



W4 - KW20

Monday, 12 May 2025 17:22

@Tech Team

Auswahl der Sprachmodelle

- Entscheidung für zwei LLMs: **Mistral** und **Phi** (werden wir so lange es geht parallel testen)
- Erste Tests durchgeführt – **Mistral** zeigt vielversprechende Ergebnisse, **Phi** ist sehr effizient
- Einsatz über **Ollama**: Modelle laufen lokal und ermöglichen flexible, offline-fähige Entwicklung
ersten Prompt-Tests zeigen: beide Modelle erkennen Unterschiede zwischen absoluten und relativen Risiken

Arbeitsumgebung

- Festlegung auf **Visual Studio Code** als zentrale Entwicklungsumgebung
- gute Integration (github, ollama), stabil, ideal für schnelle Tests und LLM-Promptentwicklung

Entwicklung einer Analyse-Matrix

- Ziel: Risikokommunikation systematisch beschreibbar und für das LLM nutzbar machen
- Anwendung der Matrix: hilft beim Labeling von Trainingsdaten und bei der späteren Ausgabe im Tool
- Die Matrix enthält zentrale Bestandteile:
 - **Grundrisiko**
 - **Neues absolutes Risiko**
 - **Relatives Risiko**
 - **Referenzklasse**
 - **Absolute Zahl**
 - **Größe der Grundgesamtheit**
 - **Inhalt der Risikobotschaft**

Im Austausch mit dem Data-Team wurden die Anforderungen an die Trainingsdaten festgelegt

@Data Team

WebScraper – Tests & Versionen

Ziel: Finden von Texten und Textabschnitten, in denen Risikokommunikation gemäß unserer obigen Definition vorkommt

Ansätze:

- Keywordbasiert: Scrapen von Texten nach Keywords wie zB "relative risk", "covid" etc...
 - Später Kombination von Zeichen wie zB "%"
- Funktionweise:
 - Komplette Artikel werden rausgesucht -> in Excel gespeichert (Anzahl beliebig)
 - Einzelne Sätze werden nochmal extra abgespeichert, sodass diese schneller gefunden werden können
- Sobald Tabelle final -> Scrapen von ca 100 Texten -> die dann gelabelt werden um erstes Training mit diesem Daten durchzuführen (hoher Arbeitsaufwand)

Datensuche Ansätze:

- Alternative 1: Webscraper
- Alternative 2: Manuelles raussuchen von besonders hochqualitativen Datensätzen -> zB statistisches Bundesamt Artikel, "gute" wissenschaftliche Artikel
- Alternative 3: Kombination aus Alternative 1 und 2

