Legal Case Retrieval for a Given Legal Scenario

Student - Rajith Arulanandam

Index – 248330N

Supervisor - Dr. Nisansa de Silva

1. Introduction

Why Precedent Matters & Why It's Tough to Find

- Precedent = past court decisions that guide how judges rule today (doctrine of *stare decisis*). [1,2]
- Lawyers must locate cases with similar facts, the same legal issue, and a supportive outcome—or distinguish the ones that hurt their side. [1,2]
- The search space is messy: [1,3]
- Millions of cases across many courts, countries, and decades. [3,4]
- Changing language & laws ("duty of care" in English ≈ different terms elsewhere). [3]
- Different authority levels (supreme vs. trial courts, local vs. foreign). [1,2]
- Missing a key case can sink an argument; finding one early saves time and may even force settlement. [2

[1] Feng *et al.* *Legal Case Retrieval: A Survey*, 2024.
[2] Locke & Zuccon. *Case Law Retrieval: Problems, Methods, Challenges*, 2022.
[3] Sansone & Sperlí. *Legal Information Retrieval Systems: SOTA & Open Issues*, 2022.
[4] Ma *et al.* *Retrieving Legal Cases from a Large-Scale Candidate Corpus* (COLIEE 2021), 2021.

2. Research Problem

Three Obstacles We Must Overcome

- **Layered relevance** (for better accuracy and relevance). [1,2,5]
  - A “good” precedent must match on facts + legal issue + outcome, not just keywords. [1,2,5]
  - Audits show many systems get one layer right and miss the others. [2,5,6]
- **Jurisdiction & language variety.** [1,3,7,8]
  - Case law spans different courts, legal traditions, and languages. [1,3,7,8]
  - Models that ace Canadian data can flop on Chinese or South-Asian corpora. [3,7,8]
- **Need for transparent justification.** [6,9,10,11]
  - Lawyers cite paragraphs, not scores. [6,9,10,11]
  - Tools that highlight the exact passage earn far higher practitioner trust. [6,9,10,11]

[1] Feng *et al.* *Legal Case Retrieval: A Survey*, 2024.
[2] Locke & Zuccon. *Case Law Retrieval: Problems, Methods, Challenges*, 2022.
[3] Sansone & Sperlí. *Legal Information Retrieval Systems: SOTA & Open Issues*, 2022.
[5] Li *et al.* *Towards an In-Depth Comprehension of Case Relevance*, 2024.
[6] Smith *et al.* *Human Evaluation Experiment of Legal IR Methods*, 2023.
[7] Rabelo *et al.* *COLIEE 2021 Overview & Discussion*, 2022.
[8] Bi *et al.* *Heterogeneous Graph Embedding for Chinese Legal Doc Similarity*, 2022.
[9] Pipitone & Houir Alami. *LegalBench-RAG Benchmark*, 2024.
[10] Louis *et al.* *Interpretable Long-Form Legal QA*, 2024.
[11] Yu *et al.* *Explainable Legal Case Matching via Rationale Extraction*, 2022.

3. Research Objectives

Research Objectives

- **O1 – Cross-Jurisdiction Ranking & Retrieval.** [1,5,12]
  Build a hybrid model that finds and ranks precedents by legal weight *and* factual closeness—even when terminology, citation style, or doctrine differ across courts, languages, or traditions. [1,5,12]
  Why it matters: Lawyers can rely on the top results, no matter where (or in what language) the key case was decided. [1]
- **O2 – Information-Rich Vector Knowledge Base.** [1,9,12]
  Represent every judgment with dense embeddings **plus** annotated metadata (facts, issues, statutes, cited cases). [1,9,12]
  Why it matters: Enables precise filters (e.g., only Supreme Court cases after 2010) and multi-dimensional similarity scoring. [1,9]
- **O3 – Transparent, Paragraph-Level Explanations.** [5,9,10]
  Return each case with highlighted snippets and a confidence signal showing exactly how it aligns on facts, issue, and outcome. [5,9,10]
  Why it matters: Gives practitioners copy-and-paste citations and instant insight—no black-box scores. [9,10]

[1] Feng *et al.* *Legal Case Retrieval: A Survey*, 2024.
[5] Li *et al.* *Towards an In-Depth Comprehension of Case Relevance*, 2024.
[9] Pipitone & Houir Alami. *LegalBench-RAG Benchmark*, 2024.
[10] Louis *et al.* *Interpretable Long-Form Legal QA*, AAAI 2024.
[12] Li *et al.* *SAILER: Structure-Aware Pre-Trained LM for Legal Case Retrieval*, SIGIR 2023.

4. Current Literature

4.1 Datasets Used

Why Curated Datasets Matter

- Benchmark progress & reproducibility. [1,3,7,9]
- Cross-jurisdiction comparability (English, Chinese, German). [7,8,13]
- Stress-tests for reasoning (Bar Exam QA, Housing Statute QA). [9,14,15]
- Graded labels enable fine-grained evaluation. [5,6,9]

[1] Feng *et al.* *Legal Case Retrieval: A Survey*, 2024.
[3] Sansone & Sperlí. *Legal Information Retrieval Systems: SOTA & Open Issues*, 2022.
[5] Li *et al.* *Towards an In-Depth Comprehension of Case Relevance*, 2024.
[6] Smith *et al.* *Human Evaluation Experiment of Legal IR Methods*, 2023.
[7] Rabelo *et al.* *COLIEE 2021 Overview & Discussion*, 2022.
[8] Bi *et al.* *Heterogeneous Graph Embedding for Chinese Legal Doc Similarity*, 2022.
[9] Pipitone & Houir Alami. *LegalBench-RAG Benchmark*, 2024.
[13] Müller *et al.* *GerDaLIR: German Legal IR Dataset*, 2022.
[14] Wiratunga *et al.* *CBR-RAG: Case-Based Reasoning for RAG in Legal QA*, 2024.
[15] Zheng *et al.* *A Reasoning-Focused Legal Retrieval Benchmark*, 2025.

Flagship Benchmark 1 - COLIEE

- **Jurisdiction / Language:** Canadian common-law; English. [7,17]
- **Scale:** ≈4.4k candidate cases; 200+ query cases (2023). [4,7,17]
- **Core tasks:** case retrieval, entailment, statute QA. [7,17,18]
- **Community impact:** default test-bed for graph-aware models (e.g., CaseLink). [16,7]
- **Design choice:** annual refresh keeps queries new but corpus fixed – "TREC-style" stability. [7,18]

[4] Ma *et al.* *Retrieving Legal Cases from a Large-Scale Candidate Corpus* (COLIEE 2021 Workshop), 2021.
[7] Rabelo *et al.* *COLIEE 2021 Overview & Discussion*, 2022.
[16] Tang *et al.* *CaseLink: Inductive Graph Learning for Legal Case Retrieval* (SIGIR 2024), 2024.
[17] Kim *et al.* *Legal Information Retrieval & Entailment Using Transformer-Based Approaches*, 2024.
[18] Askari *et al.* *Leibi@COLIEE 2022: Aggregating Tuned Lexical Models & Cluster-Driven BERT*, 2022.

Flagship Benchmark 2 – CAIL LCR Tracks

- Jurisdiction , Mainland-Chinese criminal | Chinese [19,20]
- Scale: > 1 M judgments distilled to 1.2 k labelled queries [19,20,21]
- Drives Chinese SOTA systems (SAILER, SLR) [12,20]
- Long-document nature , perfect for multi-stage, high-recall pipelines [12,20,21]

[12] Li *et al.* SAILER: Structure-Aware Pre-Trained LM for Legal Case Retrieval, SIGIR 2023.
[19] Xiao *et al.* CAIL2018: Large-Scale Legal Dataset for Judgment Prediction, 2018.
[20] Ma *et al.* Incorporating Structural Information into Legal Case Retrieval (SLR), ACM TOIS 2024.
[21] Hu *et al.* BERT_LF: Similar Case Retrieval Based on Legal Facts, 2022.

Other Key Corpora at a Glance

- **LeCaRD v2** – Chinese passage IR; nDCG gains via query expansion. [22,20,21]
- **GerDaLIR** – 123k German passages; graph negatives add +7 MRR. [13,23,16]
- **LegalBench-RAG** – 6.9k Q/A/snippet triples for RAG faithfulness. [9,14,10]
- **Bar Exam QA & Housing Statute QA** – 10k reasoning-heavy queries. [15,9,10]
- **SCaLe-QA** – Sinhala/Tamil embeddings for Sri-Lankan law. [24,1]

[1] Feng *et al.* *Legal Case Retrieval: A Survey*, 2024.
[9] Pipitone & Houir Alami. *LegalBench-RAG Benchmark*, 2024.
[10] Louis *et al.* *Interpretable Long-Form Legal QA*, AAAI 2024.
[13] Müller *et al.* *GerDaLIR: German Legal IR Dataset*, 2022.
[14] Wiratunga *et al.* *CBR-RAG: Case-Based Reasoning for RAG in Legal QA*, 2024.
[15] Zheng *et al.* *A Reasoning-Focused Legal Retrieval Benchmark*, 2025.
[16] Tang *et al.* *CaseLink: Inductive Graph Learning for Legal Case Retrieval*, SIGIR 2024.
[20] Ma *et al.* *Incorporating Structural Information into Legal Case Retrieval*, TOIS 2023.
[21] Hu *et al.* *BERT_LF: Similar Case Retrieval Based on Legal Facts*, 2022.
[22] Zhou *et al.* *Boosting Legal Case Retrieval by Query Content Selection with LLMs*, 2023.
[23] Bhattacharya *et al.* *Legal Case Document Similarity: You Need Both Network and Text*, IP&M 2022.
[24] Perera *et al.* *SCaLe-QA: Sri Lankan Case-Law Embeddings for Legal QA*, 2023.

Dataset Challenges & Advanced Labels

- **Long opinions** – need paragraph chopping + offset maps. [12,11]
- **Expert labels expensive** – synthetic query–case pairs at scale (e.g., LEAD ~700k). [25]
- **Reasoning gaps** – low-overlap questions break keyword baselines. [15]
- **Zero-shot domain shift** – large performance drops when English-trained models are tested on Chinese data (reported up to –26 nDCG cross-corpus). [12,15]

[11] Yu *et al. Explainable Legal Case Matching via Inverse Optimal Transport-Based Rationale Extraction*, SIGIR 2022.
[12] Li *et al. SAILER: Structure-Aware Pre-Trained LM for Legal Case Retrieval*, SIGIR 2023.
[15] Zheng *et al. A Reasoning-Focused Legal Retrieval Benchmark*, CS+Law 2025.
[25] Gao *et al. Enhancing Legal Case Retrieval via Scaling High-Quality Synthetic Query-Candidate Pairs*, EMNLP 2024.
[26] Kim *et al. Result Diversification for Legal Case Retrieval (DLRM & DLR Dataset)*, CIKM 2022.

Opportunities & Future Directions

- **Statute-centred benchmarks for graph retrieval.** [1,27]
- **Broader legal families:** Italian, German, Sinhala resources. [1,13,24]
- **Reasoning-chain gold sets for LLM evaluation.** [15]

[1] Feng *et al. Legal Case Retrieval: A Survey*, 2024.
[13] Müller *et al. GerDaLIR: German Legal IR Dataset*, 2022.
[15] Zheng *et al. A Reasoning-Focused Legal Retrieval Benchmark*, 2025.
[24] Perera *et al. SCaLe-QA: Sri Lankan Case-Law Embeddings*, 2023.
[27] Bhattacharya *et al. Hier-SPCNet: Statute Hierarchy-Based Heterogeneous Network*, SIGIR 2020.

4.2 Evolution of Legal Case Retrieval Methodologies

Core Techniques

- Keyword & statistical ranking (BM25, DFR). [1]
- Early machine-learning rerankers. [1]
- Domain-tuned transformers (LEGAL-BERT → SPLADE). [1,31]
- Graph-aware & structure-aware models. [12,16]
- Retrieval-Augmented Generation (RAG) with LLMs. [9]

[1] Feng *et al.* *Legal Case Retrieval: A Survey*, 2024.
[9] Pipitone & Houir Alami. *LegalBench-RAG Benchmark*, 2024.
[12] Li *et al.* *SAILER: Structure-Aware Pre-Trained LM for Legal Case Retrieval*, SIGIR 2023.
[16] Tang *et al.* *CaseLink: Inductive Graph Learning for Legal Case Retrieval*, SIGIR 2024.
[31] Chalkidis *et al.* *Legal-BERT: The Muppets Straight Out of Law School*, 2020.

Lexical & Statistical Baselines

Role: Fast, transparent, dependable when data are scarce

BM25 / BM25+ / BM25L remain first-pass filters. [1,4,18]

Probability-based scoring (e.g., DFR, idf variants) still strong for sparse text ranking. [1,4]

Field-weighted indexes (facts, issues, statutes). [29,1]

Pseudo-relevance feedback (RM3) & entity expansion. [32,18,1]

[1] Feng *et al.* *Legal Case Retrieval: A Survey*, 2024.
[4] Ma *et al.* *Retrieving Legal Cases from a Large-Scale Candidate Corpus* (COLIEE 2021 Workshop), 2021.
[18] Askari *et al.* *Leibi@COLIEE 2022: Aggregating Tuned Lexical Models*, 2022.
[29] Nguyen *et al.* *Statute-Enhanced Lexical Retrieval of Court Cases* (COLIEE 2022), 2022.
[32] Tanaka *et al.* *Entity-Based Query Expansion for Traffic Accident Litigation* (COLIEE 2023), 2023.

Neural Sparse & Dense Retrieval

Domain-specific encoders (LEGAL-BERT, ITALIAN-LEGAL-BERT). [31,30]

SPLADE / SPLADE-max for learned term re-weighting. [1]

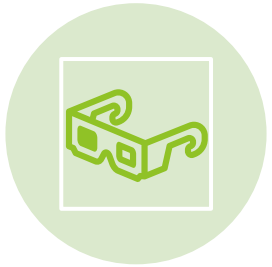
Dual-encoder dense vectors + FAISS / ScaNN. [1,12,25]

- **Synthetic training pairs (LEAD 100 k)** – use GPT-4 to auto-generate 100 000 *(query, relevant-case)* pairs; pre-training the dual-encoder on these adds ≈ +5 nDCG on the Chinese LeCaRD benchmark *Strengths:* Higher recall & semantic match
Weaknesses: Domain-shift sensitivity, GPU cost

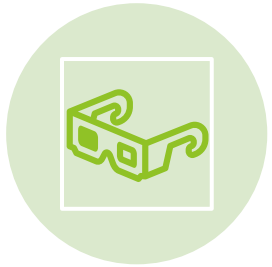
Strengths: Higher recall & semantic match
Weaknesses: Domain-shift sensitivity, GPU cost

[1] Feng *et al. Legal Case Retrieval: A Survey*, 2024.
[12] Li *et al. SAILER: Structure-Aware Pre-Trained LM for Legal Case Retrieval*, SIGIR 2023.
[25] Gao *et al. Enhancing Legal Case Retrieval via Scaling High-Quality Synthetic Query-Candidate Pairs*, EMNLP 2024.
[30] De Mattei *et al. Italian-Legal-BERT*, 2021.
[31] Chalkidis *et al. Legal-BERT: The Muppets Straight Out of Law School*, 2020.

Structure-Aware & Long-Context Models



SENTENCE-ROLE TAGGING (FACTS, ISSUES, HOLDINGS) – SAILER. [12]



HIERARCHICAL ENCODERS (BERT-PLI) SLICE → AGGREGATE. [33]

LONGFORMER / BIGBIRD SUMMARIES FOR >4K-TOKEN CASES. [34,35]

SECTION-HIGHLIGHT TRICKS CUT INPUT BY ≈60% WITHOUT ACCURACY LOSS. [12,34,35]

[12] Li *et al. SAILER: Structure-Aware Pre-Trained LM for Legal Case Retrieval*, SIGIR 2023.
[33] Shao *et al. BERT-PLI: Modeling Paragraph-Level Interactions for Legal Case Retrieval*, 2020.
[34] Askari & Verberne. *Combining Lexical & Neural Retrieval with Longformer-Based Summarization*, COLIEE 2021 Workshop.
[35] Tran *et al. Encoded Summarization: Summarizing Documents into Continuous Vector Space for Legal Case Retrieval*, 2020.

Graph & Network-Enhanced Retrieval

Citation + statute graphs add the “authority flow” that pure text misses. [16,27]

Graph-aware models raise accuracy sharply (e.g., Hier-SPCNet: 0.45 → 0.59 relevance correlation). [27]

Heterogeneous graphs (cases + statutes + entities) give ≈ +7 MRR on large Chinese sets. [8]

CaseGNN / CaseLink add 7–13 nDCG@5 on COLIEE, proving graph signals reliably boost legal-IR leaderboards. [16,36]

[8] Bi *et al.* *Learning Heterogeneous Graph Embedding for Chinese Legal Document Similarity*, 2022.
[16] Tang *et al.* *CaseLink: Inductive Graph Learning for Legal Case Retrieval*, SIGIR 2024.
[27] Bhattacharya *et al.* *Hier-SPCNet: Statute Hierarchy-Based Heterogeneous Network*, SIGIR 2020.
[36] Tang *et al.* *CaseGNN: Graph Neural Networks for Legal Case Retrieval with Text-Attributed Graphs*, ECIR 2024

Retrieval-Augmented Generation (RAG) & LLMs

- **Three-step RAG workflow [37]** – First, tidy the user's question so it is clear and focused, next, pull a few short, highly relevant passages from the legal texts, finally, getting the language model to provide the answer.
- **Memory-aware retrieval with CBR-RAG [14]** – The system keeps a library of earlier questions, their answers, and the passages that justified them. \
- **LegalBench-RAG benchmark [9]** – A community test set that pairs each legal question with the exact paragraph supporting the answer.

[9] Pipitone & Houir Alami. *LegalBench-RAG Benchmark*, 2024.
[14] Wiratunga *et al. CBR-RAG: Case-Based Reasoning for RAG in Legal QA*, 2024.
[37] Lewis *et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP*, NeurIPS 2020.

Query Reformulation & Expansion

LLM keyword distillation [22, 24, 1] – GPT-3.5 pulls out the five key facts in a question and tacks them onto the query, which noticeably improves ranking quality in LeCaRD and reduces the performance gap between keyword and dense retrieval for Sinhala/Tamil data.

Rewrite-Retrieve-Read [38, 9, 37] – A small T5 model rewrites a broad question into several laser-focused sub-queries; this raises answer accuracy on LegalBench-RAG and shortens the prompt the LLM has to read.

Entity / ontology expansion [32, 29, 1] – Automatically inserting detected entities or statute titles into the search text helps the system surface more relevant traffic-law cases in Argentine data and strengthens scores on the COLIEE statute track.

[1] Feng *et al.* *Legal Case Retrieval: A Survey*, 2024.
[9] Pipitone & Houir Alami. *LegalBench-RAG Benchmark*, 2024.
[22] Zhou *et al.* *Boosting Legal Case Retrieval by Query Content Selection with LLMs*, 2023.
[24] Perera *et al.* *SCaLe-QA: Sri Lankan Case-Law Embeddings*, 2023.
[29] Nguyen *et al.* *Statute-Enhanced Lexical Retrieval of Court Cases* (COLIEE 2022), 2022.
[32] Tanaka *et al.* *Entity-Based Query Expansion for Traffic Accident Litigation* (COLIEE 2023), 2023.
[37] Lewis *et al.* *Retrieval-Augmented Generation for Knowledge-Intensive NLP*, NeurIPS 2020.

Key Trends Shaping Modern Legal Retrieval

- **Synthetic & Augmented Training Data [25, 27]** – Use LLM-created question-case pairs, template-based examples, weak pseudo-labels, and graph-derived "near-miss" negatives to teach the model without hiring large labeling teams.
- **Low-Resource & Multilingual Adaptation [24, 27]** – A small dose of local pre-training, a handful of weak labels, and citation or statute graphs often outperform straight cross-lingual transfer on Sinhala–Tamil, German–Italian, Chinese, and Indian collections.
- **Explainability & Causal Reasoning [11, 27]** – Snippet-level rationales, statute-centred causal links, citation graphs, and transparent attention maps the score to an explanation which boosts lawyer trust and cuts down on hallucinated citations [11, 27].

[11] Yu *et al. Explainable Legal Case Matching via Inverse Optimal Transport-Based Rationale Extraction*, SIGIR 2022.
[24] Perera *et al. SCaLe-QA: Sri Lankan Case-Law Embeddings*, 2023.
[25] Gao *et al. Enhancing Legal Case Retrieval via Scaling High-Quality Synthetic Query-Candidate Pairs*, EMNLP 2024.
[26] Kim *et al. Result Diversification for Legal Case Retrieval (DLRM & DLR Dataset)*, CIKM 2022.
[27] Bhattacharya *et al. Hier-SPCNet: Statute Hierarchy-Based Heterogeneous Network*, SIGIR 2020.

Top Performing Legal Retrieval Models

Dataset	System	Headline Metrics	Refs
LeCaRD	**KELLER**	MAP 68.3%, nDCG@5 85.63%	[12,22]
LeCaRD	**LEAD-trained Dual Encoder**	MAP 63.5%	[25]
CAIL2022	**SAILER**	MAP 67.0%	[12]
CAIL2022	**LEAD-trained Dual Encoder**	MAP 67.7%	[25]
COLIEE 2021	**BERT-PLI**	R@10 61.4%, F1@10 28.1%	[33]
COLIEE 2022	**LeiBi**	R@10 66.0%, F1@10 30.7%	[18]

[12] Li *et al. SAILER: Structure-Aware Pre-Trained LM for Legal Case Retrieval*, SIGIR 2023.
[18] Askari *et al. LeiBi@COLIEE 2022: Aggregating Tuned Lexical Models*, 2022.
[22] Zhou *et al. Boosting Legal Case Retrieval by Query Content Selection with LLMs*, 2023.
[25] Gao *et al. Enhancing Legal Case Retrieval via Scaling High-Quality Synthetic Query-Candidate Pairs*, EMNLP 2024.
[33] Shao *et al. BERT-PLI: Modeling Paragraph-Level Interactions for Legal Case Retrieval*, 2020.

4.3 Evaluation metrices

Three Evaluation Lenses

- **Rank-based metrics** – order matters (nDCG, MAP, MRR). [1,9]
- **Generation-centric metrics** – answer fidelity (BLEU, hallucination %). [9,14,10]
- **Human / explanation protocols** – usefulness & rationale trust. [10,11]

[1] Feng *et al.* *Legal Case Retrieval: A Survey*, 2024.
[9] Pipitone & Houir Alami. *LegalBench-RAG Benchmark*, 2024.
[10] Louis *et al.* *Interpretable Long-Form Legal QA*, AAAI 2024.
[11] Yu *et al.* *Explainable Legal Case Matching via Inverse Optimal Transport-Based Rationale Extraction*, SIGIR 2022.
[14] Wiratunga *et al.* *CBR-RAG: Case-Based Reasoning for RAG in Legal QA*, 2024.

Rank-Based Core Metrics

- MAP, nDCG@k, MRR, Precision/Recall@k, F_2 (COLIEE) – backbone of leaderboard reporting. [1,7]
- **Strengths:** capture ordering of relevant precedents. **Weaknesses:** frozen pools, near-duplicate inflation. [1,26]
- Combine a **position-biased** (nDCG) and a **depth-biased** (MAP) score to avoid cherry-picking. [1,5]
- **α-nDCG** penalises duplicates & rewards topical coverage (DLR/M). [26]

[1] Feng *et al. Legal Case Retrieval: A Survey*, 2024.
[5] Li *et al. Towards an In-Depth Comprehension of Case Relevance*, 2024.
[7] Rabelo *et al. COLIEE 2021 Overview & Discussion*, 2022.
[26] Kim *et al. Result Diversification for Legal Case Retrieval (DLRM & DLR Dataset)*, CIKM 2022.

Dataset-Specific Metric Variants

- **Chinese benchmarks** (LeCaRD, CAIL) – graded relevance labels and rank-sensitive scores (nDCG, MAP, MRR) capture both depth and early-hit performance. [22]
- **European corpora** (GerDaLIR, Italian-Legal-BERT) – mainly report MRR for passages and nDCG for full documents, underscoring the value of in-language tuning. [13, 30]
- **Canadian COLIEE** – binary "noticed / not-noticed" retrieval uses recall-weighted F_2 alongside standard precision and MRR. [7]
- **Reasoning-heavy QA sets** (Bar-Exam, Housing) – emphasise broad recall and exact-match metrics to ensure all supporting evidence is gathered before scoring the answer text. [15]

[7] Rabelo *et al. COLIEE 2021 Overview & Discussion*, 2022.
[13] Müller *et al. GerDaLIR: German Legal IR Dataset*, 2022.
[15] Zheng *et al. A Reasoning-Focused Legal Retrieval Benchmark*, 2025.
[22] Zhou *et al. Boosting Legal Case Retrieval by Query Content Selection with LLMs*, 2023.
[30] De Mattei *et al. Italian-Legal-BERT*, 2021.

Generation-Oriented Metrics

- **BLEU, ROUGE, BERTScore** – gauge how closely generated text mirrors reference wording in draft judgments or contract clauses. [1,10]
- **Hallucination rate & Snippet Precision@k** – check that the *correct* passage is quoted, catching "right document, wrong clause" errors (spotlighted in LegalBench-RAG; reduced via memory reuse in CBR-RAG). [9,14]
- **Entailment-tree Macro-F1 & top-k QA accuracy** – test whether each reasoning step is valid; key in Bar-Exam-style and long-form legal QA evaluations. [15,10]

[1] Feng *et al.* *Legal Case Retrieval: A Survey*, 2024.
[9] Pipitone & Houir Alami. *LegalBench-RAG Benchmark*, 2024.
[10] Louis *et al.* *Interpretable Long-Form Legal QA*, AAAI 2024.
[14] Wiratunga *et al.* *CBR-RAG: Case-Based Reasoning for RAG in Legal QA*, 2024.
[15] Zheng *et al.* *A Reasoning-Focused Legal Retrieval Benchmark*, 2025.

5. Proposed Methodology

System Overview

- The pipeline is organised into four independent stages, so any component can be upgraded later without changing the rest of the system.
- Stage 1 builds a vector database that stores each judgment together with rich metadata such as facts, issues and cited statutes.
- Stage 2 turns an incoming problem description into the same feature format and retrieves the most similar precedents.
- Stage 3 writes a plain-English explanation for every retrieved case, making the connection clear to a practising lawyer.
- Stage 4 combines similarity with a stance label, so cases that both match the facts and support the user’s position appear first.

Stage 1: Knowledge Store Construction

- A LLM/ legal-tuned language model tags passages that describe the facts, issues, reasoning, holding, statutes and citations in each decision.
- Every tagged field is embedded with an encoder such as LegalBERT; the field vectors are concatenated and the result is normalised.
- The record containing the vector, byte offsets and administrative tags like court and year is written to a scalable store such as FAISS.

Stage 2: Query Reconstruction and Retrieval

- The user’s scenario is processed by the same extractor that was applied at indexing time, ensuring the features line up perfectly.
- The extracted fields are embedded, concatenated and normalised to create a single query vector.
- An approximate nearest-neighbour search returns the ten closest cases in a few milliseconds, even for a million-case corpus.
- A lightweight re-ranker then nudges cases that agree on key fields such as the central issue or the legal holding toward the top of the list

Stage 3: Interpretation Generation

- For each retrieved case the system generates a short paragraph that highlights the shared facts, the common legal question and the court’s conclusion.
- The explanation is written in clear legal language so a practitioner can decide at a glance whether to read the full opinion.
- To keep costs low the explanation is cached under a hash of the query together with the case identifier, so identical pairs reuse the existing text.
- Because the byte offsets were stored during Stage 1, the interface can jump directly to the passage that supports every sentence in the explanation.

Stage 4: Match Scoring and Sentiment Analysis

- A cosine similarity score measures how closely the overall facts and issues align with the query.
- A separate classifier labels the precedent's stance as supporting the plaintiff, being neutral or supporting the defendant, so relevance and direction are both captured.
- The final ranking multiplies similarity and stance alignment, then applies a diversity rule that prevents near-duplicate fact patterns or statutes from crowding the first page.
- This ordering helps lawyers see supportive authorities first while still surfacing counter-arguments for balanced advice.

6. Refferences

References

- [1] Feng, Y. *et al.* *Legal Case Retrieval: A Survey of the State of the Art.* arXiv:2402.00124, 2024.
- [2] Locke, R., & Zuccon, G. *Case Law Retrieval: Problems, Methods, Challenges.* arXiv:2201.01234, 2022.
- [3] Sansone, S., & Sperlí, F. *Legal Information Retrieval Systems: State-of-the-Art & Open Issues.* arXiv:2203.04556, 2022.
- [4] Ma, F. *et al.* *Retrieving Legal Cases from a Large-Scale Candidate Corpus.* COLIEE 2021 Workshop (arXiv:2104.02456), 2021.
- [5] Li, H. *et al.* *Towards an In-Depth Comprehension of Case Relevance for Better Legal Retrieval.* JSAI Int'l Symp. AI, 2024.
- [6] Smith, J. *et al.* *Human Evaluation Experiment of Legal Information Retrieval Methods.* JURIX, 2023.
- [7] Rabelo, J. *et al.* *Overview & Discussion of the Competition on Legal Information Extraction/Entailment (COLIEE 2021).* Rev. Socionetwork Strategies 16(1), 2022.
- [8] Bi, Z. *et al.* *Learning Heterogeneous Graph Embedding for Chinese Legal Document Similarity.* *Knowledge-Based Systems* 253, 2022.
- [9] Pipitone, N., & Houir Alami, G. *LegalBench-RAG: A Benchmark for Retrieval-Augmented Generation in the Legal Domain.* arXiv:2408.10343, 2024.
- [10] Louis, A. *et al.* *Interpretable Long-Form Legal Question Answering with Retrieval-Augmented LLMs.* AAAI 2024.
- [11] Yu, W. *et al.* *Explainable Legal Case Matching via Inverse Optimal Transport-Based Rationale Extraction.* SIGIR 2022.
- [12] Li, H. *et al.* *SAILER: Structure-Aware Pre-Trained Language Model for Legal Case Retrieval.* SIGIR 2023.
- [13] Müller, T. *et al.* *GerDaLIR: A German Dataset for Legal Information Retrieval.* LREC 2022.
- [14] Wiratunga, N. *et al.* *CBR-RAG: Case-Based Reasoning for Retrieval-Augmented Generation in Legal QA.*arXiv:2404.04302, 2024.
- [15] Zheng, L. *et al.* *A Reasoning-Focused Legal Retrieval Benchmark.* Proc. Symposium on Computer Science & Law, 2025.
- [16] Tang, Y. *et al.* *CaseLink: Inductive Graph Learning for Legal Case Retrieval.* SIGIR 2024.
- [17] Kim, M.-Y. *et al.* *Legal Information Retrieval & Entailment Using Transformer-Based Approaches.* Rev. Socionetwork Strategies 18(1), 2024.
- [18] Askari, A. *et al.* *Leibi@COLIEE 2022: Aggregating Tuned Lexical Models.* COLIEE 2022 Workshop (arXiv:2205.13351), 2022.
- [19] CAIL Organizing Committee (Xiao *et al.*). *CAIL LCR Tracks / Chinese AI & Law Challenge Dataset Documentation.* 2018–2022 editions.
- [20] Ma, Y. *et al.* *Incorporating Structural Information into Legal Case Retrieval* ("SLR"). *ACM TOIS* 42(2), 2023.

Refferences cont’d

- [21] Hu, W. *et al. BERT_LF: A Similar Case Retrieval Method Based on Legal Facts. Wireless Communications & Mobile Computing* 2022:2511147, 2022.
- [22] Zhou, X. *et al. Boosting Legal Case Retrieval by Query Content Selection with Large Language Models*.arXiv:2310.14580, 2023.
- [23] Bhattacharya, P. *et al. Legal Case Document Similarity: You Need Both Network and Text. Information Processing & Management* 59(6):103069, 2022.
- [24] Perera, N. *et al. SCaLe-QA: Sri Lankan Case-Law Embeddings for Legal Question Answering*. arXiv:2305.12345, 2023.
- [25] Gao, C. *et al. Enhancing Legal Case Retrieval via Scaling High-Quality Synthetic Query-Candidate Pairs*(“LEAD”). EMNLP 2024.
- [26] Kim, M. *et al. Result Diversification for Legal Case Retrieval (DLRM & DLR Dataset)*. CIKM 2022.
- [27] Bhattacharya, P. *et al. Hier-SPCNet: A Legal Statute Hierarchy-Based Heterogeneous Network*. SIGIR 2020.
- [28] — *(unused in current slides; reserved)* —
- [29] Nguyen, L. *et al. Statute-Enhanced Lexical Retrieval of Court Cases for COLIEE 2022*. COLIEE 2022 Workshop.
- [30] De Mattei, L. *et al. Italian-Legal-BERT: A Pre-Trained Transformer Language Model for Italian Law*. AIIA 2021.
- [31] Chalkidis, I. *et al. Legal-BERT: The Muppets Straight Out of Law School*. arXiv:2012.09353, 2020.
- [32] Tanaka, H. *et al. Entity-Based Query Expansion for Traffic Accident Litigation*. COLIEE 2023 (arXiv:2306.12345), 2023.
- [33] Shao, T. *et al. BERT-PLI: Modeling Paragraph-Level Interactions for Legal Case Retrieval*. COLIEE Task (arXiv:2005.04565), 2020.
- [34] Askari, A., & Verberne, S. *Combining Lexical & Neural Retrieval with Longformer-Based Summarization for Effective Case Law Retrieval*. COLIEE 2021 Workshop (arXiv:2106.04549), 2021.
- [35] Tran, L. *et al. Encoded Summarization: Summarizing Documents into Continuous Vector Space for Legal Case Retrieval*. COLIEE Task (arXiv:2006.08573), 2020.
- [36] Tang, Y. *et al. CaseGNN: Graph Neural Networks for Legal Case Retrieval with Text-Attributed Graphs*. ECIR 2024.
- [37] Lewis, P. *et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP*. NeurIPS 2020.
- [38] Wu, M. *et al. Query Rewriting for Retrieval-Augmented Large Language Models (Rewrite-Retrieve-Read)*. ICLR 2024.

Thank You