CASE BRIDGE

An AI-powered legal research system for precise case retrieval and concise, citation-backed brief generation.

B.CHARAN REDDY G474

Scope of the Project:

- Accepts a **legal research question** as input.
- Retrieves the **top 5 most relevant cases** from a large legal corpus.
- Generates a **concise analysis brief** grounded in those cases.
- Uses **IR (Information Retrieval)** + **RAG (Retrieval-Augmented Generation)** techniques inspired by CLERC.

Understanding Of The Paper

CLERC dataset provides:

- Real U.S. federal case law from CAP.
- Pre-processed for retrieval and generation tasks.

Two key contributions relevant to our tool:

1. **Retrieval task** – Finding the most relevant precedent for a given query.
2. **RAG task** – Generating factual legal analyses using retrieved citations.

Benchmarks show **fine-tuning on legal data significantly improves performance**.

Dataset Structure:

- **CLERC/doc** – Full case documents.
- **CLERC/passage** – Case documents split into 350-word passages with overlapping windows.
- **CLERC/generation** – Passages for testing legal analysis generation.
- Over **20.7M total citations** and **23.7M retrieval passages**.

Business Problems Addressed

Time-Consuming Legal Research

- Lawyers and researchers spend hours manually searching through large legal databases to find relevant precedents.

Keyword-Only Search Limitations

- Traditional legal search tools rely heavily on keywords, often missing semantically relevant cases.

Information Overload

- Large volumes of retrieved cases make it hard to identify the most relevant precedents quickly.

Manual Brief Preparation

- Writing legal briefs and inserting citations is tedious and prone to human error.

Risk of AI Hallucinations

- General AI tools may generate inaccurate or unsupported claims without proper grounding in actual legal texts..

Proposed Idea

Product Name: *Case Bridge*
Key Features:-
Search by natural language legal query:

- Supports jurisdiction-specific queries (U.S. & India).

Hybrid Retrieval:

- Uses multiple retrieval models (Semantic and lexical) for comprehensive legal coverage.

Retrieve 5 most relevant cases:

- Fine-tuned retrievers identify top matches, followed by **Context-Aware Processing** for jurisdiction relevance re-ranking.

Generate analysis briefs with citations:

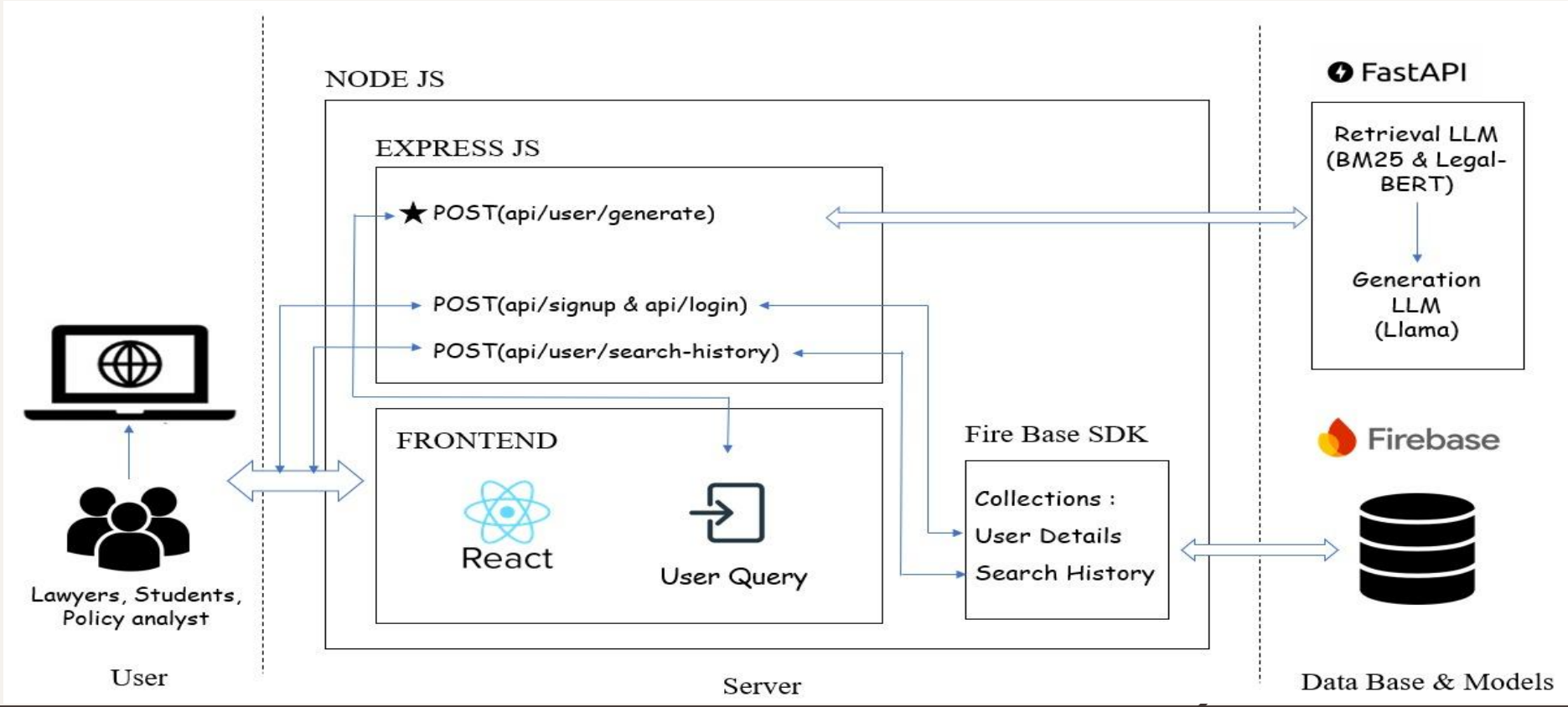
- **Brief Generation** via LLM produces concise, verified, and jurisdiction-aware legal briefs, scalable to other domains.

Minimize hallucinations:

- Grounding outputs in retrieved case text ensures factual accuracy.

End Goal: A deployable, user-friendly **web interface** for legal professionals & researchers.

Architecture Diagram:



Research Paper Idea

Case Bridge, an AI-powered legal research system that answers natural language queries by retrieving the **top 5 most relevant precedent cases** and generating **concise, citation-rich summaries**.

Our approach leverages a **hybrid retrieval architecture**:

- **BM25** for precise matching of legal terms and citations.
- **LegalBERT-DPR** for semantic understanding and legal reasoning.
- **GeneralBERT fine-tuned on CLERC** for robust handling of layperson and generalized queries.

The outputs from these models are **fused** and **jurisdiction-aware re-ranked** using a cross-encoder before passing to an **LLM** for fact-grounded summarization.

To ensure **global applicability**, we combine **U.S. legal data (via CLERC)** with a **curated Indian legal dataset**, enabling high-accuracy retrieval for both jurisdictions. This design uniquely balances **expert precision** with **broad accessibility**, serving both legal professionals and non-experts.

Thank You