Advanced Information Retrieval

Fine-Tuning and Transferability in Legal Information Retrieval

Group Number: 27

Raphael Habichler[*]   Mark Sesko[†]   Paul Brandstätter[‡]

November 11, 2025

1 Abstract

This project investigates the transferability of domain-specific fine-tuned embedding models across different legal jurisdictions. We fine-tune the Qwen3-Embedding-8B model on German legal data and evaluate its performance on Austrian, German, and Chinese legal corpora to assess cross-jurisdictional knowledge transfer.

2 Idea and Goal

Research Question: Can a legal embedding model fine-tuned on German law effectively retrieve relevant documents in other jurisdictions (Austrian and Chinese law)?

The goal is to build legal information retrieval systems that work across multiple jurisdictions without requiring extensive training data for each legal system. This addresses the challenge of cross-jurisdictional legal search and comparative law analysis.

3 Main Task

Our focus is on fine-tuning a state-of-the-art embedding model (Qwen3-Embedding-8B [4]) on German legal data and measuring how well this domain-specific knowledge trans-

---

[*]student-id: 12419578, Role: Data Processing & Fine-tuning
[†]student-id: 12114879, Role: Evaluation & Analysis
[‡]student-id: 12212566, Role: Dataset Preparation & Documentation

fers to other legal systems, specifically Austrian and Chinese law.

4   Dataset and Processing

4.1   Data Sources

- **Austrian Law Data**: Legal documents and case law from Austrian jurisdiction

- **German Law Data**: GerDaLIR dataset [3] and German legal corpus [1]

- **Chinese Law Data**: LeCaRDv2 dataset [2]

4.2   Processing

Data preprocessing includes text normalization, document segmentation, and creation of query-document pairs for training and evaluation purposes.

5   Methods and Models

We use the Qwen3-Embedding-8B model [4] and compare two experimental conditions:

1. **Baseline**: Pre-trained Qwen3-Embedding-8B evaluated directly on all three jurisdictions

2. **Fine-tuned**: Model fine-tuned on German domain-specific dataset, then evaluated on all three jurisdictions

The fine-tuning process uses cosine-loss similarity for training. Query and document texts are encoded and pooled, then their cosine similarity is computed. This trains the model to produce embeddings where relevant query-document pairs have high similarity.

6   Evaluation

We measure retrieval performance using standard information retrieval metrics:

- **Precision@k**: Proportion of relevant documents in top-k results

- **Recall@k**: Proportion of all relevant documents found in top-k results

- **nDCG@k**: Normalized Discounted Cumulative Gain, accounting for ranking quality

Performance is compared between the baseline and fine-tuned models across all three jurisdictions to quantify transferability.

7 Workflow

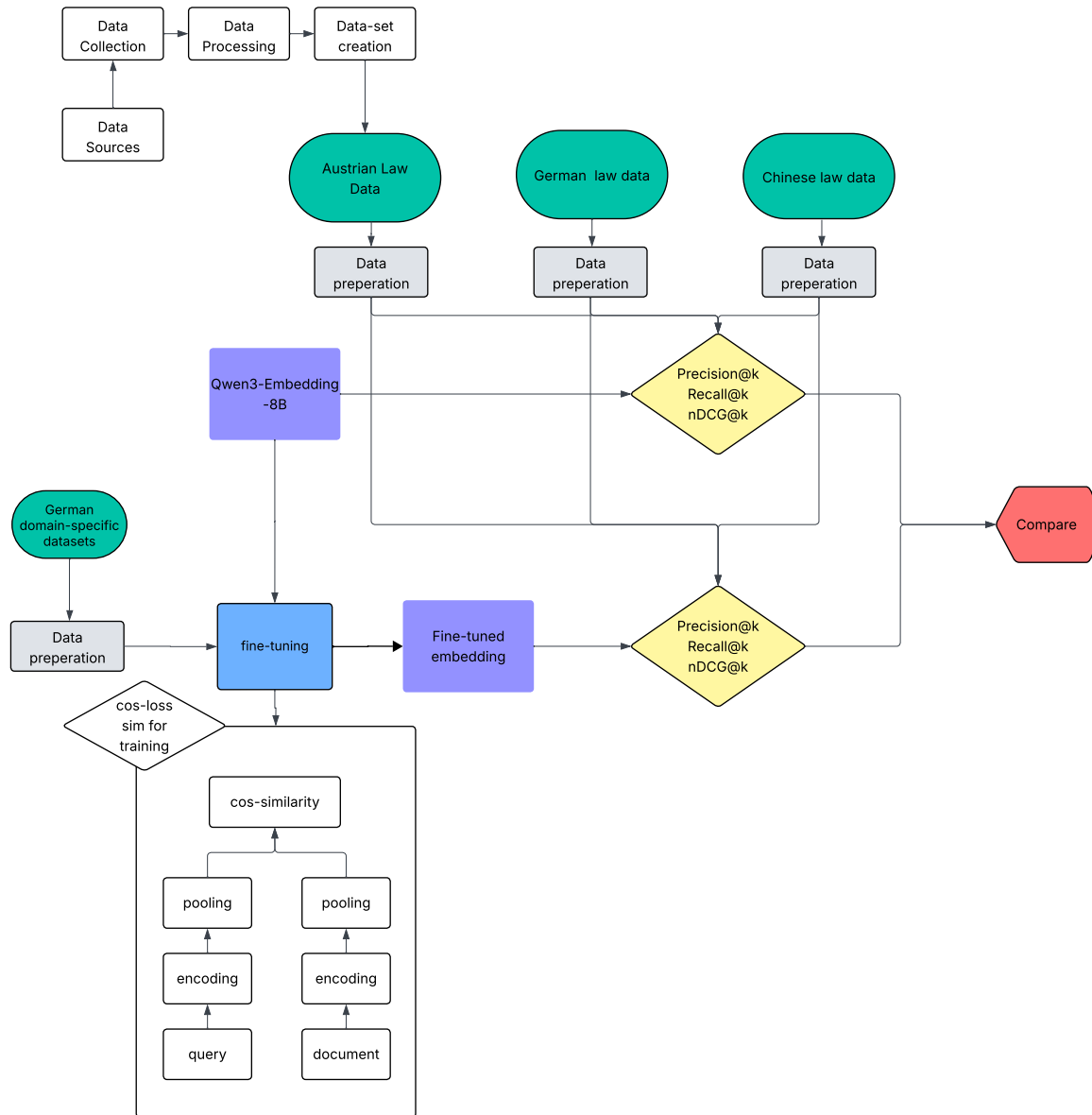Figure 1 illustrates our complete research pipeline from data collection through evaluation.

Figure 1: Project workflow and methodology

References

[1]   Christoph Hoppe et al. "Towards Intelligent Legal Advisors for Document Retrieval and Question-Answering in German Legal Documents". In: *2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*. 2021, pp. 29–32. DOI: `10.1109/AIKE52691.2021.00011`.

[2]   Haitao Li et al. *LeCaRDv2: A Large-Scale Chinese Legal Case Retrieval Dataset*. 2023. arXiv: `2310.17609 [cs.CL]`.

[3]   Marco Wrzalik and Dirk Krechel. "GerDaLIR: A German Dataset for Legal Information Retrieval". In: *Proceedings of the Natural Legal Language Processing Workshop 2021*. Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 123–128. URL: `https://aclanthology.org/2021.nllp-1.13`.

[4]   Yanzhao Zhang et al. "Qwen3 Embedding: Advancing Text Embedding and Reranking Through Foundation Models". In: *arXiv preprint arXiv:2506.05176* (2025).