Data Collection Pipeline

Query word selection with GPT4

- JSDF
- Defence
- Security
- …

Query Words

API

Japanese Legislative Speech API

Raw Sentence Segments

Sentence Segment Dataset

Supervised Fine-tuning

BERT Classifier

SBERT Embedder

Opinion-based sentence embeddings