

# Mittelstandskonferenz 2023

KMU stärken durch digitale Innovationen

FedXtract ermöglicht KMU die effiziente Nutzung von KI durch verbessertes Modelltraining mit ausreichenden Trainingsdaten unter Wahrung der Datenhoheit, sowie Erweiterung des KI-Know-hows.

## Aufgabenstellung

- KMUs möchten Deep Learning einsetzen, aber die benötigten sensiblen, verteilten Daten sind nicht aggregierbar
- Föderiertes Lernen (FL) ermöglicht ein verteiltes Training von Modellen, ohne dass Trainingsdaten die Server verlassen
- Entwicklung einer Best Practice und Beispiele zum Einsatz eines FL Frameworks
- Unterstützung für KMUs zur Einführung durch Schulung und Customizing

## Use Case Document Recognition

- Erkennung von Aufträgen, Gutachten oder Rechnungen der Automobilbranche
- Dokumente mit Adresse, Bankverbindung Fahrzeugteile in Rahmen und Tabellen
- Tools: OCR, graphische Analyse, NLP
- ca. 20 Millionen gelabelte Trainingsdaten

## Transformer-Modelle

- Aufsetzen auf Modelle: Falcon, Mistral, LayoutXLM, LLaMA
- Technologien: Python, PyTorch, Keras, Docker, GitLab, Nuance, Nvidia
- Vokabular 30.000 Token, Hidden Layers 12, Attention Heads 12

## Federated Learning (FL)–Frameworks

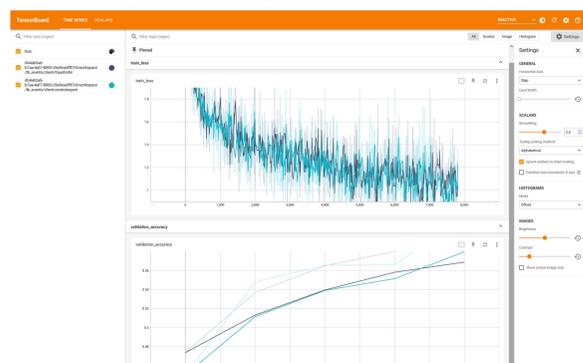
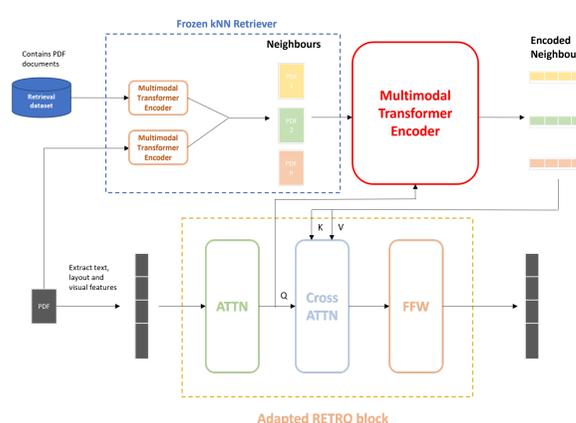
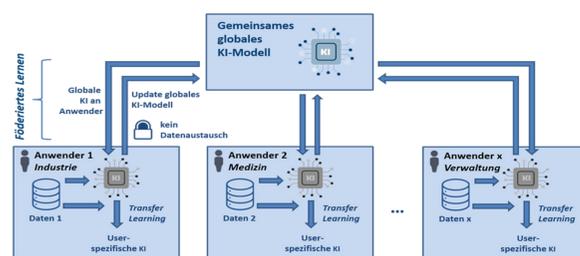
- Evaluation von 9 FL-Frameworks
- Kriterien: Open Source, Community, Stabilität, technologie-agnostisch
- Favoriten: Nvidia, Flower, PySift, Tensorflow FL

## Verteiltes Training

- Zentrale Administration und Monitoring
- 2 – n verteilte Clients
- Homogene / heterogene Datenverteilung

## Validierung

- Iterative Testläufe mit verschiedenen Modellen und Trainingsdaten
- Variable Anzahl von Clients und Epochen
- Evaluation mit TensorBoard
- Messung von Loss, Accuracy, F1-Score



# FEDXtract

**Föderierte KI-Lösung zur Verarbeitung komplexer Dokumente**

**Laufzeit: 2021-2024**

**Kontakt: Rolf Fricke, Condat AG**

**FKZ: 01IS21037A**

**Projektpartner:**



GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

