte Personen unerlässlich. Chatbots können unterstützend wirken – jedoch nicht als Ersatz für Fachkompetenz und betriebliche Erfahrung.

3 Diskussion und Fazit

Das hier beschriebene Vorgehen weist verschiedene Limitationen auf, die die Verallgemeinerbarkeit der Ergebnisse einschränken. Wie eingangs beschrieben, wurden die Fallbeispiele anhand von frei zugänglichen und kostenlosen Chatbots entwickelt, die lediglich das Anlegen eines Accounts erfordern. Insbesondere im betrieblichen Einsatz ist es denkbar, dass kostenpflichtige Versionen oder Versionen, die speziell an die Bedürfnisse des Betriebs angepasst worden sind, genutzt werden. Dies sollte einen (positiven) Einfluss auf die Antwortqualität haben. Darüber hinaus sind im Rahmen des Projekts eher Chatbots marktführender Anbieter verwendet worden, die nicht explizit einem bestimmten Konversations- oder Informationszweck dienen. Funktionen wie Bildgenerierung wurden im vorliegenden Projekt ausgeklammert, können jedoch ebenfalls eine interessante Perspektive für den Arbeitsschutz bieten. Darüber hinaus wurden alle Fallbeispiele und Dialoge in einem wohlwollenden und menschenzentrierten Ton angelegt. Die intentional negative Nutzung von Chatbots wurde nicht verfolgt. Auch und vor allem weil die Technologien einem ständigen Wandel unterliegen, sind die hier dargestellten Hinweise sehr allgemeiner Natur und bieten Leitplanken für die Anwendung. Eine stetige Überprüfung der Empfehlungen sowie fortlaufendes Monitoring bleiben dennoch erforderlich.

Zusammenfassend lässt sich feststellen, dass die Nutzung von Chatbots insbesondere beim Erkunden von The-

men sowie der Bearbeitung von Inhalten und Informationen eine erfolgsversprechende Anwendung darstellt und einen niedrigschwelligen Zugang zu Themen des Arbeits- und Gesundheitsschutzes ermöglicht. Die Verwendung natürlicher Sprache erleichtert die Anwendbarkeit und die Anpassung an verschiedene (Sprach-)Stile kann die Zufriedenheit beeinflussen (Elsholz et al. 2019). Eine wichtige Voraussetzung für die Nutzung von Chatbots besteht jedoch darin, die Möglichkeiten und Grenzen der Anwendung zu verstehen. Gezielte Eingabeaufforderungen (Koyuturk et al. 2025), beispielsweise über einen Basisprompt, und weitere Kommunikations- sowie Interaktionsstrategien zur Kalibrierung des Vertrauens (Belosevic und Buschmeier 2024) können dies unterstützen. Darüber hinaus lassen sich mehr Freiheitsgrade im Dialog mit dem Chatbot ausschöpfen, wenn die eigene fachliche Expertise in dem Themengebiet, zu dem der Dialog stattfindet, hoch ist. Dies umfasst einerseits das Erkennen von fehlenden oder falschen Informationen ebenso wie die Nutzung von Spielräumen bei der Gestaltung von Ergebnissen. In diesem Zusammenhang spielen weitere Nutzer:inneneigenschaften beim Einsatz von Chatbots eine wichtige Rolle. So zeigt die Studie von Wester et al. (2024), dass Nutzer:innen mit einer höheren technologischen Unsicherheit die Ratschläge eines Chatbots als nützlicher empfanden.

Darüber hinaus besteht eine zentrale Voraussetzung für die Nutzung von Chatbots im Kontext von Sicherheit und Gesundheit bei der Arbeit darin, dass eine betriebliche Übereinkunft über die möglichen Ziele und Zwecke der Nutzung getroffen worden ist. Auch Aspekte wie Datenschutz, IT-Sicherheit, Schulungen sowie Dokumentations- und Kennzeichnungspflichten sollten im Rahmen dessen besprochen werden. Die Nutzung von Chatbots ist damit ein Thema für viele Funktionsträger:innen im Betrieb – von der Geschäftsführung über Datenschutzbeauftragte hin zu Beschäftigtenvertretungen und Personalabteilung. Zudem eröffnet sich hier ein eindeutiges Spannungsfeld zwischen dem Schutz betrieblicher Daten und möglichst genauen Eingaben von Informationen, um erfolgreiche Dialoge führen zu können. Somit sind klare Absprachen und der gemeinsame Austausch innerhalb des Betriebs essentiell für eine zielgerichtete Nutzung von Chatbots im betrieblichen Kontext.

Von eher praktischer Seite verdeutlichte das Projekt, dass ein Nachprüfen der Antworten des Chatbots einen unerlässlichen Prozessschritt darstellt und „blindes“ Vertrauen in die Chatbots unbedingt vermieden werden sollte. Einen wichtigen Beitrag dazu kann das gezielte Prompting – also die Formulierung von Anfragen – liefern. Hier zeigte sich, wie schnell Ergebnisräume beispielsweise durch suggestive Fragen eingeschränkt werden. Chatbots sind meist so programmiert, dass sie ihren Nutzer:innen gefallen und eine angenehme Konversation ermöglichen, was Dialogverläufe

unbemerkt beeinflussen kann. Darauf aufbauend ergibt sich weiterer Forschungsbedarf, der im Folgenden kurz skizziert werden soll.

Auf übergeordneter Ebene wäre es wünschenswert, Chatbots speziell für Anwendungen im Arbeitsschutz zu schaffen. Das Training und die Verfeinerungen, die an einem Chatbot gemacht wurden, lassen sich jedoch nicht ohne Weiteres auf andere Chatbots übertragen. Kurzfristig kann es daher von Vorteil sein, Eingaben für Chatbots zu entwickeln, die die gewünschten Funktionen und Verhaltensmuster klar anweisen: Der bereits erwähnte Basisprompt kann dazu beitragen, dieses Ziel zu erreichen und bietet die nötige Flexibilität für den Einsatz in verschiedenen Chatbots. Weiteres Potenzial wird in *Retrieval Augmented Generation* gesehen: Hier werden Textdaten so abgelegt, dass Fragmente von ihnen, die zu der Anfrage eines/einer Nutzer:in passen, gefunden werden können. Diese Fragmente werden dann automatisch in die Eingabe des Chatbots eingeschleust. So ist es dem Chatbot z.B. möglich Quellen verlässlicher zu zitieren. Erste Studien im Kontext von Unfallanalysen weisen hier auf großes Potenzial hin (Baek et al. 2025).

Weiterer Forschungsbedarf wird zudem darin gesehen, die Antwortqualität von Chatbots im Bereich von Sicherheit und Gesundheit bei der Arbeit besser bewerten zu können. Darüber hinaus sind empirische Studien mit Nutzer:innen essenziell, um deren Verhalten fundiert zu analysieren und daraus evidenzbasierte Grundlagen für die Gestaltung zielgerichteter Schulungsmaßnahmen abzuleiten.

4 Zentrale Empfehlungen für die Praxis

Das Kooperationsprojekt hat gezeigt, dass der Einsatz von modernen Chatbots durch das Anbieten von Ideen und Informationen einen positiven Beitrag zur Effizienz und Effektivität von Arbeitssicherheit und Gesundheitsschutz leisten kann. Gleichwohl sind Chatbots anfällig dafür, Fehlinformationen zu erteilen oder eingeschränkte Perspektiven abzubilden, wodurch mitunter zusätzliche Gefährdungslagen entstehen. Das kritische Hinterfragen und Überprüfen der Antworten von Chatbots durch betriebliche Experten und Fachkräfte sowie Wissen über die Möglichkeiten und Grenzen der Anwendungen sind daher sowohl bei frei zugänglichen als auch bei spezialisierten Chatbots eine wichtige Grundvoraussetzung bei der Nutzung. Zusammengefasst lassen sich folgende Empfehlungen für Praktiker:innen formulieren:

● Betriebliche Vorgaben

Prüfen Sie, ob es in Ihrer Institution interne Richtlinien zur Nutzung von öffentlich zugänglichen Chatbots gibt, und setzen Sie diese Vorgaben um.

• **Datenschutz und Datensicherheit**

Achten Sie bei der Eingabe darauf, dass in Ihrer Frage bzw. Ihrem Prompt weder personenbezogene Daten von Beschäftigten noch Betriebsinterna enthalten sind.

• **Abschätzung der potenziellen Chancen und Risiken**

Überlegen Sie im Vorfeld, für welchen Anwendungsfall Sie den Chatbot einsetzen möchten, welchen Zusatznutzen Sie sich davon erwarten und welche Risiken durch falsche oder mangelhafte Ausgaben entstehen können.

• **Vorwissen zum Themengebiet**

Achten Sie darauf, dass Sie ausreichend Vorwissen zum Themengebiet Ihrer Frage bzw. Ihres Prompts haben, und unterziehen Sie die Antwort einem Faktencheck in Hinblick auf Korrektheit und Vollständigkeit.

• **Formulierungen und Antwortqualität**

Achten Sie auf die Qualität Ihrer Eingabe bzw. Ihres Prompts, indem Sie u. a. die Situation und die Rolle, die der Chatbot einnehmen soll, genau beschreiben. Überprüfen Sie, ob Sie durch die Wortwahl Ihrer Eingabe bzw. Ihres Prompts die Ausrichtung der Antwort unbeabsichtigt einschränken oder in eine bestimmte Richtung lenken.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Al-Mhdawi MKS, Qazi A, Alzarrad A, Dacre N, Rahimian F, Buniya MK, Zhang H (2023) Expert evaluation of chatGPT performance for risk management process based on ISO 31000 standard. SSRN 4504409
- Arici N, Gerevini AE, Olivato M, Putelli L, Sigalini L, Serina I (2023) Real-world implementation and integration of an automatic scoring system for Workplace safety courses in Italian. *Future Internet* 15:268. <https://doi.org/10.3390/fi15080268>
- Arntz M, Baum M, Brüll E, Dorau R, Hartwig M, Lehmer F, Matthes B, Meyer SC, Schlenker O, Tisch A, Wischniewski S (2025) Digitalisierung und Wandel der Beschäftigung (DiWaBe 2.0): Eine Datengrundlage für die Erforschung von Künstlicher Intelligenz und anderer Technologien in der Arbeitswelt, 1. Aufl. baa: Bericht. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund
- Baek S, Park CY, Jung W (2025) Automated safety risk management guidance enhanced by retrieval-augmented large language model. *Autom Constr* 176:106255
- Belosevic M, Buschmeier H (2024) Calibrating trust and enhancing user agency in LLM-based Chatbots through conversational styles. *CUI@ CHI 2024: Building Trust in CUIs—From Design to Deployment*.
- Carlini N, Ippolito D, Jagielski M, Lee K, Tramer F, Zhang C (2022) Quantifying memorization across neural language models. In: *The Eleventh International Conference on Learning Representations*. arXiv:2202.07646
- Dong C, Li Y, Gong H, Chen M, Li J, Shen Y, Yang M (2022) A survey of natural language generation. *ACM Comput Surv* 55(8):1–38 (insbesondere Abschnitt 8, External Knowledge Engagement und Abschnitt 1.1)
- Elsholz E, Chamberlain J, Kruschwitz U (2019) Exploring language style in chatbots to increase perceived product value and user engagement. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, S 301–305
- EU-OSHA (2025) First findings of the fourth European survey of enterprises on new and emerging risks (ESENER 2024). European Agency for Safety and Health at Work (EU-OSHA), Bilbao
- Gao M, Lu T, Yu K, Byerly A, Khashabi D (2024) Insights into LLM long-context failures: when transformers know but don't tell. In: *Findings of the Association for Computational Linguistics: EMNLP 2024*, S 7611–7625
- Haque MA, Li S (2024) Exploring ChatGPT and its impact on society. *AI Ethics*: 1–13
- Kanaka NP (2025) Comparative investigation of Grokking in transformers and beyond: a study on reasoning. The Ohio State University (Master's thesis)
- Kereopa-Yorke B (2024) Building resilient SMEs: Harnessing large language models for cyber security in Australia. *J AI Robot Workplace Autom* 3(1):15–27
- Koyuturk C, Theophilou E, Patania S, Donabauer G, Martinenghi A, Antico C, Ognibene D (2025) Understanding learner-LLM Chatbot interactions and the impact of prompting guidelines. arXiv preprint arXiv:2504.07840
- Kriesel D (2006) Neuronale Netze. Rheinische Friedrich-Wilhelms-Universität, S 18
- Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, Akhtar N, Barnes N, Mian A (2023) A comprehensive overview of large language models. *ACM Trans Intell Syst Technol*
- Ooi KB, Tan GWH, Al-Emran M, Al-Sharafi MA, Capatina A, Chakraborty A, Wong LW (2023) The potential of generative artificial intelligence across disciplines: perspectives and future directions. *J Comput Inf Syst*. <https://doi.org/10.1080/08874417.2023.2261010>
- Open AI (2025) Expanding on what we missed with sycophancy. <https://openai.com/index/expanding-on-sycophancy/>. Zugegriffen: 30. Mai 2025
- Oviedo-Trespalacios O, Peden AE, Cole-Hunter T, Costantini A, Haghani M, Rod JE, Reniers G (2023) The risks of using ChatGPT to obtain common safety-related information and advice. *Saf Sci* 167:106244
- Padovan M, Cosci B, Petillo A, Nerli G, Porciatti F, Scarinci S, Palla A (2024) ChatGPT in occupational medicine: a comparative study with human experts. *Bioengineering* 11(1):57
- Padovan M, Palla A, Marino R, Porciatti F, Cosci B, Carlucci F, Nerli G, Petillo A, Necciari G, Dell'Amico L, Lucisano VC, Scarinci S, Foddìs R (2025) ChatGPT-4 vs. Google Bard: Which Chatbot Better Understands the Italian Legislative Framework for Worker Health and Safety? *Appl Sci* 15(3):2076–3417. <https://doi.org/10.3390/app15031508>

- Salguero-Caparrós F, Pardo-Ferreira MDC, Martínez-Rojas M, Rubio-Romero JC (2020) Management of legal compliance in occupational health and safety. A literature review. *Saf Sci* 121:111–118
- dos Santos ML, Victória VNG (2024) Critical evaluation of applications of artificial intelligence based linguistic models in Occupational Health. *Rev Bras Med* 2:6
- Smetana M, Salles de Salles L, Sukharev I, Khazanovich L (2024) Highway construction safety analysis using large language models. *Appl Sci* 14:1352. <https://doi.org/10.3390/app14041352>
- Tran S, Khan N, Kimito EC, Pedro A, Soltani M, Hussain R, Park C (2023) Extracting information from construction safety requirements using large language model. In: *Proceedings of the 23rd international conference on construction applications of virtual reality*, S 761–767
- Uddin SJ, Albert A, Ovid A, Alsharif A (2023) Leveraging ChatGPT to aid construction hazard recognition and support safety education and training. *Sustainability* 15(9):7121
- Wester J, De Jong S, Pohl H, Van Berkel N (2024) Exploring people's perceptions of LLM-generated advice. *Comput Hum Behav Artif Hum* 2(2):100072
- Wolfram S (2023) What is ChatGPT doing ... and why does it work? Stephen Wolfram writings. <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>. Zugriffen: 26. Mai 2025
- Hinweis des Verlags** Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.