



### **Inhalt**

Einleitung	4
Wie funktionieren moderne Chatbots?	5
Was sind Large Language Models (LLM)?	
Wie verlässlich sind Antworten von LLM-basierten Chatbots?	5
Welche ökologischen und datenschutzrechtlichen Bedenken gibt es?	6
Wie werden Chatbots verwendet?	6
Kurz und bündig	
Chatbots im Arbeits- und Gesundheitsschutz	
Stand der Forschung	8
Betriebliche Voraussetzungen	g
Besonderheiten	g
Kurz und bündig	10
Anwendungsfälle	11
Themen erkunden/Brainstorming/Informieren	11
Inhalte und Informationen bearbeiten	13
Komplexe Analysen und Faktenchecks	
Kurz und bündig	18
Das Wesentliche zum Abschluss	19
Anhang I: Glossar – Fachbegriffe kurz erklärt	20
Anhang II: Literatur	22



Moderne Chatbots, die auf großen Sprachmodellen beruhen, sind mittlerweile in unserem Alltag angekommen. Fragen an Chatbots zu stellen, um sich über bestimmte Themen zu informieren oder Inspirationen zu erhalten, wird auch im Arbeitskontext zunehmend üblich. Der Bereich von Arbeitssicherheit und Gesundheitsschutz ist davon nicht ausgenommen. Mit dieser Entwicklung sind Chancen, aber auch Herausforderungen für die Sicherheit und Gesundheit bei der Arbeit verbunden.

Die deutsche Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) und die österreichische Allgemeine Unfallversicherungsanstalt (AUVA) haben es sich daher zur Aufgabe gemacht, die positiven Möglichkeiten der Chatbot-Nutzung zu stärken, ohne problematische Aspekte aus den Augen zu verlieren. Mit dem vorliegenden Leitfaden sollen Arbeitsschutz-Fachleute dabei unterstützt werden, moderne Chatbots im Rahmen ihrer Arbeitstätigkeit bewusst und reflektiert zu nutzen. Der Fokus liegt auf sprachbasierten Chatbots. Anwendungen im Bereich der Bild- und Videokommunikation o. Ä. sind ausgeklammert.

Wenngleich der Leitfaden keine wissenschaftliche Studie darstellt, erfolgte seine Entwicklung unter Berücksichtigung von aktuellen wissenschaftlichen Artikeln zur Chatbot-Nutzung im Bereich von Arbeitssicherheit und Gesundheitsschutz. Zudem flossen experimentelle Anwendungsbeispiele der Chatbot-Nutzung und Erkenntnisse aus Feedback-Schleifen mit Fachleuten in den Leitfaden ein. Darauf aufbauend wurden praxisorientierte Empfehlungen für Personen entwickelt, die

eine professionelle Rolle im Bereich von Sicherheit und Gesundheitsschutz am Arbeitsplatz einnehmen.

Der Leitfaden gibt zunächst eine Basisinformation zur Funktionsweise von modernen Chatbots. Nach allgemeinen Überlegungen zur Nutzung von Chatbots steht ihre Anwendung im Arbeits- und Gesundheitsschutz im Mittelpunkt. Das Herzstück des Leitfadens bilden Anwendungsfälle, die in die drei Bereiche "Themen erkunden / Brainstorming / Informieren", "Inhalte und Informationen bearbeiten" und "Komplexe Analysen und Faktenchecks" gegliedert sind. Praktische Beispiele veranschaulichen die Anwendungsfälle, die wesentlichen Erkenntnisse sind in Tabellenform zusammenfasst. Wichtige Informationen sind in Kästchen und Grafiken übersichtlich dargestellt. Die Unterkapitel "Kurz und bündig" am Ende eines jeden Kapitels geben einen Überblick über die wesentlichen Inhalte. Erklärungen von Fachbegriffen finden Sie im Glossar. Fachbegriffe sind bei ihrer ersten Verwendung im Text unterstrichen und in der PDF-Version dieses Leitfadens mit ihrer Erklärung im Glossar verlinkt.

### Wie funktionieren moderne Chatbots?

Seit der Veröffentlichung von <u>ChatGPT</u> Ende 2022 sind <u>Chatbots</u> aus dem Alltag vieler Menschen nicht mehr wegzudenken. In vielen Kontexten gibt es Programme, mit denen Informationen in Form eines Chatverlaufs ausgetauscht werden können. Diese Programme, die sogenannten Chatbots, sind in der Lage, teils komplexe Fragestellungen sinnvoll zu beantworten.

In den folgenden Abschnitten wird die Funktionsweise solcher Programme kurz erklärt. Herzstück moderner Chatbots ist ein großes Sprachmodell, im Englischen "Large Language Model" oder kurz LLM. In diesem Dokument beziehen wir uns ausschließlich auf Chatbots, denen ein LLM zugrunde liegt.

### Was sind Large Language Models (LLM)?

Ein LLM ist ein Computerprogramm, welches einen unvollständigen Text als Eingabe bekommt und als Ausgabe jenes Wort an den Text anfügt, welches ihn am passendsten weiterführt. Stellt man einem Chatbot eine Frage, wird diese Frage als Eingabetext für ein LLM weitergegeben. Dieses LLM erzeugt Wort für Wort die Antwort, die vom Chatbot ausgegeben wird. Die Aufgabe, das passendste Wort zu finden, ist nicht einfach zu erfüllen. Für das passendste Wort ist nicht nur entscheidend, dass es grammatikalisch den Regeln der Sprache folgt, sondern auch, dass es inhaltlich sinnvoll zum Text passt.

Um sinnvolle Wörter finden zu können, sind LLM der Struktur unseres Nervensystems nachempfunden. Im Kern ist ein LLM ein großes, <u>künstliches neuronales Netz</u> – eine Software also, die, grob vereinfacht, wie ein tatsächliches Nervensystem eines Organismus funktioniert. Dieses künstliche neuronale Netz lernt die richtigen Worte vorherzusagen, indem es mit Lückentexten <u>trainiert</u> wird. Ein Modell, welches diese Lückentexte erfolgreich

vervollständigt, wird <u>Basismodell</u> (engl. Foundation Model) genannt. Die Fähigkeit, das nächste Wort in einem beliebigen Lückentext vorherzusagen, setzt voraus, dass bis zu einem gewissen Maß sinnvolle Antworten gegeben werden können. Es befähigt noch nicht dazu, gute Gespräche zu führen. Zur Feinabstimmung wird das Basismodell in einer zweiten Trainingsphase mithilfe von menschlichem Feedback darauf trainiert, gute Gespräche zu führen.

Um alle Aspekte unserer Sprache abzudecken, müssen LLM sehr groß sein und es müssen viele Texte verwendet werden, um sie zu trainieren. GPT-4, das LLM hinter ChatGPT 4, hat 1,7 Billionen¹ verschiedene Neuronen und es ist davon auszugehen, dass der gesamte im öffentlichen Internet verfügbare Text zu seinem Training benutzt wurde. Durch die enorme Größe der Systeme gibt es nur wenige Firmen, die in der Lage sind, LLM zu erzeugen, und die in Frage kommenden äußern sich nicht konkret über ihre Trainingsstrategie oder die zum Training genutzten Texte.

## Wie verlässlich sind Antworten von LLM-basierten Chatbots?

Durch die Größe der Systeme ist es selbst den Wissenschaftlern:Wissenschaftlerinnen, die diese Systeme erzeugt haben, nicht möglich zu sagen, wie genau das neuronale Netz in der Lage ist, diese Aufgabe zu erfüllen. Das ist vergleichbar mit neuronalen Netzen in der Biologie: Wie einzelne Nervenzellen funktionieren, hat der Mensch gut verstanden. Trotzdem ist es ihm z. B. nicht möglich, im Detail zu erklären, wie das menschliche Denken funktioniert.

Da nicht restlos geklärt ist, wie LLM Antworten generieren, ist es auch nicht möglich, Aussagen zu treffen wie z.B.: "Die Antwort des LLM ist immer korrekt (wenn

die Frage wie folgt formuliert ist)." Zurzeit ist die einzige Möglichkeit sicherzustellen, ob die Aussage eines LLM richtig ist oder nicht, sie selbst zu überprüfen. Insbesondere wird in der Problemstellung, für die das LLM eine Lösung finden soll ("Was ist das passendste nächste Wort?"), nicht berücksichtigt, dass die Ausgabe inhaltlich korrekt sein muss. Wenn also eine Anfrage gestellt wird wie: "Mein Name ist Max Mustermann, ich wohne in Göttingen. Nenne mir den Namen der Rockband, in der ich spiele, und den Titel unseres größten Hits", dann liefert das System als Antwort irgendeinen Bandnamen und den Namen irgendeines Songs, einfach, weil das zu der Anfrage passt.

<sup>1</sup> Genau genommen handelt es sich um 1,7 Billionen (1,7×10<sup>12</sup>) <u>Parameter</u>; nicht alle davon beschreiben ein Neuron.

# Welche ökologischen und datenschutzrechtlichen Bedenken gibt es?

Bei jedem Wort der Ausgabe, das ein Chatbot generiert, muss das gesamte LLM einmal durchlaufen werden. Je größer das LLM, desto mehr Rechenoperationen sind dafür erforderlich – dies kostet viel Energie. Es wird daher empfohlen abzuwägen, wie notwendig die Nutzung eines LLM zur Lösung eines Problems ist. Viele Anwendungsfälle bzw. Antworten können ebenso mit etablierten Suchmaschinen oder im Gespräch mit Kollegen:Kolleginnen gefunden werden.

Der 18-tägige Betrieb eines LLM setzt (bei moderater Auslastung) etwa 340 kg  ${\rm CO}_2$  frei. (Luccioni et al., 2023).

Dies entspricht etwa der CO<sub>2</sub>-Menge, die ein SUV auf der Fahrt von Wien nach Berlin und zurück emittiert.

Außerdem gibt es Datenschutzaspekte, die im Umgang mit öffentlich im Netz verfügbaren Chatbots oft diskutiert werden. Um diese einzuordnen, sei hier der Schritt beleuchtet, der ein LLM zu einem Chatbot werden lässt. Ein LLM ist ein ganz normales Computerprogramm, welches lokal auf einem Rechner ausgeführt wird. Es nimmt Text entgegen und gibt das "passendste" Wort zurück. Darüber hinaus sind klassische LLM "statisch", d. h. das Modell lernt nur, wenn es auf Basis der Daten, die im Trainingsdatensatz hinterlegt sind, trainiert wird.

Es lernt nicht eigenständig und auch nicht durch Interaktion mit Nutzenden. Um das LLM als Chatbot im

Internet nutzbar zu machen, binden viele herstellende Unternehmen ihr LLM in eine Website ein. Diese Kombination aus LLM und der umgebenden Website ist dann der Chatbot.

Die meisten der diskutierten Datenschutzaspekte hängen mit dieser Einbettung zusammen. Firmen, die Chatbots anbieten, haben die Möglichkeit, die Eingaben der Nutzenden zu speichern, um diese z. B. für das Training der nächsten Chatbot-Version zu verwenden. Dieser Umstand ist die Hauptursache für die Datenschutzbedenken.

Die Forschung zu künstlichen neuronalen Netzen zeigt, dass beim Training verwendete Daten wieder aus dem Modell hervorgeholt werden können. Das heißt, dass jede:r Nutzende einer späteren Version potenziell die vorangegangenen Eingaben aller anderen Nutzenden einsehen könnte.

Es wird daher empfohlen, den Chatbot wie einen firmenexternen beruflichen Kontakt zu behandeln, etwa wie unbekannte Kunden:Kundinnen, die auf einer Messe angetroffen werden. Der Dialog mit einem öffentlich zugänglichen Chatbot sollte grundsätzlich nicht als vertraulich betrachtet werden (Details siehe AGB der Anbieter:innen).

Darüber hinaus gibt es die Möglichkeit, offen zugängliche und vortrainierte LLM (z.B. Mixtral von Mistral Al) zu nutzen und darauf aufbauend selbst einen Chatbot, beispielsweise im Firmenintranet, bereitzustellen. Die technische Umsetzung ist allerdings anspruchsvoll.

### Wie werden Chatbots verwendet?

Die Anwendung von Chatbots erfordert in der Regel eine persönliche Registrierung zur Nutzung der Dienste. Die Details sind in den jeweiligen Nutzungsbedingungen geregelt. Die Kommunikation mit Chatbots erfolgt dialogorientiert. Demnach gestaltet sie sich wie ein Gespräch oder eben ein Chat im Frage-Antwort-Stil. Das Fragestellen – in diesem Kontext "Prompting" genannt – kann in verschiedenen Sprachen erfolgen und beliebig lang oder kurz sein.

### Fragen bzw. Prompting

Das Prompting hat einen großen Einfluss auf die Art und Qualität der Antworten des LLM. Es lassen sich bereits zahlreiche Tipps dazu finden, wie Dialoge mit Sprachmodellen verbessert werden können.

#### Die richtigen Fragen stellen





"Was ist bei der Prävention von Unfällen mit herunterfallenden Gegenständen zu beachten?"



"Können Sicherheitsschuhe bei der Prävention von Unfällen mit herunterfallenden Gegenständen helfen?"

Die zweite Formulierung hat einen einschränkenden Charakter und fokussiert die Antworten direkt auf den Begriff Sicherheitsschuhe, während andere Lösungsansätze vernachlässigt werden. Diesen gemein ist die Empfehlung, die Frage und die zu ihrer Beantwortung benötigten Informationen (z. B. die Ist-Situation und Hintergründe) möglichst präzise zu formulieren. Gleichzeitig ist darauf zu achten, die Frage so offen zu halten, dass die Antwortmöglichkei ten nicht vorab eingeschränkt oder in eine bestimmte Richtung gelenkt werden. Ein nachträgliches Spezifizieren von Fragen und Antworten ist in der Kommunikation mit dem System möglich.

### **Antworten**

Sprachmodelle unterscheiden sich unter anderem darin, ob und wie häufig neue Versionen verfügbar sind. Sprachmodelle spiegeln den Wissensstand und die Aktualität der Trainingsdaten wider. Einige Modelle können tagesaktuelle Informationen über die Anbindung an Suchmaschinen abrufen. Des Weiteren können Nutzende über zusätzliche Informationseingaben in den Prompt das Antwortverhalten beeinflussen. Es ist zu beachten, dass Sprachmodelle verschiedener Anbieter:innen einen unterschiedlichen Umgang damit pflegen, wie prominent Informationsquellen dar-

gestellt werden. Während beispielsweise der Chatbot Perplexity in den Antworten hinterlegt, woher die Aussagen stammen, wird diese Angabe bei anderen LLM nur auf Nachfrage geteilt. In manchen Fällen kann es auch zu sogenannten Halluzinationen des Systems kommen, bei denen fiktive Inhalte generiert werden wie etwa das Anführen einer Quelle, die real nicht existiert. Das nachfolgende Beispiel zeigt eine solche Halluzination, die aufgrund fehlender Quellenangaben nicht auf den ersten Blick erkennbar ist.



Kannst du mir für die OSHA-Studie und die HSE-Studie jeweils die Titel nennen, damit ich sie im Internet suchen kann?

Natürlich, hier sind die Titel der Studien, die Ihnen bei der Suche nach detaillierten Informationen helfen können:



- 1. OSHA (Occupational Safety and Health Administration)
- Titel: "Culinary and Food Service Workers Safety"
- Beschreibung: Diese Ressourcen und Berichte der OSHA behandeln die Sicherheit von Mitarbeitern in der Gastronomiebranche und enthalten häufige Verletzungsarten sowie Präventionsstrategien.
- 2. HSE (Health and Safety Executive)
- Titel: "Slips and trips in the food and drink industries"
- Beschreibung: Diese Studie konzentriert sich auf das Auftreten und die Prävention von Ausrutsch-, Stolper-, und Sturzunfällen in der Lebensmittel- und Getränkeindustrie.

Bei beiden Angaben handelt es sich nicht direkt um wissenschaftliche Studien, obwohl die Wortkombination in Dokumenten der Institutionen zu finden ist.



Abbildung 1: Chatprotokoll (gekürzt) vom 20.08.2024 mit ChatGPT

### Kurz und bündig

Ein Chatbot ist ein Computerprogramm, welches mit Nutzenden in einem Chatverlauf interagiert. Die Teile des Chatverlaufs, die der Chatbot beiträgt, stammen bei modernen Chatbots aus einem Large Language Model (LLM). Dieses ist mit Trainingstexten bestückt, die mutmaßlich aus dem öffentlich zugänglichen Teil des Internets stammen. Die Antwort des LLM stellt –

basierend auf den Trainingsdaten – eine wahrscheinliche Fortsetzung der vorangegangenen Konversation dar. Im Speziellen bedeutet das, dass die Ausgaben des Chatbots nicht richtig sein müssen und kritisch hinterfragt werden sollten. Die Interaktion mit dem Chatbot erfolgt über ein Frage-Antwort-Format. Fragen sollten dabei möglichst präzise und offen gestellt werden.



### Stand der Forschung

Mit den beschriebenen Fähigkeiten und dem Wissen des Internets bieten Chatbots auch ein enormes Potenzial für Anwendungen in der Arbeitswelt. Vereinzelte Studienbeiträge betrachten den Einsatz von Chatbots im Bereich von Sicherheit und Gesundheit bei der Arbeit und beschreiben deren Potenzial. Die inhaltlichen Schwerpunkte sind derzeit noch sehr unterschiedlich gelagert und spiegeln erste Erfahrungen wider. Weitere Forschung ist erforderlich.

In den vergangenen Jahren sind verschiedene Anwendungsfälle erprobt und dokumentiert worden wie beispielsweise der Einsatz von Chatbots in der Arbeitsmedizin (Padovan et al., 2024), Potenziale von generativen KI-Modellen in verschiedenen Berufsfeldern, darunter dem Gesundheitswesen, dem HRund dem Bildungsbereich (Ooi et al., 2023), oder der Einsatz von KI-basierten Chatbots (u. a. für Arbeitsschutzfragen) in Gewerkschaftsorganisationen (Flanagan / Walker, 2020). Chatbots unterstützen zudem bei sicherheitsrelevanten Professionen in der Ausbildung. Sie bieten einen einfachen und abwechslungsreichen Zugang zu Themen wie Gefahrenerkennung (Uddin et al., 2023). Anwendungen wie die automatisierte Auswertung von Testfragen bei der Überprüfung von Arbeitsschutzwissen, die beispielsweise im Rahmen von Unterweisungen zum Einsatz kommen können (Arici et al., 2023), führten zu einer beachtlichen Zeitersparnis im Vergleich zur "händischen" Auswertung. Im Bereich des Risikomanagements, das hinsichtlich der Gefährdungsbeurteilung vergleichbare Züge aufweist, konnte ein Chatbot relevante Strategien zur Verminderung von Risiken (hier in der Baubranche) identifizieren, jedoch fielen die Konsistenz der Antworten sowie die

Priorisierung vermeintlicher Risiken weniger zufriedenstellend aus (Al-Mhdawi, 2023). Auch eine Studie zur Auswertung von Unfalldatenbanken und die nachfolgende Entwicklung von Präventionskonzepten zur Identifikation von Unfallschwerpunkten (Smetana et al., 2024) zeigten die große Bandbreite an Möglichkeiten der Modelle auf.

Es bleibt aber zu beachten, dass die Modelle nur so gut sein können wie das Wissen, an dem sie gelernt haben und auf dessen Basis sie trainiert worden sind. Auch die zur Verfügung gestellten Kontextinformationen spielen in diesem Zusammenhang eine Rolle (Padovan et al., 2024). All dies wirkt sich auf die Vollständigkeit und Genauigkeit der Antworten aus – für einen Anwendungsfall im Bereich Cybersicherheit waren sie beispielsweise nicht zufriedenstellend (Kereopa-Yorke, 2024). Die Einführung und Nutzung von Chatbots sind zudem immer im Zusammenspiel von Mensch, Technik und Organisation zu betrachten. In diesem Zusammenhang stellen Santos und Victoria fest (2024), dass im Kontext von arbeitsmedizinischen Fragestellungen eine weitere Technologisierung die Beziehung zu den Patienten:Patientinnen mitunter schwächen kann.

### Betriebliche Voraussetzungen

Bevor frei zugängliche Chatbots im betrieblichen Kontext eingesetzt werden, sollte zunächst geprüft werden, ob innerbetriebliche Vereinbarungen oder Richtlinien zum Umgang mit diesen vorliegen. Diese betreffen häufig den Schutz von personenbezogenen Daten der Beschäftigten und von Betriebsinterna, die durch Eingaben in das System möglicherweise unkontrolliert an die Öffentlichkeit gelangen. Grundsätzlich gehören sensible Daten nicht in einen Chatbot. Das bedeutet, dass Klarnamen und weitere zuordenbare Daten von Personen, aber auch betriebsinterne Informationen und Dokumente nicht in den Chatbot eingegeben werden dürfen. Auch urheberschutzrechtliche Fragen sind bei Eingaben in das System zu bedenken und sollten vor einer betriebsinternen Nutzung geklärt sein. Wenn von Chatbots generierte Inhalte genutzt werden, müssen diese grundsätzlich immer kenntlich gemacht werden. In diesem Zusammenhang kann es hilfreich sein, die Dialoge abzuspeichern und verfügbar zu halten, um die Nachvollziehbarkeit zu gewährleisten.

Datenschutzrechtliche Bedenken oder das Verarbeiten kritischer Informationen können womöglich dazu beitragen, dass die betriebliche Nutzung von Chatbots verboten oder eingeschränkt wird.

### **Eingabe sensibler Daten**





"Diese E-Mail mit der Personalkalkulation des nächsten Jahres hat mir mein Kollege Stefan geschickt. Wie kann ich gut darauf antworten?"



"Was muss ich beachten, wenn ich eine E-Mail zur Personalplanung schreiben möchte?"

Die erste Anfrage beinhaltet Klarnamen (Stefan) sowie zuordenbare Daten von Personen und betriebsinterne Informationen (Personalkalkulation).

### Besonderheiten

Beim Zugriff auf frei verfügbare LLM, wie ChatGPT oder Gemini, liegt keine vorherige Begrenzung der Inhalte auf Arbeitsschutzthemen oder dergleichen vor, was sowohl die Passgenauigkeit als auch die Qualität der Antworten einschränken kann. Eine zentrale Aufgabe bleibt somit auch bei der Nutzung von Chatbots im Arbeitsschutz die Qualitätskontrolle der Antworten.

Zentrale Aufgabe bei der Nutzung von Chatbots im Arbeitsschutz ist die Kontrolle der Antwortqualität.

Handlungen zum betrieblichen Arbeitsschutz unterliegen gesetzlichen Vorschriften, die verschiedene Rechte und Pflichten insbesondere für den:die Arbeitgeber:in festschreiben. Doch auch die betrieblichen Fachkräfte für Sicherheit und Gesundheit bei der Arbeit sind durch das Regelwerk an verschiedene Aufgaben gebunden. Da diese umfassenden Rechte und Pflichten von allen Beteiligten nachweislich wahrgenommen und eingehalten werden müssen, sind diese häufig in einer bestimmten Form zu erfüllen (z.B. Dokumentationsverpflichtungen und Nachweise). Diese Anforderungen sind modernen Chatbots meist nicht bekannt, denn wie beschrieben basieren die Antworten auf der Wahrscheinlichkeit von Wortfolgen und nicht auf sachlicher Korrektheit. Während Formulierungsvorschläge der Chatbots oft klar verständlich sind, können insbesondere bei rechtlichen Belangen Kleinigkeiten in der Formulierung einen großen Unterschied in der Bedeutung machen.

Sprachmodelle sind womöglich in der Lage, Inhalte und Anforderungen aus dem Rechts- und Regelwerk wiederzugeben oder gar anwender:innenfreundlich zu erklären, die Zusammenhänge zwischen verschiedenen Regelungsbereichen (z. B. Gesetze auf Bundes- und Landesebene) sowie die zugrundeliegende Systematik können jedoch nicht vollumfänglich abgebildet werden.

Chatbots können den Zugang zum Rechts- und Regelwerk erleichtern, ersetzen jedoch keine Fachkundige Person.

Nachfolgend sind einige Leitfragen angeführt, die auf zentrale Aspekte bei der Interaktion mit Chatbots eingehen:

#### Expertise / Vorwissen

- Welches Vorwissen haben Sie selbst in dem Themenbereich der Frage an den Chatbot?
- Wären Sie grundsätzlich in der Lage, einen fachlichen Fehler in der Antwort des Chatbots auszumachen?

#### Datensouveränität

- Wie sensibel sind die Informationen, die Sie in den Chatbot eingeben?
- Wären Sie bereit, mit einer fremden Person das gleiche Gespräch zu führen und dieselben Informationen zu teilen?

#### Konsequenzen

Welche Konsequenzen hat es, wenn Sie sich auf die Ausgabe des Chatbots verlassen und diese inkorrekt ist?

Sowohl beim Einholen von Informationen als auch bei der Entwicklung von konkreten Maßnahmen für den Arbeits- und Gesundheitsschutz ist darauf zu achten, welche Betroffenheiten vorliegen. Anfragen und Antworten sollten mit der entsprechenden Sorgfalt formuliert und interpretiert werden. Interventionen, Datenerhebungsmethoden oder Präventionsangebote, die im Dialog mit der KI entwickelt werden, sollten nicht ohne Prüfung übernommen werden und erfordern – wenn die Inhalte über das Wissen der Nutzenden hinausgehen – die Rücksprache mit Experten:Expertinnen.

Beispiel: Wer trägt die Verantwortung? Eine Sicherheitsfachkraft nutzt einen Chatbot für die Erstellung von Schulungsunterlagen zur richtigen Handhabung einer Leiter. Der Chatbot gibt falsche Informationen zum Anstellwinkel aus. Kurz danach stürzt ein:e Arbeitnehmer:in wegen einer zu flach angelegten Leiter ab und verletzt sich. Sind Sie sich bewusst, dass die Sicherheitsfachkraft hier die Verantwortung trägt?

Arbeitsplatzbezogene Risiken oder auch Dialoge rund um das Thema Gesundheit im Betrieb sollten möglichst neutral und ohne Personenbezug eingegeben werden. In ähnlicher Weise gilt es zu berücksichtigen, dass Forschende zwar Potenzial darin sehen, LLM auch für diagnostische und therapeutische Zwecke einzusetzen, im privaten Gebrauch jedoch unbedingt davon abgesehen werden sollte, medizinische Informationen (eigene oder bei anderen beobachtete) einzugeben und Therapievorschläge ungeprüft und ohne ärztliche Rücksprache anzunehmen.

#### **Chatbots im Arbeitsschutz:**

zwischen Hilfe und Halbwissen

#### **Vom Neuling zum Profi:**

#### **Wie Expertise die Chatbot-Nutzung beeinflusst**

- Hohe Expertise = mehr Spielraum: Kritisches Prüfen & gezieltes Nachfragen sind möglich.
- Laien:Laiinnen profitieren von allgemeinen Infos, aber komplexe Fachfragen erfordern das Wissen von Experten:Expertinnen.

#### **Fakten und Feinschliff:**

#### Wann ein Chatbot sinnvoll ist

- Einfache Fakten lassen sich meist gut abrufen aber immer gegenprüfen!
- Textoptimierung (z. B. Stil & Korrekturen) funktioniert oft gut, dennoch Sinnzusammenhang prüfen!

#### **Fehler mit Folgen:**

#### Abwägen der Risiken bei Chatbot-Nutzung

- Welche Konsequenzen wären mit der Nutzung von Fehlinformationen verbunden?
- Bei geringem Risiko können Chatbots als Hilfe geeignet sein.
- Bei hohem Risiko sollten unbedingt Fachleute konsultiert werden.

Ein weiterer Aspekt, der bei der Nutzung von LLM im Arbeitsschutz berücksichtigt werden sollte, bezieht sich auf das Vertrauen, welches den betrieblichen Experten:Expertinnen entgegengebracht wird. Die Verbreitung von fehlerhaften Informationen kann nicht nur mit Vertrauensverlust, sondern ggf. auch mit schädigenden Folgen für die Beschäftigten verbunden sein, sodass eine besonders hohe Sorgfaltspflicht besteht und der Einsatz von Chatbots bei Fragen zu Sicherheit und Gesundheit nicht den betrieblichen Dialog ersetzen darf.

### Kurz und bündig

Die Forschung zum Einsatz von Chatbots im Bereich von Sicherheit und Gesundheit bei der Arbeit steckt derzeit noch in den Kinderschuhen. Betriebliche Regeln zum Umgang mit Chatbots helfen dabei, geeignete Rahmenbedingungen zu definieren, denn insbesondere Aspekte wie Datenschutz oder auch die Verwendung von Ergebnissen und Inhalten, die mit Chatbots erzeugt worden sind, sollten geklärt sein. Die Beantwortung komplexer Arbeitsschutzfragen erfordert ein hohes Maß an Expertise und Fachwissen. Vor diesem Hintergrund sollten bei Dialogen mit Chatbots wichtige Aspekte wie die eigene Expertise, mögliche (ungewollte) Konsequenzen für unterschiedlicher Personengruppen berücksichtigt werden.

### Ähnlich und doch anders

Was in Österreich das ArbeitnehmerInnenschutzgesetz und die Arbeitsplatzevaluierung sind, sind in Deutschland das Arbeitsschutzgesetz und die Gefährdungsbeurteilung. Sprachliche und inhaltliche Unterschiede sind zu berücksichtigen, jedoch werden diese von LLM z. T. übergeneralisiert. Das bedeutet, dass Inhalte vermischt werden und im Internet (in den Trainingsdaten) häufig vorkommende Informationen das Antwortverhalten dominieren.



Die nachfolgende Klassifikation von Anwendungsfällen soll als Unterstützung dabei dienen, die Interaktionen mit Chatbots besser einordnen zu können. Sie erhebt jedoch keinen Anspruch auf Vollständigkeit und weitere Klassifikationen sind möglich. Dennoch bieten die folgenden Themenbereiche und Beispiele einen Zugang dazu, wie Interaktionen mit Chatbots gestaltet werden können und welche Fallstricke zu berücksichtigen sind.

### Themen erkunden/Brainstorming/Informieren

Das intuitive Dialogformat von Chatbots kann eine Unterstützung sein, um sich einem neuen Thema anzunähern und ohne Sorge vor Bewertung Fragen zu stellen. In kurzen Dialogen können Themenfelder erkundet, Informationen gesammelt, erstes Wissen erlangt und Fragestellungen aus anderen Perspektiven betrachtet werden. Gleichwohl eignen sich die offenen Dialogformate auch, um zu einem bekannten Themen-

komplex Ideen bzw. Argumentationslinien zu entwickeln. Das folgende Beispiel auf Seite 13 aus dem Bereich der Arbeitsmedizin bzw. der Interaktion mit Betriebsverantwortlichen zeigt, dass der Chatbot zunächst schlüssig und logisch argumentiert. Bei genauerer Betrachtung fehlen jedoch zentrale Argumente, die aus der Perspektive des Arbeits- und Gesundheitsschutzes relevant sind.



Ich bin Arbeitsmediziner einer Baufirma. Mit welchen Argumenten kann ich den Arbeitgeber von Maßnahmen zum Hautschutz der Beschäftigten auf der Baustelle überzeugen?

Als Arbeitsmediziner können Sie den Arbeitgeber von Maßnahmen zum Hautschutz der Bauarbeiter mit den folgenden Argumenten überzeugen:



- 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

Γ 1

3. Rechtliche Anforderungen und Haftung

[...]

4. Finanzielle Vorteile

[...]

5. Betriebsklima und Mitarbeiterbindung

[...]

6. Verbesserung des Unternehmensimages

[...]

Gesundheit und Wohlbefinden der Beschäftigten wurden als zentrales Argument erkannt. Obwohl die Antwort vollständig wirkt, fehlen jedoch wichtige Aspekte wie z.B. Hautkrebsrisiken durch Sonneneinstrahlung und Handhabung von Stäuben und Stoffen.

Abbildung 2: Fallbeispiel "Brainstorming" – Chatprotokoll (gekürzt) vom 24. Juni 2024 mit ChatGPT

Zusammenfassend lässt sich für Interaktionen im Bereich des Brainstormings feststellen, dass Chatbots die Perspektivenvielfalt oft erhöhen (siehe Tabellen 1 und 2). Dennoch kann es zu einer Überbetonung bestimmter Aspekte kommen, insbesondere wenn die Fragestellung eine Suggestion (z. B. einen bestimmten Begriff, der beispielhaft angeführt ist) enthält und dadurch die Antwort in eine bestimmte Richtung lenkt.

Das Erkennen unpraktikabler, fehlerhafter oder fehlender Inhalte erfordert weiterhin das Kontextwissen der Anwender:innen. Der von Chatbots vermittelte Eindruck der Vollständigkeit erschwert allerdings das Erkennen fehlender Inhalte. Zudem sollte nicht ver

nachlässigt werden, dass der kreative Austausch unter Kollegen:Kolleginnen ebenso fruchtbar sein kann.

### Haben Sie schon Rollenspiele mit einem Chatbot ausprobiert?

Nutzen Sie die Möglichkeit, mit einem Chatbot in den Dialog zu treten. Weisen Sie dem Chatbot unterschiedliche Rollen zu, z. B.: "Nimm die Rolle eines Sicherheitsbeauftragten in einer kleinen Tischlerei (mit 5 Mitarbeitenden) ein. Führe mit mir ein Fachgespräch über das Thema …"

Anwendungsfall	Brainstorming	
Beschreibung	Generierung kreativer Ideen und Ansätze in verschiedenen Kontexten, um innovative Lösungen für Probleme zu entwickeln oder neue Perspektiven zu eröffnen.	
Beispiele	<ul> <li>"In der Firma soll eine neue Standbohrmaschine angeschafft werden. Woran muss ich denken?"</li> <li>"Welche kreativen Lösungen könnten helfen, die Unfallrate in einem Lagerhaus zu senken?"</li> <li>"Wie kann ich den:die Arbeitgeber:in davon überzeugen, Hautschutzmaßnahmen auf der Baustelle anzubieten."</li> </ul>	
Vorteile / Chancen	<ul><li>Bietet vielfältige Perspektiven.</li><li>Fördert kreatives Denken.</li><li>Spart Zeit bei der Ideenfindung.</li></ul>	
Nachteile / Gefahren	<ul> <li>Risiko unpraktikabler Ideen.</li> <li>Mögliche Vernachlässigung eigener Kreativität.</li> <li>Übermäßiges Vertrauen in maschinell generierte Vorschläge.</li> <li>"Versteckte" Falschinformation bzw. fehlende Informationen.</li> </ul>	
Tipps	<ul> <li>Variieren Sie die Prompts, um unterschiedliche Ideen und Perspektiven zu erhalten.</li> <li>Bewerten Sie die generierten Ideen kritisch und prüfen Sie deren Umsetzbarkeit.</li> <li>Kombinieren Sie maschinelle Vorschläge mit menschlichem Input, um die Kreativität und Relevanz zu erhöhen.</li> <li>Integrieren Sie zusätzliche Informationen zur Problemstellung (wer ist involviert, warum soll das Thema behandelt werden usw.). Besonders effektiv lassen sich diese Informationen in ein Rollenspiel mit dem Chatbot integrieren.</li> </ul>	

Tabelle 1: Hinweise zum Umgang mit Chatbots beim Brainstorming

Anwendungsfall	Themen erkunden
Beschreibung	Unterstützung bei der Erkundung und Analyse neuer Themenbereiche, einschließlich der Bereitstellung relevanter Informationen und der Identifizierung von Schlüsselkonzepten.
Beispiele	<ul> <li>"Erkläre mir bitte die wesentlichen Prinzipien des Arbeitsschutzes in der Industrie."</li> <li>"Was muss ich bei der Einrichtung eines ergonomischen Heimarbeitsplatzes beachten?"</li> <li>"Was ist eine Gefährdungsbeurteilung / Arbeitsplatzevaluierung psychischer Belastungen?"</li> </ul>
Vorteile / Chancen	<ul> <li>Schnelles Eintauchen in neue Themen.</li> <li>Zugriff auf umfassende Informationen.</li> <li>Effiziente Strukturierung komplexer Inhalte.</li> <li>Das Vokabular eines Themengebiets kennenlernen.</li> </ul>
Nachteile / Gefahren	<ul> <li>Risiko von Fehlinformationen oder Halluzinationen.</li> <li>Oberflächliche Abdeckung komplexer Themen.</li> <li>Übermäßige Abhängigkeit von generierten Inhalten.</li> </ul>
Tipps	<ul> <li>Stellen Sie gezielte Nachfragen, um die Genauigkeit der Informationen zu verbessern.</li> <li>Verwenden Sie spezifische und präzise Prompts, um detaillierte und relevante Antworten zu erhalten.</li> <li>Validieren Sie die erhaltenen Informationen mit externen und zuverlässigen Quellen, um Fehlinformationen zu vermeiden.</li> </ul>

 Tabelle 2: Hinweise zum Umgang mit Chatbots beim Erkunden neuer Themen

### Inhalte und Informationen bearbeiten

Mit entsprechender Vorbereitung können Chatbots zudem dazu genutzt werden, Inhalte in ein neues Format zu bringen, beispielsweise mit der Aufforderung, Zusammenfassungen von Texten zu erstellen oder auch

Inhalte in Quizform aufzubereiten. Das nachfolgende Beispiel verdeutlicht, wie Inhalte nach einer Unterweisung abgefragt werden können. Der Chatbot greift in seinen Fragen den inhaltlichen Kontext auf.



Bitte erstelle mir ein kurzes Quiz von 10 Fragen, anhand dessen ich das Wissen zum Thema PSA (speziell Sicherheitsschuhe) in einer Restaurantküche abfragen könnte.

Natürlich, hier sind zehn Fragen zum Thema PSA (Persönliche Schutzausrüstung), speziell Sicherheitsschuhe in einer Restaurantküche:



#### Frage 1: Warum sind Sicherheitsschuhe in einer Restaurantküche wichtig?

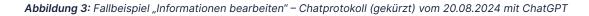
- a) Um den Mitarbeitern Komfort zu bieten
- b) Um das Risiko von Verletzungen durch rutschige Böden oder fallende Gegenstände zu minimieren
- c) Um einen professionellen Look zu gewährleisten
- d) Um die Mitarbeiter vor Sonnenbrand zu schützen

[...]

### Frage 3: Welche Sicherheitsmerkmale sollten Sicherheitsschuhe für eine Restaurantküche haben?

- a) Wasserdicht und atmungsaktiv
- b) Stahlkappe und rutschfeste Sohle
- c) Integrierter Flaschenöffner
- d) Eingebaute Uhr

Die Fragen sind zwar inhaltlich relevant und praxisnah, allerdings mangelt es den Antwortoptionen an Seriosität und Tiefe. Auch wird der Kontextbezug (hier: Gastronomie) nicht hinreichend aufgegriffen.



Insbesondere der Blick auf die Antwortoptionen zeigt eine mangelnde Seriosität und Tiefe in der Umsetzung der Aufgabe. Auch der Kontextbezug der Anfrage wird in den Antworten nicht hergestellt. Gezieltes Prompting (z. B. Vorgaben dazu, wie viele Antwortoptionen pro Frage generiert werden, Einfachantwort oder Mehrfachantwort, angenommener Kenntnisstand bei der Zielgruppe) führt zu einer präziseren Antwort des Sprachmodells.

Werden dem Sprachmodell Texte und Inhalte zur Verfügung gestellt, die umgeformt, zusammengefasst oder beantwortet werden sollen, trägt dies zur Antwortqualität bei – vorausgesetzt die bereitgestellten Inhalte und Informationen sind korrekt. Dennoch lassen sich das Risiko inhaltlicher Fehler wie auch mögliche Verzerrungen von Originalinhalten nicht ausschließen (siehe Tabelle 3).

Anwendungsfall	Inhalte und Informationen bearbeiten	
Beschreibung	Bearbeitung und Anpassung von Texten oder Daten, um sie für spezifische Zwecke zu optimieren, einschließlich der Umstrukturierung, Zusammenfassung oder Detaillierung von Inhalten.	
Beispiele	<ul> <li>"Fasse mir bitte diesen Artikel über Künstliche Intelligenz in 200 Wörtern zusammen."</li> <li>"Formatiere diese Richtlinien zur Notfallvorsorge nach den folgenden Vorgaben: Übersicht, Maßnahmen, Kontaktinformationen."</li> <li>"Überführe die Inhalte der Unterweisung in ein Quiz zur Leistungskontrolle."</li> </ul>	
Vorteile / Chancen	<ul><li>Schnelle Anpassung an spezifische Bedürfnisse.</li><li>Automatisierte Strukturierung und Formatierung.</li><li>Verbesserung der Textqualität.</li></ul>	
Nachteile / Gefahren	<ul> <li>Risiko inhaltlicher Fehler.</li> <li>Verlust von Nuancen bei komplexen Anpassungen.</li> <li>Mögliche Verzerrung von Originalinhalten.</li> </ul>	
Tipps	<ul> <li>Geben Sie klare und detaillierte Anweisungen zur gewünschten Struktur und Form der Inhalte.</li> <li>Überprüfen Sie die bearbeiteten Inhalte auf Konsistenz und Genauigkeit, um inhaltliche Fehler zu vermeiden.</li> <li>Ergänzen Sie die maschinellen Anpassungen mit persönlichem Feinschliff, um die Qualität zu erhöhen.</li> </ul>	

Tabelle 3: Hinweise zum Umgang mit Chatbots bei der Bearbeitung von Informationen und Inhalten

Potenzial für die Anwendung im Arbeitsschutz wird zudem darin gesehen, dass Texte in mehrere Sprachen übersetzt werden können. Mithilfe von Prompts können auch Festlegungen z.B. zur Sprachkomplexität, der Länge der Sätze oder des Textumfangs insgesamt getroffen werden.

Anwendungsfall	Übersetzungen und Mehrsprachigkeit	
Beschreibung	Übersetzung von Texten in andere Sprachen und Unterstützung bei der Erstellung und Pflege mehrsprachiger Inhalte, um eine breitere Zielgruppe zu erreichen.	
Beispiele	<ul> <li>"Bitte übersetze mir diesen Text aus dem Englischen ins Deutsche."</li> <li>"Welche idiomatischen Ausdrücke werden in der spanischen Sprache für 'Guten Morgen' verwendet?"</li> <li>"Erstelle mir eine mehrsprachige Broschüre für ein internationales Publikum in den Sprachen Englisch, Deutsch und Italienisch."</li> </ul>	
Vorteile / Chancen	<ul><li>Schnelle und kosteneffiziente Übersetzungen.</li><li>Zugang zu einem globalen Publikum.</li><li>Unterstützung bei der Mehrsprachigkeit.</li></ul>	
Nachteile / Gefahren	<ul> <li>Ungenaue Übersetzungen können zu Missverständnissen führen.</li> <li>Kulturelle Nuancen können verloren gehen.</li> <li>Gefahr der eingeschränkten Überprüfbarkeit der Ergebnisse.</li> <li>Nicht alle Sprachen werden vom Chatbot gleich gut beherrscht.</li> <li>Eingeschränkte Übersetzbarkeit von juristischen Begriffen und Fachbegriffen.</li> </ul>	
Tipps	<ul> <li>Überprüfen Sie die Übersetzungen auf kulturelle und sprachliche Nuancen, um Missverständnisse zu vermeiden.</li> <li>Nutzen Sie spezifische Prompts für Fachübersetzungen, um die Genauigkeit in speziellen Bereichen zu gewährleisten.</li> </ul>	

Tabelle 4: Hinweise zum Umgang mit Chatbots bei Übersetzungen und Mehrsprachigkeit

Anwendungsfall	Sprachliche Optimierung und Korrekturlesen	
Beschreibung	Verbesserung der Sprachqualität und Korrektur von Rechtschreib-, Grammatik- und Stilfehlern, um Texte präziser und ansprechender zu gestalten, ohne Inhalt und Struktur der zugrundeliegenden Texte weitreichend zu verändern.	
Beispiele	<ul> <li>"Korrigiere die Rechtschreib- und Grammatikfehler in diesem Absatz."</li> <li>"Bitte überarbeite den Schreibstil dieses Textes, um ihn formeller zu gestalten."</li> <li>"Überprüfe diesen Leitfaden für Sicherheitsprozeduren auf Klarheit und Verständlichkeit."</li> </ul>	
Vorteile / Chancen	<ul><li>Verbesserte Textqualität.</li><li>Zeitersparnis bei der Korrektur.</li><li>Einheitlicher Stil in Dokumenten.</li></ul>	
Nachteile / Gefahren	<ul> <li>Verlust des persönlichen Stils und Generierung gleichförmiger Texte.</li> <li>Übermäßige Abhängigkeit von Korrekturen.</li> <li>Risiko automatischer Fehlkorrekturen.</li> </ul>	
Tipps	<ul> <li>Überprüfen Sie die Korrekturen sorgfältig, um den persönlichen Stil des Textes zu wahren.</li> <li>Verwenden Sie spezialisierte Prompts, um spezifische stilistische Präferenzen und Anforderungen zu berücksichtigen.</li> <li>Kombinieren Sie maschinelle Korrekturen mit menschlicher Überprüfung, um Fehler zu minimieren.</li> </ul>	

 Tabelle 5: Hinweise zum Umgang mit Chatbots bei sprachlichen Optimierungen und Korrekturlesen

### Komplexe Analysen und Faktenchecks

Mithilfe von Chatbots lassen sich auch komplexere Fragestellungen bearbeiten. Hier wird beispielsweise bei der Identifikation und Analyse von Mustern in großen Datenmengen sowie bei der Datenaggregation und der nachfolgenden Strategieentwicklung besonders großes Potenzial vermutet. In diesem Zusammenhang ist jedoch die Bedeutung von Kontextinformationen, Datenqualität und zuvor angelegte "Leitplanken" zu berücksichtigen. Wie können komplexe Fragestellungen im Bereich des Arbeits- und Gesundheitsschutzes

mit einem Chatbot bearbeitet werden? Im folgenden Beispiel, das das Chemikalienrecht streift, verweist das Sprachmodell zwar auf wichtige Aspekte, dennoch wird deutlich, dass das Rechts- und Regelwerk mit seinen vielschichtigen Verzahnungen, Ausnahmen und regionalen Besonderheiten eine besondere Herausforderung darstellt. Der hohe Abstraktionsgrad und die "Pauschalität" der Antwort müssen bearbeitet und die Anforderungen im weiteren Dialog konkretisiert werden.



Hi ChatGPT, ich habe einen Betrieb für Badezimmerarmaturen. Normalerweise stellen wir Armaturen aus Messing her. Jetzt wollen wir auch einige von ihnen verchromen. Worauf muss ich dabei achten?

Das Verchromen von Badezimmerarmaturen kann das Erscheinungsbild verbessern und die Haltbarkeit erhöhen. Hier sind einige wichtige Punkte, die Sie beachten sollten:



- 1. Vorbereitung der Armaturen
- 2. Beschichtungsverfahren [...]
- 3. Qualitätskontrolle [...]
- 4. Umwelt- und Arbeitsschutz [...]
- 5. Zulieferer und Qualität der Materialien [...]
- 6. Technologische Aspekte [...]
- 7. Rechtliches und Normen:
- Regulierungen: Informieren Sie sich über die geltenden Vorschriften und Normen für verchromte Produkte in Ihrer Region [...]

In einem ersten Überblick werden zentrale Punkte angerissen – auch Umweltund Arbeitsschutz. Der Hinweis auf regional geltende Vorschriften ist wichtig.



Welche Sicherheitsvorkehrungen muss ich beachten, wenn ich die Armaturen verchromen will?

- 1. Persönliche Schutzausrüstung [...]
- 2. Chemikalienhandhabung
- Sicherheitsdatenblätter (SDS):
   Machen Sie sich mit den Sicherheitsdatenblättern der verwendeten Chemikalien vertraut und halten Sie sie in der Nähe des Arbeitsbereichs bereit.
- Lagerung: Lagern Sie Chemikalien gemäß den Vorschriften sicher und getrennt, um chemische Reaktionen zu verhindern.
- Kennzeichnung: Stellen Sie sicher, dass alle Behälter korrekt gekennzeichnet sind, um Verwechslungen zu vermeiden

- 3. Arbeitsplatzsicherheit
- 4. Schulung und Unterweisung
- 5. Entsorgung und Umweltschutz
- 6. Maschinen- und Anlagensicherheit
- 7. Dokumentation und Kontrolle
- 2. Beschichtungsverfahren [...]
- 3. Qualitätskontrolle [...]
- 4. Umwelt- und Arbeitsschutz [...]
- 5. Zulieferer und Qualität der Materialien [...]
- 6. Technologische Aspekte [...]
- 7. [...]

Die Nachfrage zu erforderlichen Sicherheitsvorkehrungen konkretisiert das Anliegen. Die Bedeutsamkeit und rechtliche Verpflichtung bspw. des Zugangs zu SDS für Beschäftigte wird nicht hinreichend deutlich gemacht.



Die hier erhaltene Antwort wird dem Umfang der gesetzlichen Pflichten, die aus der Einführung eines neuen Verfahrens und neuer Arbeitsstoffe resultieren, nicht gerecht.

Ein möglicher Weg zur Überprüfung von Fakten – insbesondere von gesetzlichen Anforderungen – wird im folgenden Fallbeispiel dargestellt.

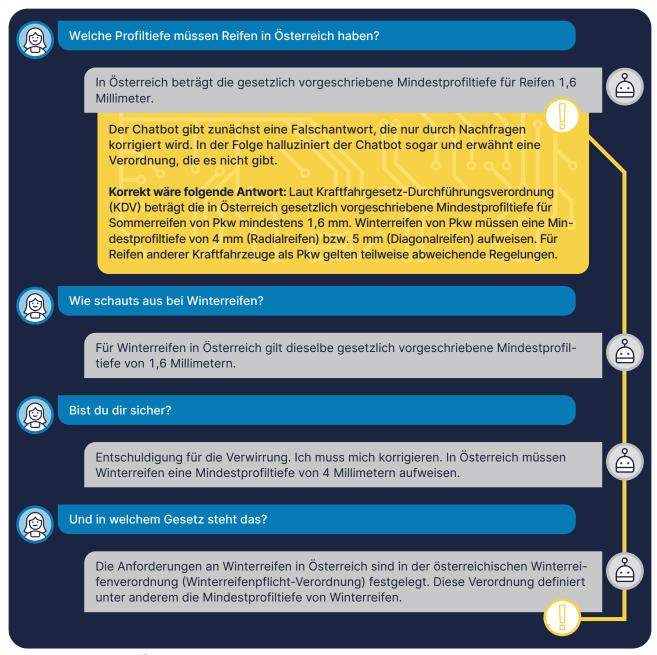


Abbildung 5: Fallbeispiel "Überprüfung gesetzlicher Anforderungen" - Chatprotokoll (gekürzt) vom 12.01.2025

Anwendungsfall	Recherche und Faktenprüfung
Beschreibung	Sorgfältige Überprüfung von Informationen, um deren Richtigkeit sicherzustellen, sowie Durchführung gezielter Recherchen zur Unterstützung von Projekten oder Entscheidungsprozessen.
Beispiele	<ul> <li>"Welche sind die neuesten Vorschriften zur Arbeitssicherheit im Bauwesen?"</li> <li>"Stimmen die folgenden Informationen über die Auswirkungen von Lärm auf die Gesundheit der Arbeitnehmenden mit aktuellen Studien überein?"</li> <li>"Gibt es aktuelle Berichte über Sicherheitsvorfälle in der Chemieindustrie?"</li> </ul>
Vorteile / Chancen	<ul><li>Unterstützt fundierte Entscheidungen.</li><li>Spart Zeit bei der Informationsbeschaffung.</li></ul>
Nachteile / Gefahren	<ul> <li>Risiko der Übernahme unzuverlässiger Quellen.</li> <li>Mögliche Bestätigung von Fehlinformationen.</li> <li>Gefahr, dass Fakten aus dem Kontext gerissen werden.</li> <li>Beachten Sie, dass nicht alle Chatbots in der Lage sind, tagesaktuelle Informationen zu berücksichtigen.</li> </ul>
Tipps	<ul> <li>Validieren Sie die Informationen mit mehreren Quellen, um die Zuverlässigkeit der Daten zu gewährleisten.</li> <li>Nutzen Sie präzise und spezifische Prompts, um genauere und relevantere Ergebnisse zu erzielen.</li> <li>Achten Sie auf die Quelle der bereitgestellten Informationen und prüfen Sie deren Glaubwürdigkeit.</li> <li>Informieren Sie sich über den Wissensstand, der vom Chatbot berücksichtigt wird (welche Trainingsdaten sind bis zu welchem Zeitpunkt eingeflossen und werden tagesaktuelle Daten genutzt).</li> </ul>

Tabelle 6: Hinweise zum Umgang mit Chatbots bei der Recherche und Faktenprüfung

Die Analyse komplexer Fragestellungen oder Datenbestände mithilfe von LLM kann durch ein nachtrainiertes Modell oder die Berücksichtigung zusätzlicher Informationen via <u>retrieval augmented generation</u> verbessert werden. Sowohl die Entwicklung als auch das Training dieser Modelle erfordern eine hohe Expertise, aber auch die spätere Anwendung sollte nur durch geschulte Fachkräfte erfolgen. Aufgrund ihrer technischen Eigenschaften sind Sprachmodelle in der Regel nur so gut,

wie das Wissen, an dem sie gelernt haben und mit dem sie trainiert worden sind, umfangreich und fundiert ist. Da verschiedene Themenbereiche unterschiedlich gut bekannt sind, kann sich das auf die Vollständigkeit und Genauigkeit der Antworten auswirken. Betriebliche Realitäten sind in der Regel hochkomplex und auch im Bereich des Arbeits- und Gesundheitsschutzes können selten alle Rand- und Rahmenbedingungen in einem kurzen Dialog abgebildet werden.

### Kurz und bündig

Die Anwendungsfälle zeigen die hohe Bandbreite an Einsatzmöglichkeiten von Chatbots auf und verdeutlichen die jeweiligen Vor- und Nachteile bei der Nutzung. In allen Dialogen ist zu berücksichtigen, dass Chatbots auf eine angenehme Kommunikation ausgelegt sind und "gefallen wollen". Dies zeigt sich auch, wenn sie als Informations- und Analysewerkzeug im Bereich von Sicherheit und Gesundheit bei der Arbeit genutzt werden. Daher ist der Umgang mit den Antworten des Chatbots auf zwei Ebenen besonders her

ausfordernd: Einerseits kommt es in manchen Fällen zu offensichtlichen Fehlern und Falschantworten. Diese können mit der entsprechenden Expertise identifiziert werden. Andererseits kann es bei den Antworten auch zu versteckten Fehlern kommen, etwa durch Auslassen, Unvollständigkeit oder Überbetonung bestimmter Aspekte. Diese können insbesondere durch suggestive Fragen der Anwender:innen entstehen und sind nur schwer aufzudecken.

### Das Wesentliche zum Abschluss

Der Einsatz von modernen Chatbots kann durch das Anbieten von Ideen und Informationen einen positiven Beitrag zur Effizienz und Effektivität von Arbeitssicherheit und Gesundheitsschutz leisten. 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 sowie Wissen über die Möglichkeiten und Grenzen der Anwendungen sind daher eine wichtige Grundvoraussetzung bei der Nutzung.

Hier ist das Wesentliche nochmals kurz zusammengefasst:

#### Betriebliche Vorgaben

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

- Nähere Informationen in Kapitel "Betriebliche Voraussetzungen"
- 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!

- → Nähere Informationen in den Kapiteln "Welche ökologischen und datenschutzrechtlichen Bedenken gibt es?" und "Chatbots im Arbeits- und Gesundheitsschutz"
- 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!
  - → Nähere Informationen in Kapitel "Anwendungsfälle"

#### 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!

- → Nähere Informationen in Kapitel "Anwendungsfälle", insbesondere "Komplexe Analysen und Faktenchecks"
- 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!

→ Nähere Informationen in Kapitel "Anwendungsfälle"

# Anhang I: Glossar – Fachbegriffe kurz erklärt

### Basis Modell / engl.: Foundation Model und Nachtrainieren / Fine Tuning

Das Generieren von Sprache oder Erkennen von Bildern sind sehr komplexe Tätigkeiten. KI-Programme, die diese Aufgaben erfüllen, sind sehr groß und benötigen viele Daten, um trainiert zu werden. In einigen Fällen reichen die verfügbaren Daten nicht aus, um ein Modell für eine Aufgabe zu trainieren. Ein Beispiel: Es soll ein Chatbot erzeugt werden, der Standardanfragen für ein Unternehmen beantwortet. Der Umfang an gut beantwortbaren Anfragen ist anfangs aber gering. Hier wäre eine mögliche Vorgehensweise, ein großes Modell zu nehmen und es zunächst darauf zu trainieren, sinnvolle Sprache zu generieren. Dieses Modell ist dann das Basismodell: Es beherrscht die grundlegende Aufgabe "Sprache generieren". Durch eine Feinabstimmung - ein Fine Tuning oder Nachtrainieren anhand der Anfragen – kann das Modell in der Folge lernen, die spezifischen Anfragen des Unternehmens zu beantworten.

#### Chatbot

Ein Chatbot ist ein Computerprogramm, welches als Gesprächspartner:in in einem Chatverlauf mit dem:der Nutzenden interagiert. Chatbots lassen sich ohne KI als klassisches Computerprogramm gestalten. Diese traditionellen Chatbots sind jedoch wesentlich weniger leistungsfähig als die KI-basierten Versionen, welche ihre Antworten auf die Eingaben der Nutzenden mithilfe eines Large Language Models finden.

#### **Deep Learning**

Die künstlichen neuronalen Netze (KNN), die in der Praxis meist zum Einsatz kommen, bestehen aus Ebenen von Neuronen, die miteinander verbunden sind. Dabei wird die Ausgabe einer Ebene immer als Eingabe der nächsten verwendet (der Fachbegriff ist "Fully Connected Feed Forward Network"). Die Anzahl der Ebenen, die auf diese Weise verbunden sind wird als Tiefe des Netzwerks bezeichnet.

Die Bezeichnung Deep Learning leitet sich davon ab, dass sehr große, tiefe neuronale Netze mit vielen ineinander geschalteten Ebenen und vielen Neuronen verwendet werden. Zu sagen, dass Deep Learning das Trainieren von großen Netzen ist, wäre jedoch zu kurz gegriffen. Beim Deep Learning geht es um die Art und Weise, wie mit dem Problem, welches von den KNN gelöst werden soll, umgegangen wird. Bei der klassischen Entwicklung von KI wird sehr viel Zeit damit ver

bracht, das zu lösende Problem so gut wie möglich zu begreifen. So kann ein KNN gestaltet werden, das für die Lösung gut geeignet ist. Deep Learning verfolgt einen gänzlich anderen Ansatz. Hier geht es darum, viele Daten und leistungsfähige, große Netze zu verwenden. Man hofft, dass das Netz jene Zusammenhänge, die der:die Programmierende im klassischen Fall verstehen muss, um das optimale Netz zu finden, beim Training selbst entdeckt und entsprechend reagiert.

#### **Generative KI**

Es gibt verschiedene Aufgaben, die KI-Programme erfüllen können, etwa Vorhersagen tätigen oder Eingaben klassifizieren (Beispiel: "Handelt es sich bei dem handschriftlich geschriebenen Buchstaben um ein 'i' oder ein 'I'?"). Generative KI ist ein Sammelbegriff für Programme, die basierend auf einer Eingabe (z. B. "Erstelle ein Bild von einem blauen Schmetterling!") ein digitales Gut wie Texte, Bilder oder Videos erstellen.

### GPT-3, GPT-4, LLAMA, Mixtral, Bard, Gemini, Copilot, DeepSeek

Bei den genannten Begriffen handelt es sich um Chatbots oder LLM verschiedener Hersteller:innen. Sie werden oft als Synonym für Chatbots verwendet. So ist GPT-4 z. B. das LLM, welches ChatGPT 4 von OpenAl zugrunde liegt. LLAMA ist das LLM von Meta, Mixtral das offene LLM von MistralAl, Copilot ist der Chatbot von Microsoft usw.

### Großes Sprachmodell / engl.: Large Language Model (LLM)

Ein großes Sprachmodell ist ein Computerprogramm, das dazu geschrieben wurde, einen Text einzulesen und anschließend jenes Wort an den Text anzufügen, welches ihn logisch weiterführt. Große Sprachmodelle verfügen über sehr viele Parameter (daher der Zusatz "groß" im Namen).

#### Halluzination (von großen Sprachmodellen)

Chatbots können je nach Eingabe des:der Nutzenden Ausgaben erzeugen, die logisch wirken, jedoch frei erfunden und nicht auf die Trainingsdaten zurückzuführen sind. So kann z.B. auf einen Text oder eine Quelle verwiesen werden, die es nicht gibt, oder ein historischer Sachverhalt wird nicht der Realität entsprechend dargestellt. Diese Falschausgaben nennt man Halluzinationen.

### Künstliche Intelligenz (KI) / engl.: Artificial Intelligence (AI)

Aufgrund der Bandbreite und Relativität des Begriffs Intelligenz ist es nicht möglich, eine allgemeingültige KI-Definition zu formulieren. Unter Künstlicher Intelligenz kann man sich in hiesigem Kontext ein Computerprogramm vorstellen, welches basierend auf Daten, die es verarbeitet, zu seiner konkreten Funktionsweise kommt. Bei einem klassischen Computerprogramm muss der:die Programmierende jedes mögliche Verhalten des Programms im Vorhinein spezifizieren. Bei KI kann das Programm in einem gewissen Spielraum, der von dem:der Programmierenden vorgegeben wird, und basierend auf spezifischen Daten (u. a. Beispiele für erfolgreiches Handeln) sein Verhalten selbstständig definieren.

### Künstliches Neuronales Netz (KNN) / engl.: Artificial Neural Network (ANN)

Ein Künstliches Neuronales Netz (KNN) ist ein Computerprogramm, welches modellhaft dem Nervensystem eines natürlichen Organismus nachempfunden ist. Eine Verkettung von mathematischen Funktionen simuliert das Verhalten von Nervenzellen, die miteinander verbunden sind. Wie sein biologisches Vorbild kann das KNN durch das Anlegen neuer "Nervenverbindungen" darauf trainiert werden, verschiedene Aufgaben zu erfüllen.

#### **Parameter**

Aus technischer Sicht handelt es sich bei einem KI-Programm um eine sehr große mathematische Funktion, die durch Verarbeitung von Daten so verändert wird, dass sie das Problem bzw. die Aufgabe löst, für welche das Programm geschaffen wurde. Ein Beispiel für eine Familie solcher Funktionen sind die Künstlichen Neuronalen Netze. Die Parameter sind verschiedene Zahlenwerte in dieser Funktion, die während des Trainings angepasst werden, um den gewünschten Effekt zu erzielen. Bei einem Künstlichen Neuronalen Netz sind die Parameter Zahlen, die bestimmen, wie stark die künstlichen Neuronen miteinander verbunden sind. Sie sind also das Gegenstück von Synapsen in biologischen Nervensystemen. Meist ist die Anzahl der Parameter eine konstante Eigenschaft des verwendeten Programms und wird während des Trainings nicht verändert.

#### **Prompt / Prompten / Prompt Engineering**

Als Prompt bezeichnet man die Eingabe des:der Nutzenden in einen Chatbot. Die Interaktion mit dem Chatbot wird auch als Prompten bezeichnet. Unter dem Begriff Prompt Engineering versteht man das gezielte Formulieren von Anfragen an den Chatbot, die das technische Wissen über die Funktionsweise des Chatbots mitberücksichtigt, um bessere Antworten zu erhalten. So ist z. B. bekannt, dass Chatbots bei komplizierten Herleitungen von Sachzusammenhängen weniger oft falsch liegen, wenn man sie bittet, die Herleitung schrittweise zu erklären. Diese Forderung nach einer Erklärung mit dem Ziel, eine bessere finale Aussage zu erhalten, wäre hier Prompt Engineering.

#### **Retrieval Augmented Generation (RAG)**

Beim RAG werden im Vorfeld Texte zu einer Datenbank verarbeitet, in der sich Textstücke, deren Inhalt ähnlich ist, gemeinsam auffinden lassen<sup>2</sup>. Das LLM arbeitet dann nicht direkt mittels Eingabe des:der Nutzenden. Stattdessen werden Textstücke aus der Datenbank, die zu der Eingabe passen, gefunden. Diese Textstücke werden zusammen mit der Eingabe zu einem Prompt zusammengestellt, der die Information enthält, welcher Teil Eingabe und welcher Datenbankwissen ist. Das LLM arbeitet dann mit diesem Prompt und kann so zusätzliches Fachwissen berücksichtigen, das nicht im Training enthalten war. Im komponierten Prompt kann es auch dazu aufgefordert werden, aus den Textstücken der Datenbank zu zitieren. Dies ist eine Möglichkeit, tatsächlich existierende Quellen in eine Chatbotausgabe einzufügen.

#### **Training von KI**

KI-Programme stellen die Regeln, wie sie auf bestimmte Eingaben reagieren, durch das Verarbeiten von sogenannten Trainingsdaten selbst auf. Dieses Verarbeiten der Daten und Ableiten von Regeln wird als Training bezeichnet. Die meisten Programme, die im Moment praktische Anwendung finden, werden anfänglich über einen gewissen Zeitraum hinweg trainiert. Nach dieser sogenannten Trainingsphase ist die Art, wie das Programm einen bestimmten Input verarbeitet, festgelegt. So lernt ein Programm, welches Verkehrsschilder erkennt, anfangs an 1.000 Beispielen, wie ein Stoppschild aussieht. Es lernt danach allerdings im Einsatz nicht weiter – die Fähigkeit, "im Feld" zu lernen, ist Teil aktueller Forschung.

<sup>2</sup> Für den Aufbau so einer Datenbank gibt es verschiedene Technologien aus dem Bereich des Natural Language Processing. Die Stichworte hier sind "Word Embedding" oder "Sequence Embedding".

### **Anhang II: Literatur**

Al-Mhdawi, M. K. S.; Qazi, A.; Alzarrad, A.; Dacre, N.; Rahimian, F.; Buniya, M. K.; Zhang, H. (2023). *Expert Evaluation of ChatGPT Performance for Risk Management Process Based on ISO 31000 Standard*. Available at SSRN 4504409.

Arici, N.; Gerevini, A. E.; 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 2023, 15(8), 268. https://doi.org/10.3390/fi15080268.

dos Santos, M. L.; Victória, V. N. G. (2024). Critical evaluation of applications of artificial intelligence based linguistic models in Occupational Health. Rev Bras Med Trab. 2024 Aug 5; 2, 6.

Flanagan, F.; Walker, M. (2021). How can unions use Artificial Intelligence to build power? The use of AI chatbots for labour organising in the US and Australia. New Technology, Work and Employment 36(2): 159–176. https://doi.org/10.1111/ntwe.12178.

Kereopa-Yorke, B. (2024). Building resilient SMEs: Harnessing Large Language Models for Cyber Security in Australia. Journal of AI, Robotics & Workplace Automation, 3(1), 15–27.

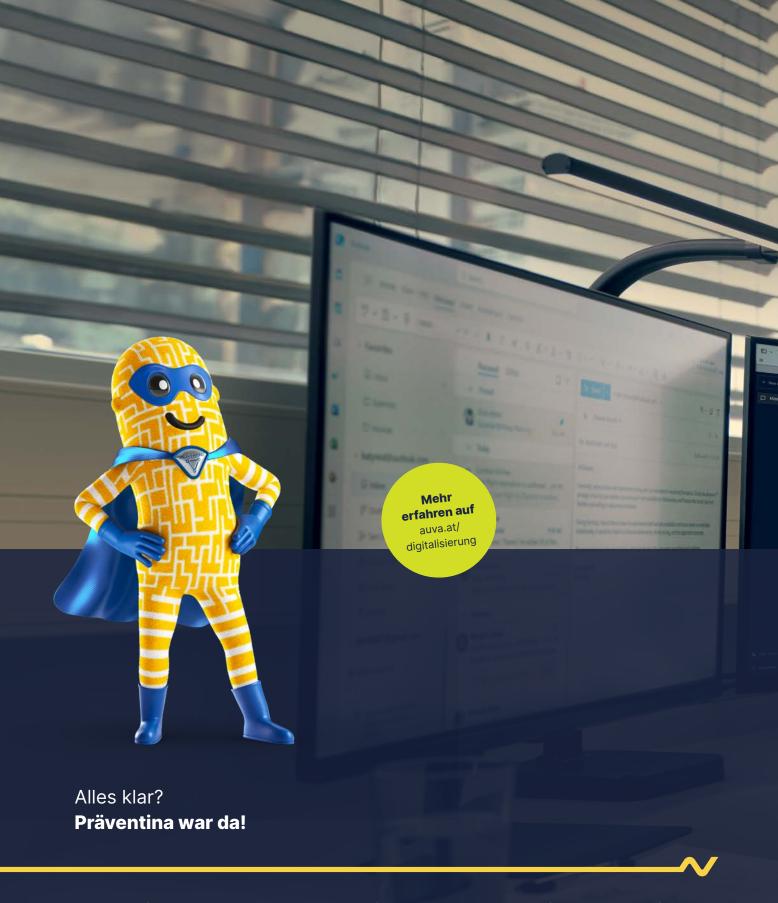
Luccioni, A. S.; Viguier, S.; Ligozat, A. L. (2023). *Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model.* Journal of Machine Learning Research, 24(253), 1–15.

Ooi, K. B.; Tan, G. W. H.; Al-Emran, M.; Al-Sharafi, M. A.; Capatina, A.; Chakraborty, A.; Wong, L. W. et al. (2023). *The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions.* Journal of Computer Information Systems, 1–32. https://doi.org/10.1080/08874417.2023.2261010.

Padovan, M.; Cosci, B.; Petillo, A.; Nerli, G.; Porciatti, F.; Scarinci, S.; Palla, A. et al. (2024). ChatGPT in Occupational Medicine: A Comparative Study with Human Experts. Bioengineering, 11(1), 57.

Smetana, M.; Salles de Salles, L.; Sukharev, I.; Khazanovich, L. (2024). *Highway Construction Safety Analysis Using Large Language Models*. Appl. Sci. 2024, 14(4), 1352. https://doi.org/10.3390/app14041352.

Uddin, S. J.; Albert, A.; Ovid, A.; Alsharef, A. (2023). Leveraging ChatGPT to Aid Construction Hazard Recognition and Support Safety Education and Training. Sustainability 2023, 15(9), 7121.



Das barrierefreie PDF dieses Dokuments gemäß PDF/UA-Standard ist unter <u>auva.at/publikationen</u> abrufbar.

**Medieninhaber und Hersteller:** Allgemeine Unfallversicherungsanstalt, Wienerbergstraße 11, 1100 Wien **Verlags- und Herstellungsort:** Wien

HUB - 08/2025 - tev

Titelbild: digitalwerk & Vienna Paint Studios

Layout: Lukas Hofreiter