



## W3 - KW19

Monday, 12 May 2025 17:22

### @Tech Team

#### Diskussion zur technischen Architektur

- Überlegungen zur Nutzung von Encoder/Decoder-Modellen zur Informationsextraktion

#### Auswahl möglicher Sprachmodelle

- Evaluierung von BERT, DeBERTa, DistilBERT
- Diskussion über Mistral-7B als praktikable Option und erste Tests (<https://github.com/kaihag250/Riscom>)
- Erwägung von LLaMA – nähere Prüfung steht noch aus

#### Vergleich IDEs

- Erstmal für VS Code entschieden
- PyCharm (stabil, aber potenziell langsam)
- JupyterLab gut für LoRA, aber nicht ideal für API-gesteuerte Anwendungen

#### Evaluation unterstützender Tools

- Analyse von Ollama: Lokaler LLM-Server, nutzt Mistral offline
- HuggingFace erkundet: Modell- und Datensatzauswahl

### @Data Team

#### Umsetzung automatischer Beschaffung von Newsartikeln

- Festlegung auf Umsetzung mithilfe von News APIs und Python Bibliotheken
- Evaluierung von verschiedenen APIs und Vergleich bezüglich Funktionalität und Limitationen
- Entwickeln eines Prototypen und Weiterentwicklung

#### Auswahl passender Textdaten

- Herausarbeiten von Schlüsselthemen, welche oft im Kontext Risikokommunikation vorkommen (z.B Medizin, Naturkatastrophen, Finanzmärkte)
- Prompting der APIs mit entscheidenden Keywords (z.B Covid + Risk)
- Extraktion erster Datensätze zu gegebenen Themen in Excel File

#### Erste Überlegungen bzgl. Labeln der Datensätze

- Welche Informationen über Risikokommunikation sind für ein Modell relevant
- Sammeln von Potentiell relevanten Kategorien --> Festlegung tbd.