

„Der Einsatz von KI muss menschliche Entfaltung erweitern und darf sie nicht vermindern. KI darf den Menschen nicht ersetzen.“

Alena Buyx, Ethikrat (2023)

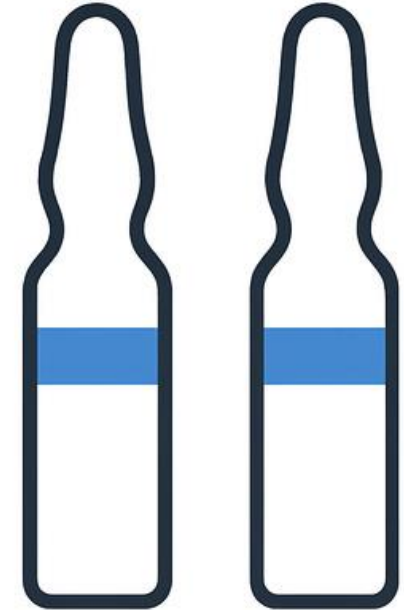
Künstliche Intelligenz trifft auf Fehlermeldesysteme – neue Perspektiven?

Dr. med Carlos R. Hölzing



Agenda

- 01 Hintergrund Fehlermeldesysteme (CIRS)
- 02 Herausforderungen der aktuellen Systeme
- 03 Potenziale durch Künstliche Intelligenz
- 04 Ausblick & Diskussion

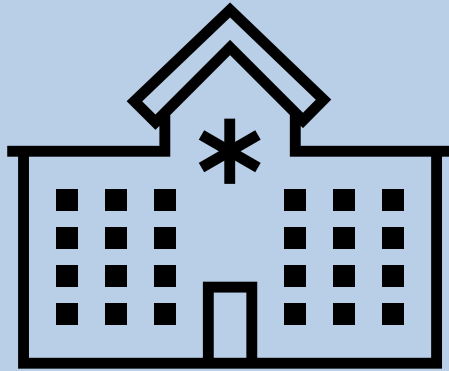


Vermeidbare Risiken müssen sichtbar werden,
bevor sie zu Schäden führen

Critical Incident Reporting System

Berichtssystem zur anonymisierten Meldung von kritischen Ereignissen (critical incidents) und Beinahe-Schäden





**Einrichtung internes
CIRS**

Pflicht

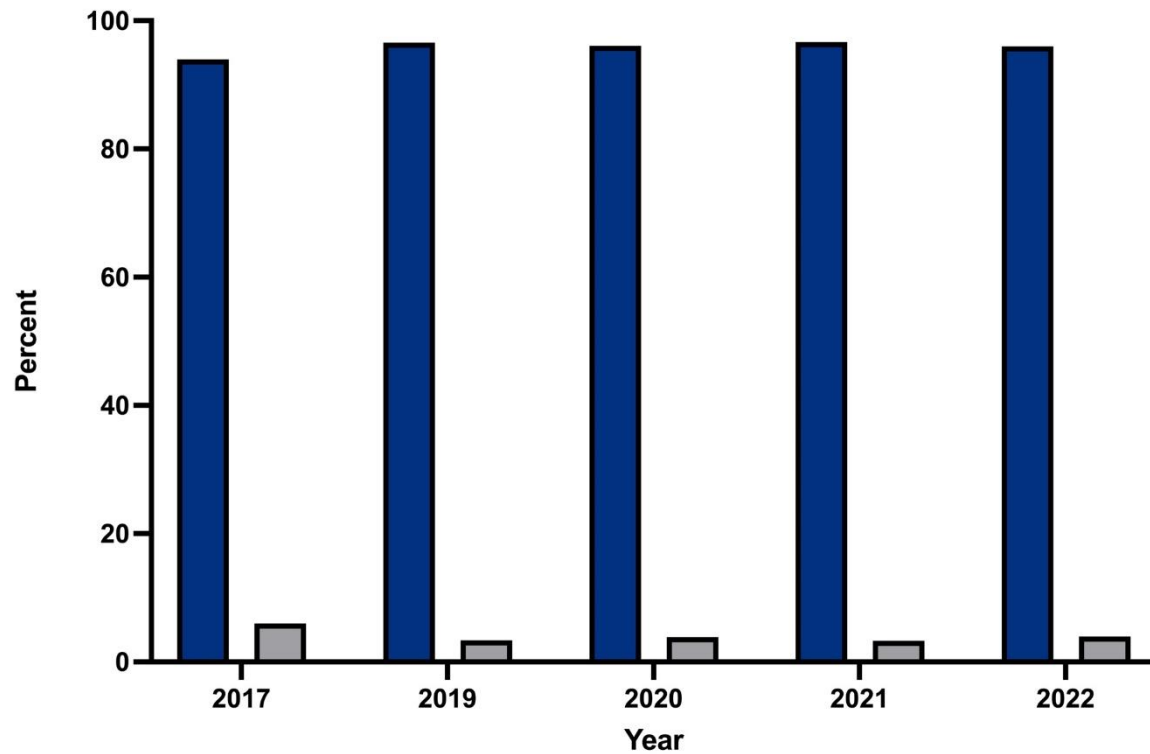


**Einrichtungübergreifendes
CIRS**

Freiwillig

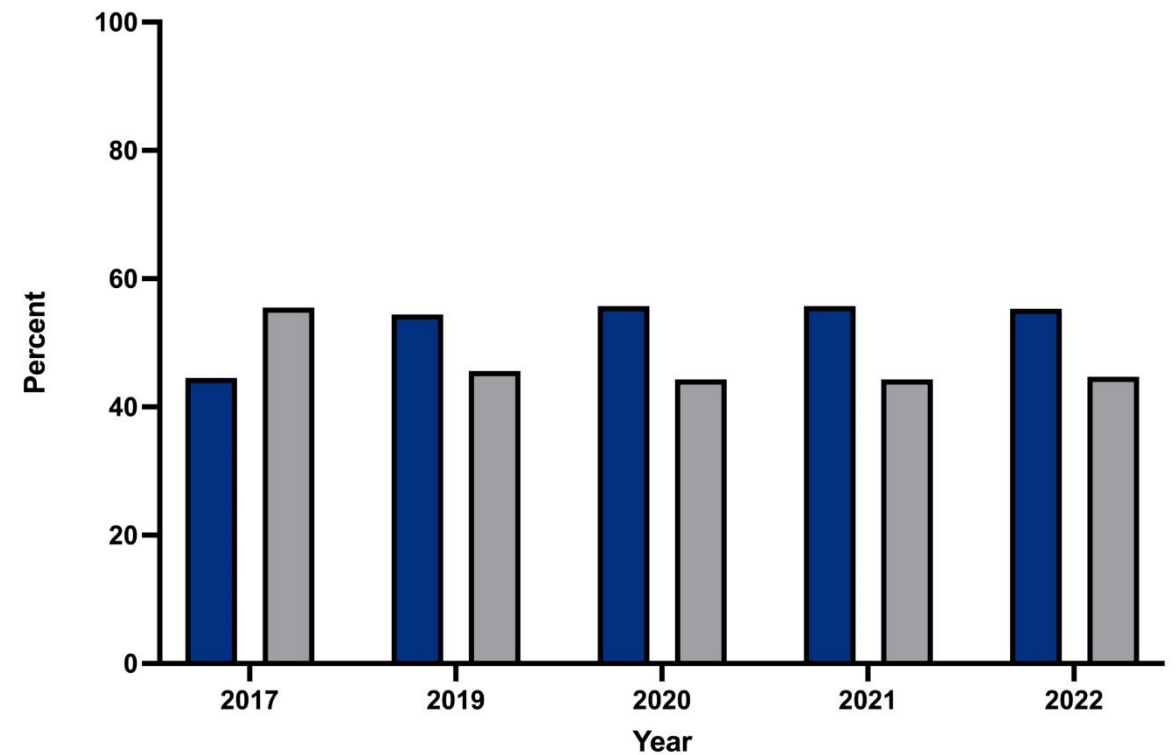
Participation in the internal incident reporting system

■ Participation ■ Non-participation



Participation in the cross-institutional incident reporting system

■ Participation ■ Non-participation



- ▶ **Kaum externe Weitergabe:** 93 % übermitteln <10 % der Meldungen an einrichtungsübergreifende CIRS
- ▶ 71 % nutzen **keine strukturierten Analysevorgaben**
- ▶ In 57 % dauert es >4 Wochen bis zur Veröffentlichung
- ▶ **Barrieren & Herausforderungen:** Zeitmangel (29 %), geringer wahrgenommener Nutzen (36 %), Vertrauen in Anonymität (14 %)

- ▶ **CIRS** wird oft nur als **Schlagwort** wahrgenommen
- ▶ **Unklarheit über Abläufe** und Prozesse
- ▶ Zweifel, ob die eigene Meldung überhaupt relevant ist
- ▶ **Mangel an psychologischer Sicherheit** beim Melden
- ▶ **Negative Assoziationen** mit CIRS (z. B. Angst vor Schuldzuweisung)

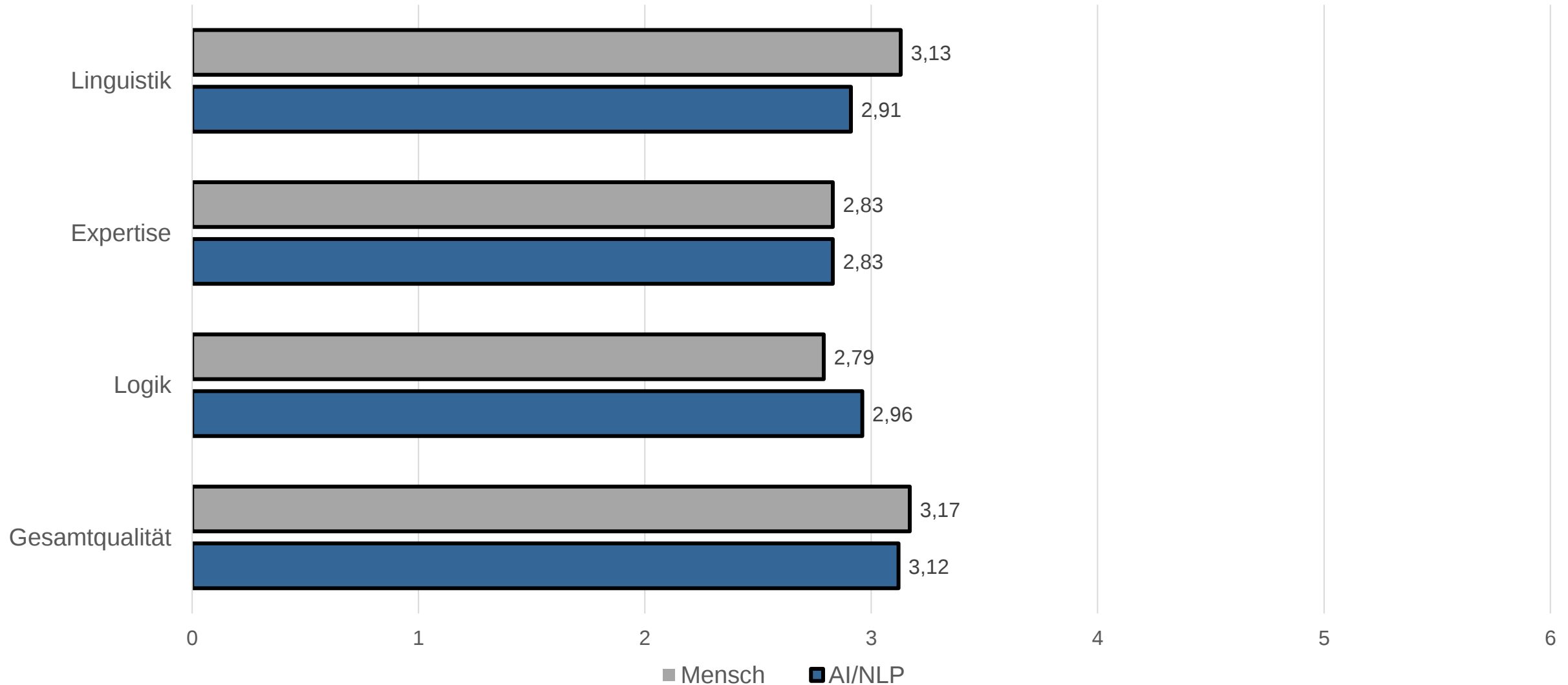


CIRS bleibt hinter Erwartungen zurück

Potenziale durch Künstliche Intelligenz

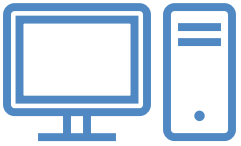


Machbarkeitsanalyse der Fallrückmeldung





- ▶ Generative KI ist stark in Sprachgenerierung



- ▶ Abhängigkeit vom Input
 - Ergebnisse hängen stark von der **Qualität des Fallberichts** ab



- ▶ Kein echtes Verstehen
 - KI generiert Text auf Basis statistischer Wahrscheinlichkeiten (Transformer-Architektur)

KI kann generative Rückmeldungen erstellen

Das Medikament war falsch

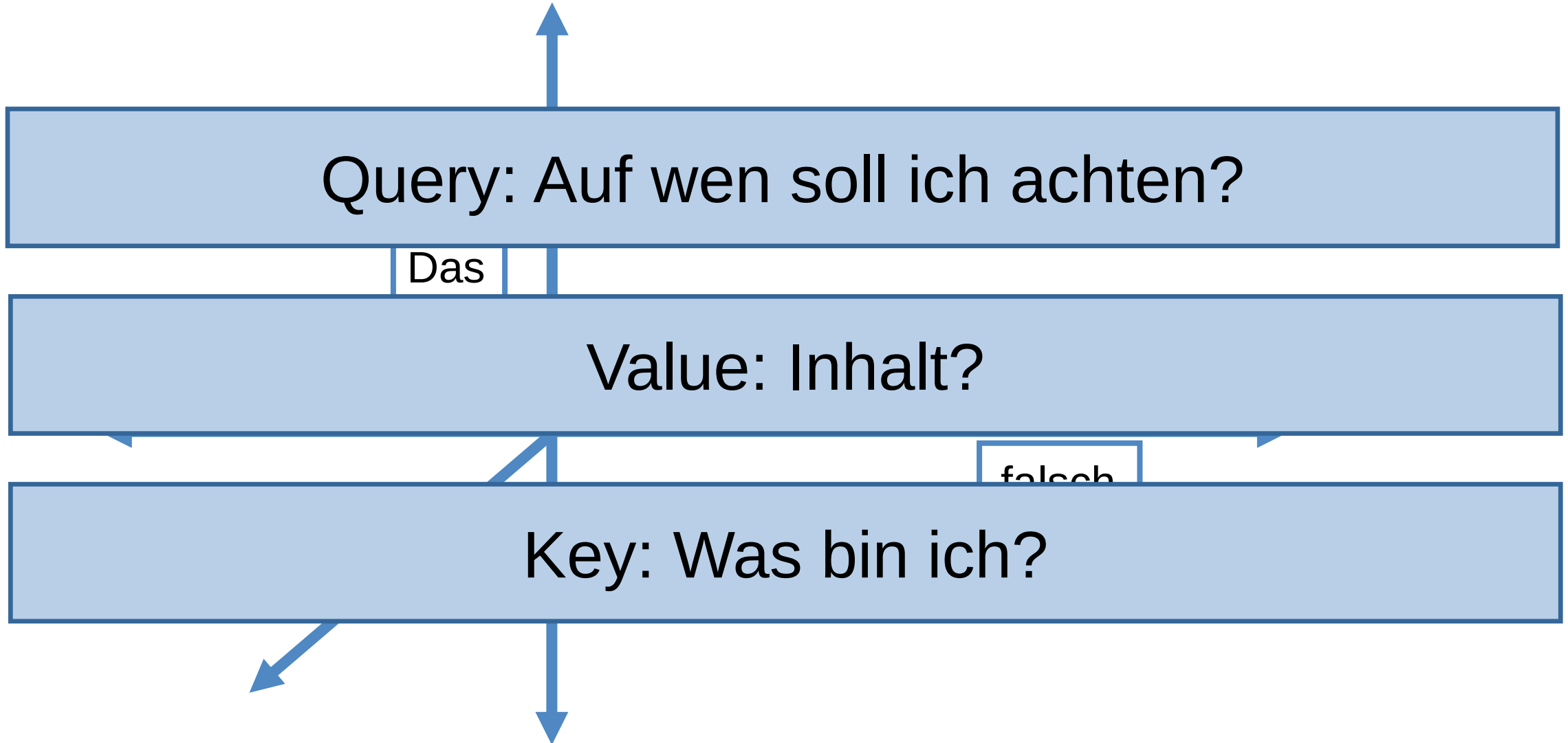
Das

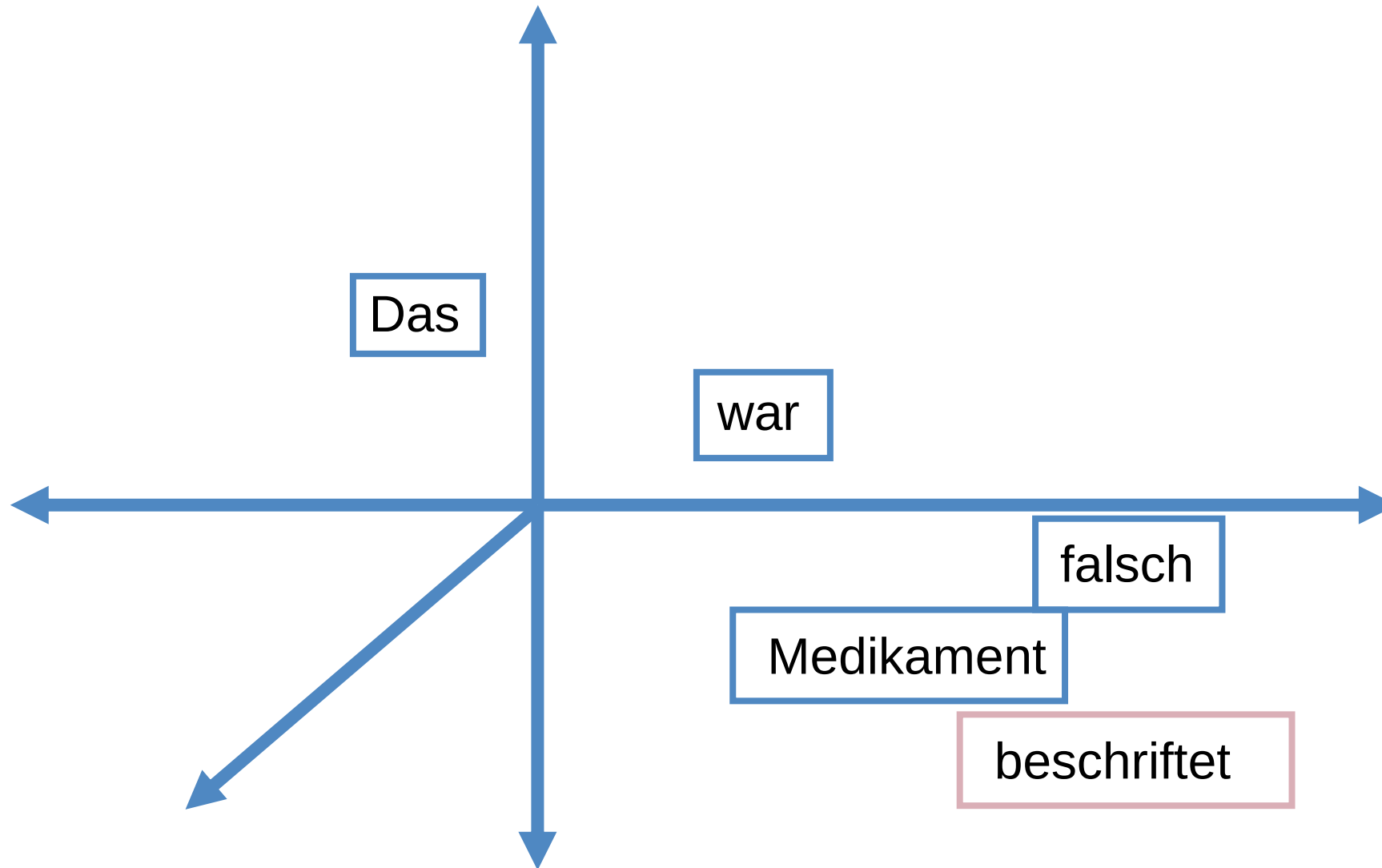
Medikament

war

falsch

Embedding = Übersetzung von Text in Vektoren





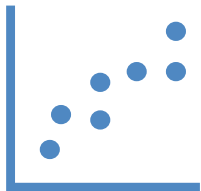
Das Medikament war falsch beschriftet



Fallklassifizierung



Emotionserkennung

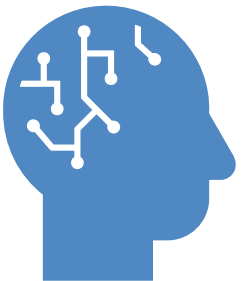


Mustererkennung



► Ausgangsproblem:

- CIRS = **unstrukturierte Freitextmeldungen**
- Keine einheitliche Klassifizierungsgrundlage



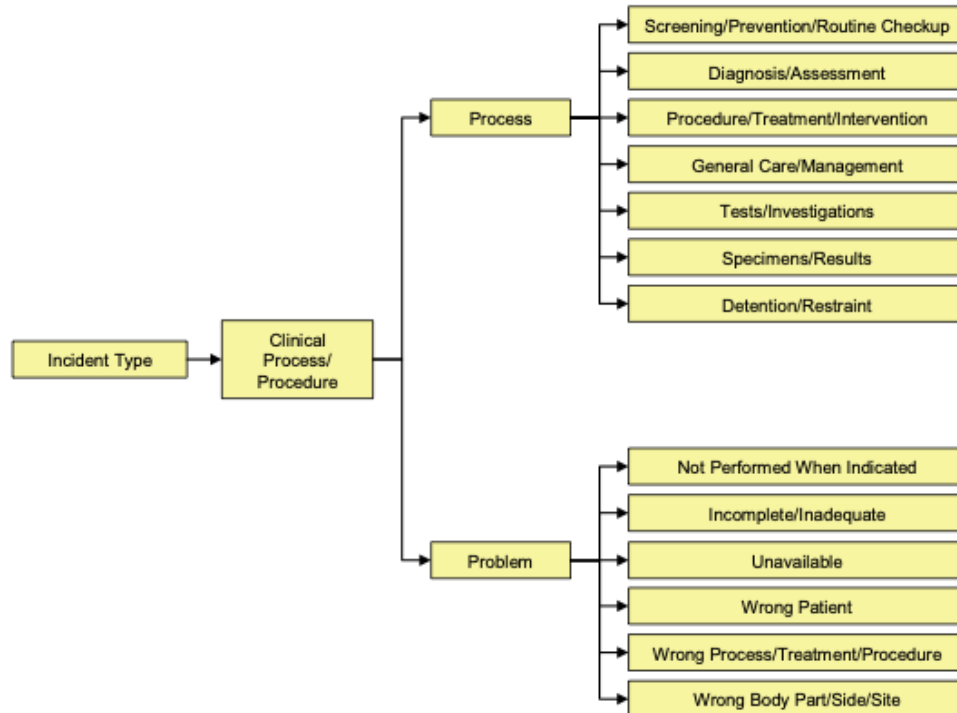
► Modelle

- Baseline: Term Frequency – Inverse Document Frequency
- Fortgeschritten: **Transformer** (GBERT)

► Unterschiedliche und undurchsichtige Klassifizierungstaxonomie

World Health Organization

Conceptual Framework for the International
Classification for Patient Safety



Uniklinikum Würzburg

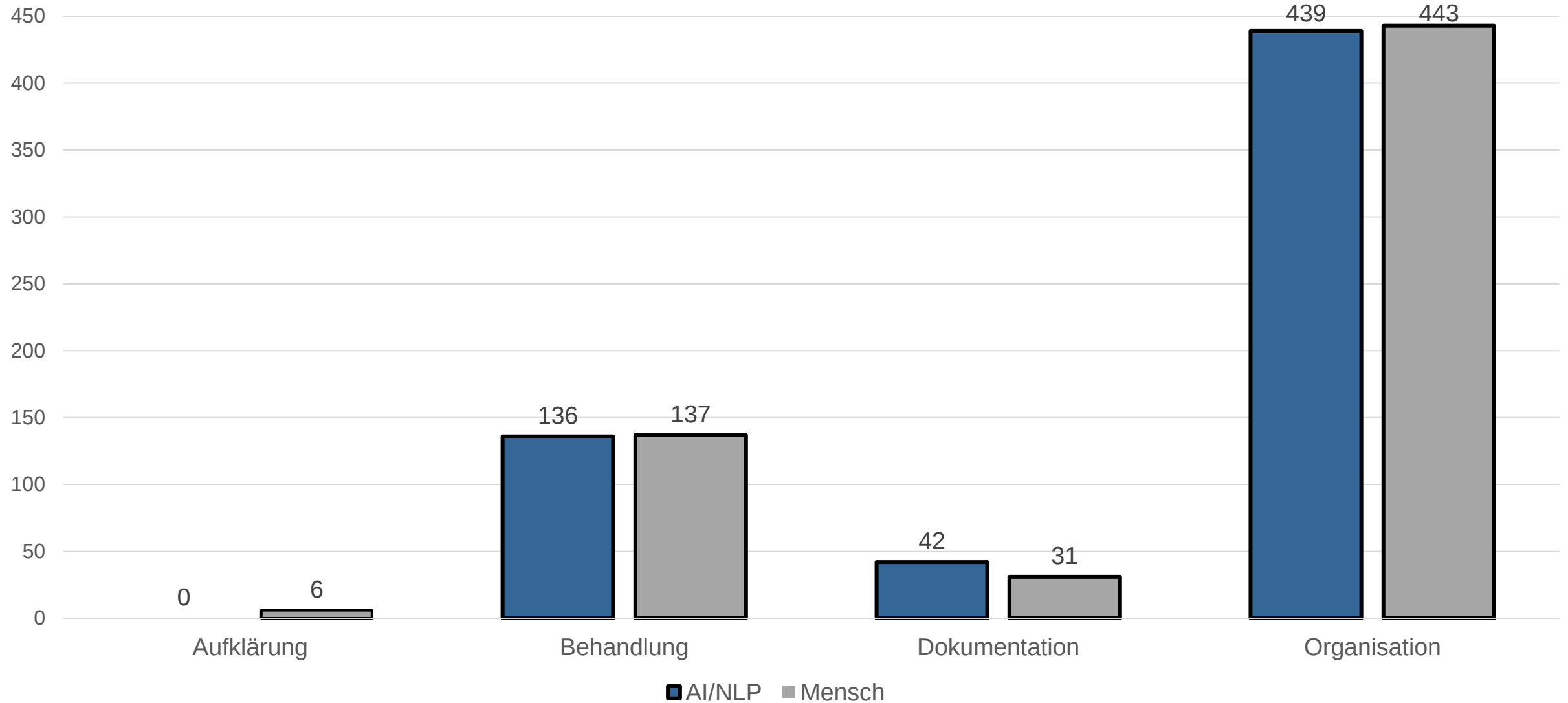
Stabsstelle medizinisches Qualitäts-, Risiko- und
Datenmanagement

Aufklärung

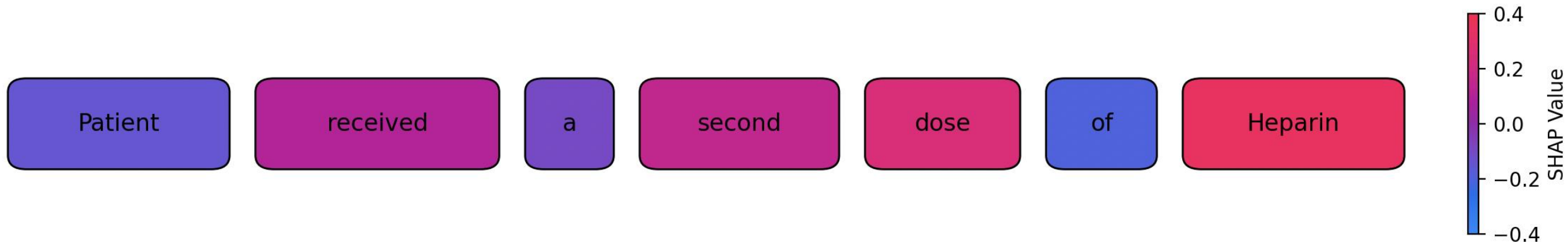
Behandlung

Dokumentation

Organisation



- ▶ SHapley Additive ExPlanations (SHAP) → Verfahren zur Modellinterpretation
- ▶ Misst den Beitrag jedes Wortes zur Klassifikation
- ▶ Farben zeigen positiven (magenta) oder negativen (blau) Einfluss





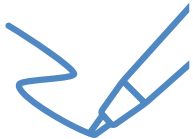
Aufklärung

Tokens wie meldete, Chirurg, danach
→ externe Zuschreibung, Hierarchie-Aspekte



Behandlung

Tokens wie Morphin, Medikationsverwechslung



Dokumentation

Selbstreflexive Wörter (ich, überprüfen, geschah)



Organisation

Tokens wie Hygiene, Datenschutz, Standards
→ zeigen systemische Schwächen

- 
- ▶ Transformer (GBERT) deutlich überlegen
 - ▶ Erklärbarkeit durch SHAP → klinisch plausible Einsichten

KI-Klassifizierung ist möglich, schwierig bei seltenen Klassen und unvollständigen Input

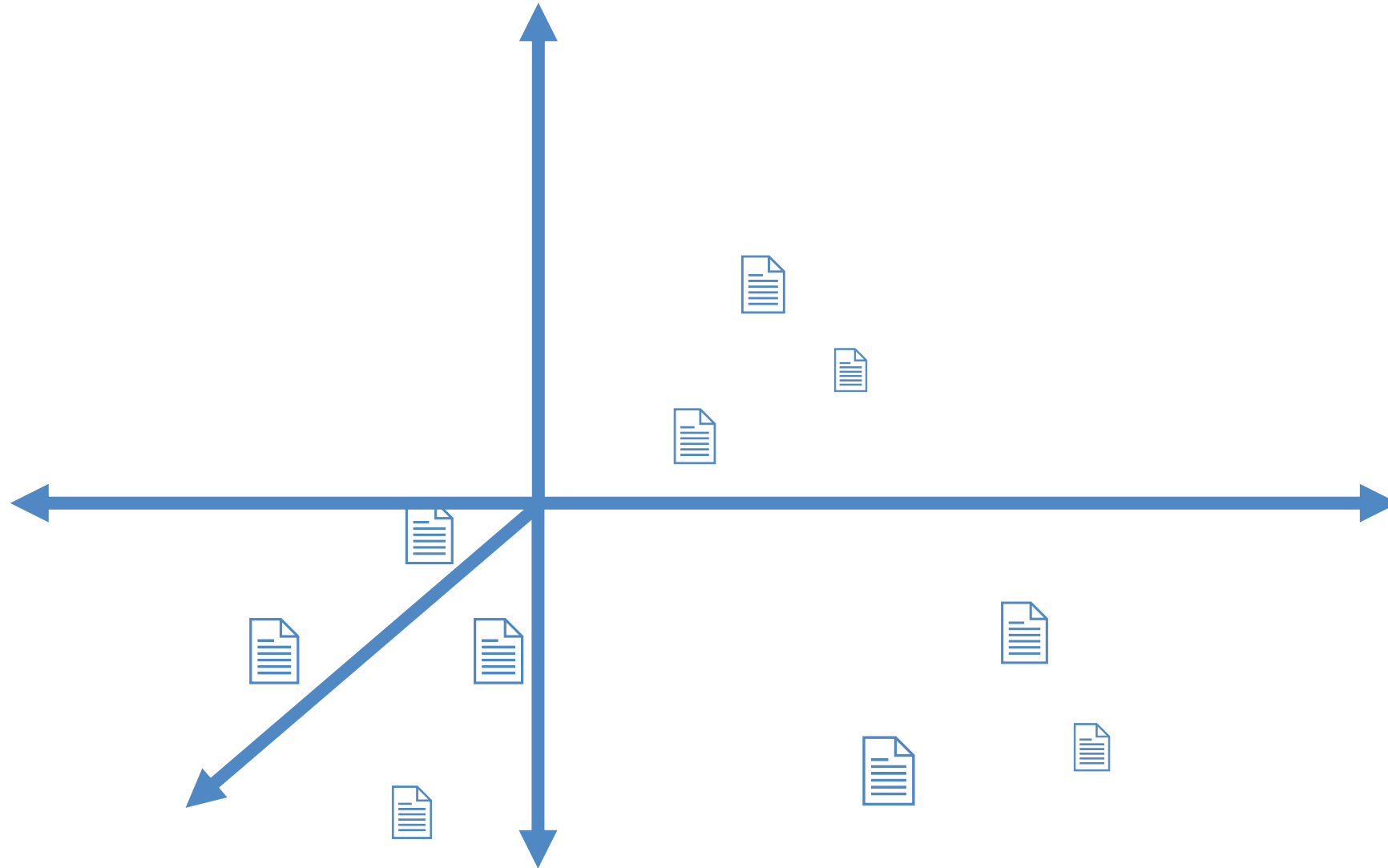
- 
- ▶ Datenlimit (seltene Klassen) → unsicheres Modell
 - ▶ Black-Box-Charakter

- ▶ CIRS = bisher Fokus auf **Struktur & Inhalte**
- ▶ Emotionale Signale kaum untersucht, obwohl entscheidend für **Risiko-Wahrnehmung & Safety-Kultur**
- ▶ Analyse von 11.056 Berichten
- ▶ **Generative pre-trained transformer** (GPT-4) für Emotionsanalyse

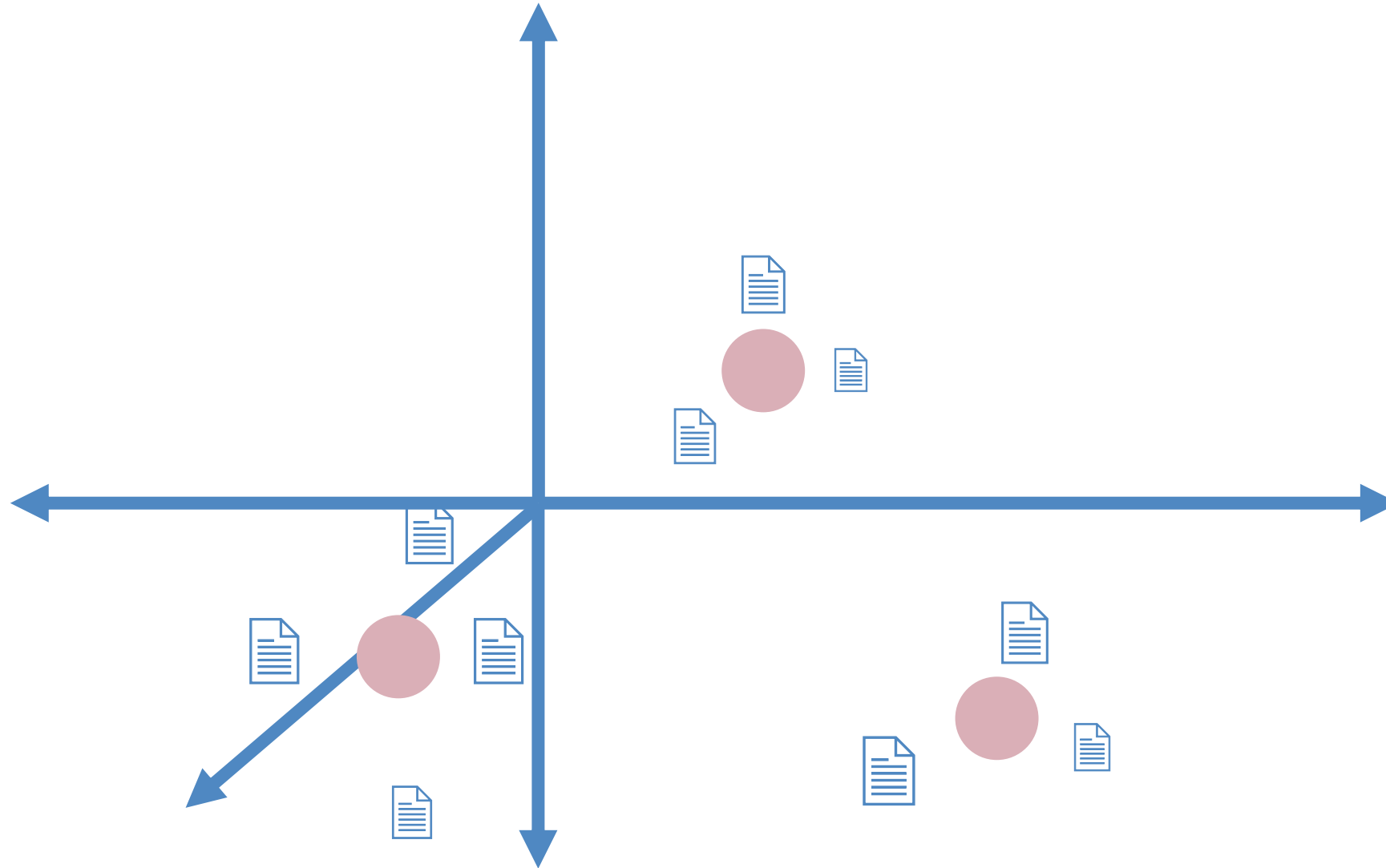
cirsmedical®

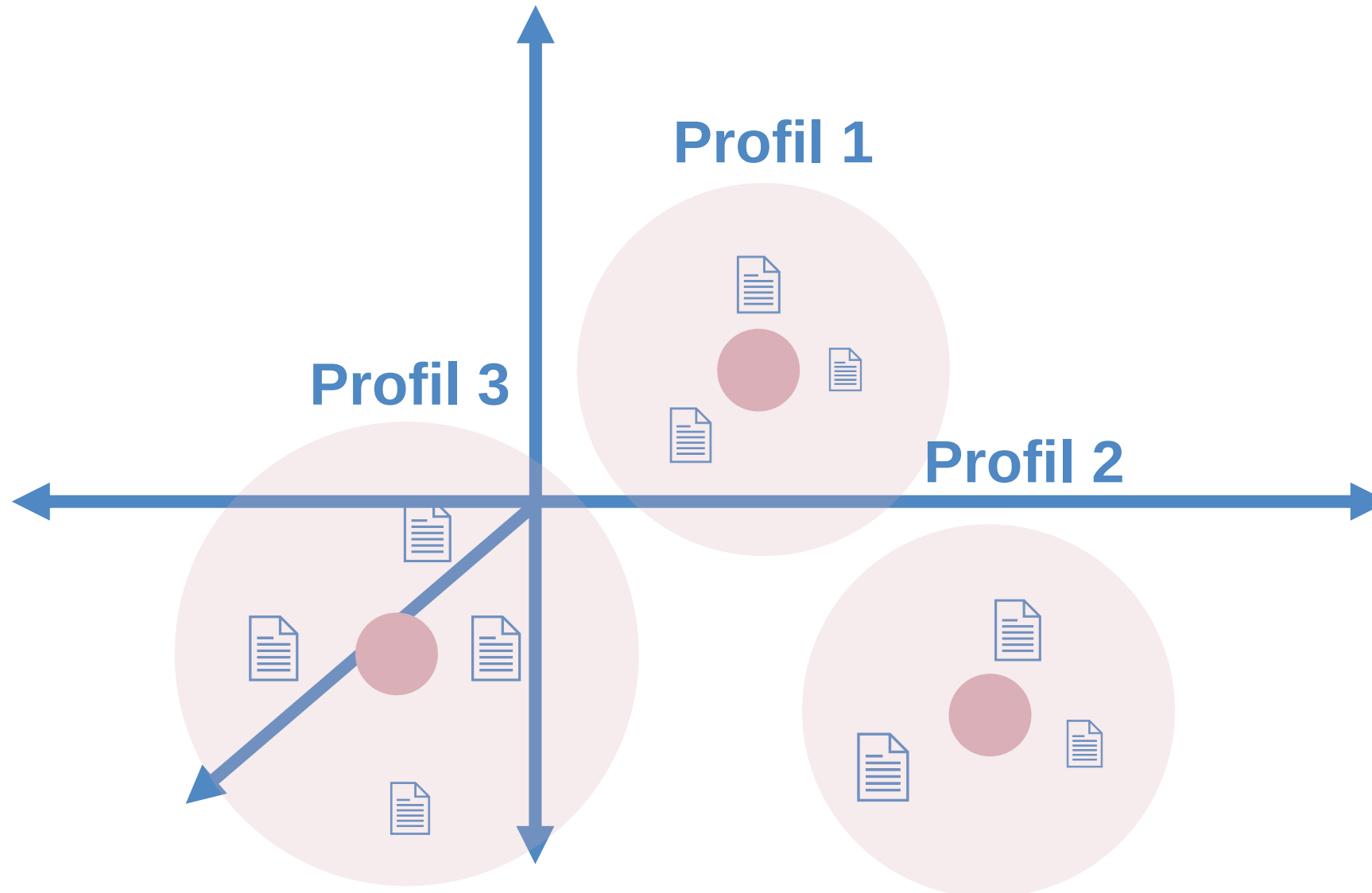
- ▶ Mithilfe des Transfordermodells wurde 5 Emotionen
 - (Angst, Frustration, Ärger, Trauer, Schuld) bewertet
- ▶ Validiert durch Annotationen
 - Cohen's κ bis 0.85
- ▶ induktiv exploriert mithilfe von **k-Means Clustering**
- ▶ **Latent Dirichlet Allocation** (LDA) zur Themenmodellierung





Exkurs: K-Means Clustering und LDA





Profil 1



„Bei der **Intubation** fehlte der **Tubus**, das **Notfallteam** kam zu **spät**.“



„Während der **Reanimation** war kein **Defibrillator** im **Raum**, das **Gerät** musste geholt werden.“



„Im **OP** kam es zu einem **Kreislaufstillstand**, doch die **Medikamente** waren nicht verfügbar.“

- ▶ **95.6 % negative** Grundstimmung
- ▶ Dominant: Angst (0.63) & Frustration (0.59)
- ▶ Höhere Angst & Schuld in Notfallsettings ($p < .05$ / $p < .001$)
- ▶ **Emotionale Berichte** (v. a. mit Angst, Schuld, Trauer) bekamen **seltener Feedback**
- ▶ **Keine** starke Korrelation mit **Schadensschwere**
- ▶ **4 Emotionale Profile** identifiziert



Hoch negativ

Hohe Angst, Frustration, Trauer, Ärger
Typisch: Notfallsituationen, hoher Stress



Neutral

Niedrige Emotionswerte, eher technisch
Typisch: Routineereignisse, standardisierte Abläufe



Frustration

Hohe Frustration & Ärger, wenig Angst
Typisch: organisatorische / prozedurale Probleme



Gemischt

Mittlere Werte in Angst, Frustration, Trauer
Typisch: komplexe, aber nicht hochakute Fälle

- ▶ Zusätzliche Dimension neben sachlicher Klassifikation
- ▶ **Frühwarnsignale:** Angst/Schuld deuten auf komplexe oder riskante Situationen hin
- ▶ **Triage-Unterstützung:** prioritäre Meldungen schneller sichtbar machen
- ▶ **Kultur-Sensibilität:** Emotionen spiegeln Belastung des Personals, Ansatz für Just Culture & Peer Support (**Second Victim**)
- ▶ **Hinweise auf Systemprobleme**

- ▶ **Emotion ≠ Schaden**: keine direkte Korrelation mit objektiver Schwere
- ▶ **Bias im Reporting**: Emotionalität hängt auch von Schreibstil, Kultur und Selbstselektion ab
- ▶ Black-Box-Systeme mit möglicher Fehlklassifikation

KI kann Emotionen aus Fallberichten erkennen. Als zusätzlicher Indikator für Belastung, Priorisierung oder Second-Victim-Risiken.

- ▶ **Künstliche Intelligenz** kann:
 - **Feedback geben, Klassifizieren, Emotionen erkennen**
- ▶ Aber nur **auf Grundlage vorhanden Informationen**
 - Weitere Probleme: unstrukturierte Daten, Imbalance

KI kann Risikomanager:innen unterstützen, indem sie deren Informationen ergänzt, bei Klassifizierung und Feedback hilft und zusätzliche Metriken zur Triage bereitstellt.

Vielen Dank für Ihre Aufmerksamkeit!

Dr. med. Carlos R. Hölzing

E-Mail: Hoelzing_C@ukw.de

Tel: 0931/201-30037