

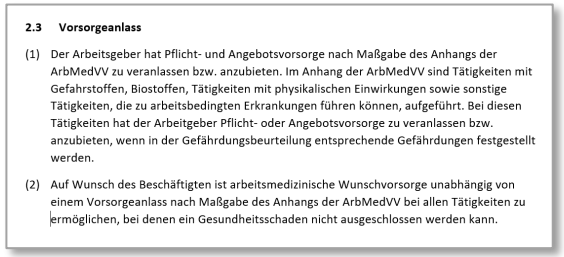
Proof of concept - Verarbeitung von Normtexten mit LLMs

CHATBOT - STAATLICHE ARBEITSSCHUTZ-REGELN

Testumgebung in Form eines Chatbots

- Entwicklung eines Chatbots als praxisnahe Evaluationsumgebung für Normtexte
- Bewertung der Leistungsfähigkeit entlang realistischer Use-Cases
- Dialogbasierte Interaktion ermöglicht Rückfragen oder Präzisierungswünsche
- Demo macht die Ergebnisse und Grenzen der Verarbeitung anschaulich
- Einbindung folgender staatlicher Regeln
  - *TRGS 402 „Ermitteln und Beurteilen der Gefährdungen bei Tätigkeiten mit Gefahrstoffen: Inhalative Exposition“ (2023)*
  - *AMR 3.3 „Ganzheitliche arbeitsmedizinische Vorsorge unter Berücksichtigung aller Arbeitsbedingungen und arbeitsbedingten Gefährdungen“ (2022)*
  - *ASR A6 „Bildschirmarbeit“ (2024)*
  - *TRBA 250 „Biologische Arbeitsstoffe im Gesundheitsdienst und in der Wohlfahrtspflege“ (2025)*
  - *TRBS 1201 „Prüfungen und Kontrollen von Arbeitsmitteln und überwachungsbedürftigen Anlagen“ (2019)*

Betrachtung von 4 zentralen Inhaltsarten



2.3 Vorsorgeanlass

(1) Der Arbeitsgeber hat Pflicht- und Angebotsvorsorge nach Maßgabe des Anhangs der ArbMedVV zu veranlassen bzw. anzubieten. Im Anhang der ArbMedVV sind Tätigkeiten mit Gefahrstoffen, Biostoffen, Tätigkeiten mit physikalischen Einwirkungen sowie sonstige Tätigkeiten, die zu arbeitsbedingten Erkrankungen führen können, aufgeführt. Bei diesen Tätigkeiten hat der Arbeitgeber Pflicht- oder Angebotsvorsorge zu veranlassen bzw. anzubieten, wenn in der Gefährdungsbeurteilung entsprechende Gefährdungen festgestellt werden.

(2) Auf Wunsch des Beschäftigten ist arbeitsmedizinische Wunschvorsorge unabhängig von einem Vorsorgeanlass nach Maßgabe des Anhangs der ArbMedVV bei allen Tätigkeiten zu ermöglichen, bei denen ein Gesundheitsschaden nicht ausgeschlossen werden kann.

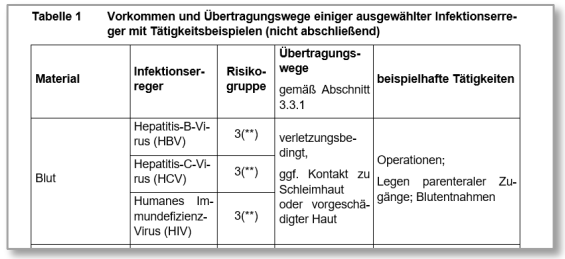
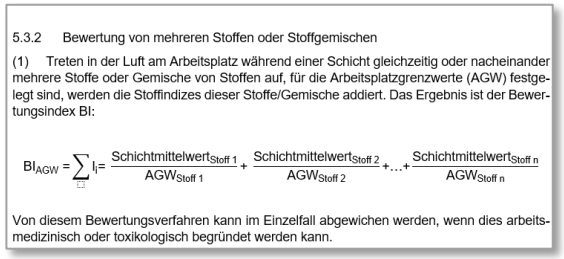


Tabelle 1 Vorkommen und Übertragungswege einiger ausgewählter Infektionserreger mit Tätigkeitsbeispielen (nicht abschließend)

Material	Infektionserreger	Risikogruppe	Übertragungswege gemäß Abschnitt 3.3.1	beispielhafte Tätigkeiten
Blut	Hepatitis-B-Virus (HBV)	3(**)	verletzungsbedingt, ggf. Kontakt zu Schleimhaut oder vorgeschädigter Haut	Operationen; Legen parenteraler Zugänge; Blutentnahmen
	Hepatitis-C-Virus (HCV)	3(**)		
	Humanes Immundefizienz-Virus (HIV)	3(**)		



5.3.2 Bewertung von mehreren Stoffen oder Stoffgemischen

(1) Treten in der Luft am Arbeitsplatz während einer Schicht gleichzeitig oder nacheinander mehrere Stoffe oder Gemische von Stoffen auf, für die Arbeitsplatzgrenzwerte (AGW) festgelegt sind, werden die Stoffindizes dieser Stoffe/Gemische addiert. Das Ergebnis ist der Bewertungsindex BI:

$$BI_{AGW} = \sum_{i} I_i = \frac{\text{Schichtmittelwert}_{\text{Stoff 1}}}{AGW_{\text{Stoff 1}}} + \frac{\text{Schichtmittelwert}_{\text{Stoff 2}}}{AGW_{\text{Stoff 2}}} + \ldots + \frac{\text{Schichtmittelwert}_{\text{Stoff n}}}{AGW_{\text{Stoff n}}}$$

Von diesem Bewertungsverfahren kann im Einzelfall abgewichen werden, wenn dies arbeitsmedizinisch oder toxikologisch begründet werden kann.

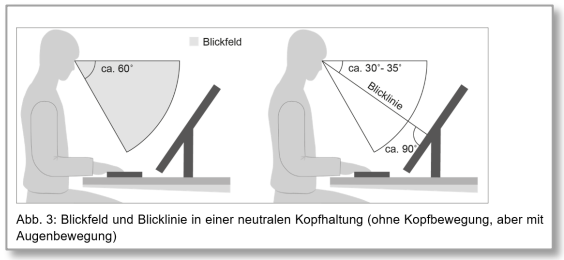


Abb. 3: Blickfeld und Blicklinie in einer neutralen Kopfhaltung (ohne Kopfbewegung, aber mit Augenbewegung)

1. Text
2. Tabellen (+ Tabellen aus dem Bereich Gartenbau)
3. Mathematische Formeln
4. Abbildungen

1. Text

- Auffinden relevanter Passagen und inhaltliche Beantwortung funktionieren in der Regel **zuverlässig**
- Textinhalte werden wortgenau übernommen und wiedergegeben
- Sehr viele Quellen und Bezüge!

Herausforderungen

- Normtexte verweisen teilweise auf **externe Quellen** (z. B. weitere Normen, Richtlinien, Anhänge), die **nicht** im hochgeladenen Dokumentenkorpus enthalten sind
- Ohne Zugriff auf diese **Referenzen** bleiben Antworten an solchen Stellen **unvollständig** oder müssen eingeschränkt werden
- **Versionierung** muss geklärt und den Bots beigebracht werden (Zeiträume vereinbaren)

Empfehlung:

- Sicherstellen, das referenzierte Quellen ebenfalls bereitgestellt und zugänglich sind
- Einfache Sprache ergänzen

2. Tabellen

- Die Verarbeitung einer Tabelle als zusammenhängender Block macht das gezielte Auffinden einzelner Einträge **nicht** zuverlässig
- Die Bedeutung eines Tabellenwerts ergibt sich aus seiner **Zuordnung zu Zeile und Spalte**, insbesondere über Zeilenbezeichnungen und Spaltenüberschriften

Implementierter Lösungsansatz:

- Zusätzliche **Verarbeitung auf Zeilenebene**: Jede Tabellenzeile wird separat erfasst und mit den Spalten-überschriften angereichert, sodass Einträge eigenständig interpretierbar sind

Empfehlung:

- Möglichst **einfache Tabellenstrukturen** verwenden; komplexe Header-Konstruktionen (mehrzeilig, verschachtelt, mehrdimensional) erschweren die robuste Verarbeitung und sollten vermieden werden

3. Mathematische Formeln

- Formeln sind im speziellen Office Math Markup Language (OMML) Format hinterlegt

- Ohne **explizite Aufbereitung** lassen sich Formeln nur eingeschränkt verarbeiten, die Darstellung und mathematische Bedeutung werden dabei **nicht zuverlässig** übernommen

Implementierter Lösungsansatz:

- Umwandlung der Formel in **LaTeX** als standardisierte, lesbare Repräsentation zur zuverlässigen Speicherung, Suche und Darstellung in Antworten

- Nutzung des **umliegenden Textkontextes** zur besseren Auffindbarkeit

Empfehlungen:

- Formeln möglichst konsistent in einer eindeutigen, textbasierten Notation wie LaTeX hinterlegen

- Formeln mit textueller Beschreibung **räumlich nah (!) annotieren**

4. Bilder

- Bilder sind **nicht** automatisch **in gleicher Qualität** nutzbar wie Fließtext, da der Inhalt häufig **nicht** direkt als Text vorliegt

- Unterschiedliche Arten von Bildern erfordern **unterschiedliche Vorgehensweisen**

4.1. Bild mit reinen Textinhalten



Abb. 2: Hierarchie von Gestaltungskriterien in der Softwareergonomie (in Anlehnung an DGUV Information 215-450)

Herausforderung:

- Inhalt ist visuell vorhanden, aber **nicht unmittelbar** verarbeit- und nutzbar
- Typisch: Abbildungen nur mit Textkästen → auflösen

Möglicher Lösungsansatz:

- Texterkennung (OCR) zur **Umwandlung in maschinenlesbaren Text**

Empfehlung:

- Textinformationen nach Möglichkeit **direkt als Text hinterlegen** und „Text-als-Bild“ vermeiden
- Wichtige Anforderungen an die zukünftige grundsätzliche Ausgestaltung von staatlichen Regeln → **Barrierefreiheit**
- Abbildungen vermeiden vor Abbildungen beschreiben

- **4.2. Schematische Abbildungen**

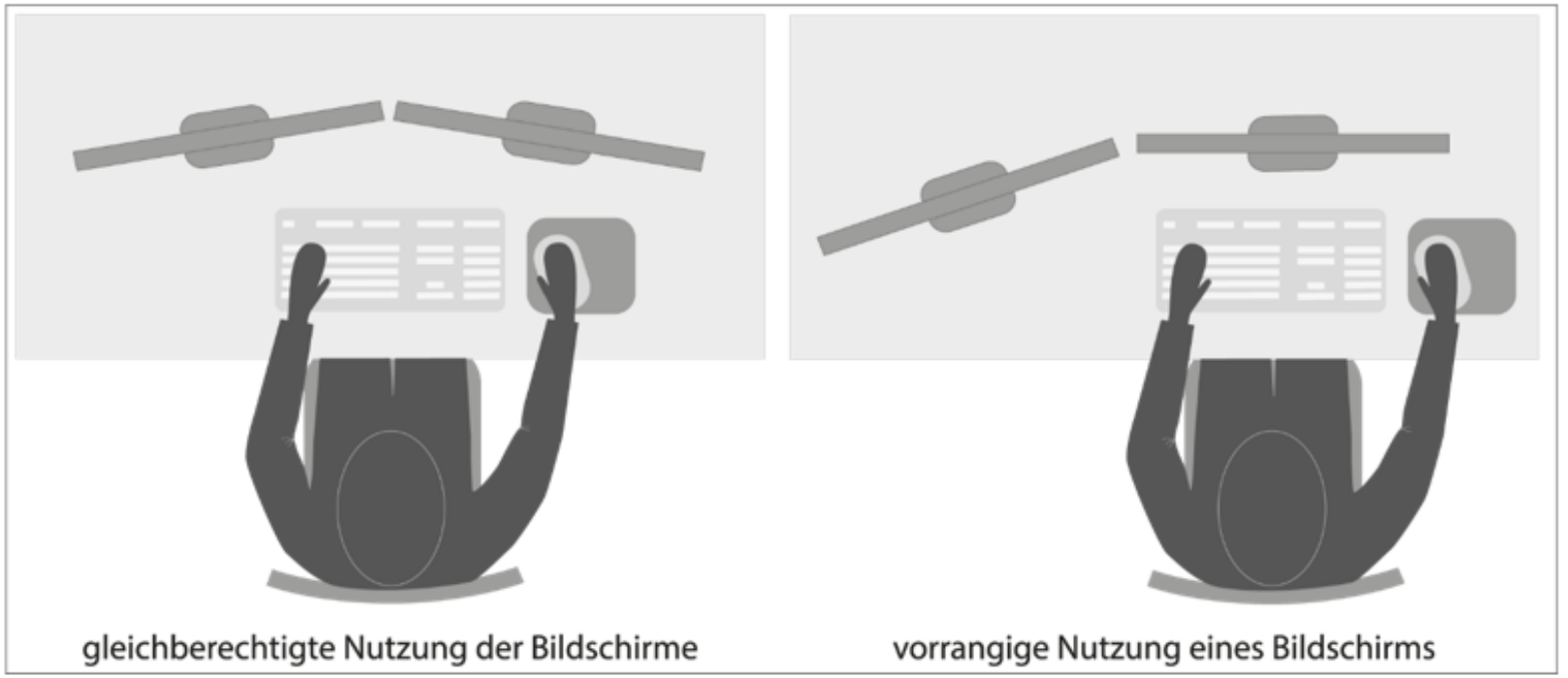


Abb. 4: Anordnung mehrerer Bildschirme

Herausforderung:

- Bedeutung entsteht aus dem **visuellen Zusammenhang**, nicht nur aus einzelnen Wörtern

Möglicher Lösungsansatz:

- **KI-basierte Bildanalyse** zur Erstellung einer kurzen, fachlichen Beschreibung als Grundlage für die Verarbeitung und Auffindbarkeit

Empfehlung:

- Schematische Abbildungen von vornherein **mit einer Beschreibung versehen** (auch relevant für Barrierefreiheit, z. B. für blinde Nutzer)

Fazit

- Formeln und beschreibende Formeltexte sollten **räumlich sehr nah** stehen
- Wie sieht es mit der **Versionierung** der Regeln aus von? Ist zu klären. Useranfrage
- ***docx*** funktioniert reibungslos, ***pdf*** mittels techn. Unterstützung in ***docx*** **umwandeln,** visuelle Kontrolle erforderlich
- Chatbot kann **kontextnahe Themen** gut aufgreifen
- Bei zu vielen Quellen, **Limitierung** oder festlegen des Scores (Qualität / Passfähigkeit der Quelle in Abh. zur Frage), z.B. 10 „beste“ Quellen
- Fließtext lässt sich zuverlässig verarbeiten, auffinden und korrekt wiedergeben
- Tabellen und Formeln benötigen **gezielte Aufbereitung**, damit Inhalte eindeutig nutzbar bleiben
- Bilder sind je nach Typ nur eingeschränkt auswertbar und benötigen **textuelle Beschreibung**
- Externe Verweise erfordern die **Bereitstellung der referenzierten Quellen**
- Aktueller Prototyp ist **nicht skalierbar**



Proof of concept - Verarbeitung von Normtexten mit LLMs

VIELEN DANK FÜR IHR INTERESSE!