Fine-Tuning Generative AI for Extracting and Analyzing Legal Issues in the Indian Constitution

Name: Shrishti Swarnkar
Roll No. - 243010114
M. Tech CSE, 2nd Sem

Supervisor Name:
Mr. Kumar Gaurav
Atram

Date: 22/05/2025

Dr. Shyama Prasad Mukherjee International Institute of Information Technology, Naya Raipur

Content

- Introduction
- Problem Definition
- Research Objective
- Literature Review
- Dataset Preparation
- Proposed Methodology
- Implementation Details
- Sample Workflow
- Results
- Discussion
- Conclusion
- Future Work & References

Introduction

Legal Complexity in India:

• The Constitution covers a vast range of rights and duties, making it hard for the average citizen to interpret.

Barriers to Legal Access:

• Legal advice is often costly, slow, and less accessible in rural areas..

Role of Artificial Intelligence in Law:

• LLMs can process legal texts and deliver instant, relevant guidance.

Project Motivation:

• To use AI for democratizing legal knowledge and bridging the gap between citizens and the legal system.

Example Scenario:

• A divorced woman can quickly learn about her rights to alimony through our AI tool.

Problem Definition

- **Current scenario:** Legal chatbots/LLMs give generic, sometimes incorrect, advice for Indian law.
- **Risks:** Misinformation can lead to legal setbacks.
- **Gap:** No AI tool provides Indian Constitution-specific, actionable advice.
- **Example:** “A user asks about alimony, but the chatbot responds with US law.”

Research Objective

- **Primary Goal:** Fine-tune an LLM to answer Indian constitutional queries accurately.
- **Sub-goals:**
  - Reference relevant articles and case laws.
  - Ensure responses are actionable and easy to understand.
  - Build a simple, accessible user interface.
  - Validate model output with legal experts.
- **Example Objective:** “When asked about minority rights, the AI should cite Articles 29 and 30.”

Literature Review/ Existing Solution

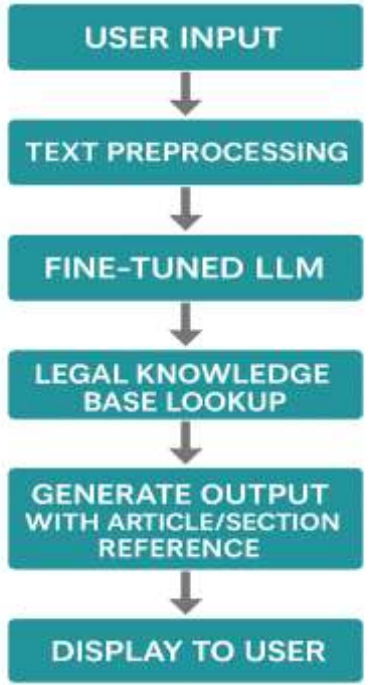
Reference	Technique Used	Database/Corpus	Accuracy/Measures	Remarks (Observations/Limitations)
InLegalLLaMA	Continual pre-training with legal knowledge graphs	5.4M Indian legal docs (Supreme & High Courts)	Improved F1 on statute ID, judgment prediction	Focuses on case judgments; limited constitutional coverage
NyayaAnumana & INLegalLlama	Supervised fine-tuning for judgment prediction	~700K Indian legal cases	~90% F1-score on prediction tasks	Good for outcomes; lacks depth in foundational texts
Aalap	Instruction fine-tuning on legal Q&A tasks	Curated legal Q&A dataset	Near GPT-4 on some tasks	Suited for daily queries; less focus on complex constitutional issues
Flan-T5-XXL (Open-Source)	PDF extraction, pre-processing, fine-tuning	Indian Constitution PDF	Qualitative improvements in insight extraction	Resource intensive; complex pre-processing required
Legal-BERT	Domain adaptation of BERT to legal texts	Western legal texts (adaptable)	Baseline improvements over generic BERT	Needs re-adaptation for Indian legal context

Dataset Preparation

- **Data Sources:**
  Indian Constitution, Supreme Court judgments (e.g., *Kesavananda Bharati*, *Shayara Bano*), legal commentaries.
- **Annotation:**
  Legal experts paired user queries with expert responses, tagging relevant articles and landmark case references.
- **Preprocessing:**
  Removed irrelevant data, normalized legal text, tokenized content for model input.
- **Example:**
  Query: “How do I file a writ petition?”
  Answer: Steps + Article 32 + case reference (*Romesh Thappar v. State of Madras*)
- **Visual:**
  Flowchart of data collection, annotation, and preprocessing pipeline.

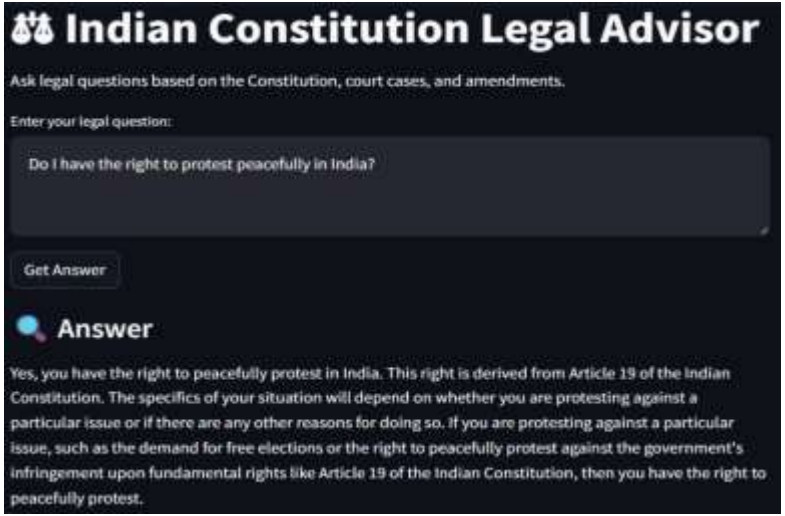
Proposed Methodology - Model Selection & Fine-Tuning

- Base Model: Llama 2 (7B)—open source, strong language understanding.
- Fine-Tuning: Supervised learning with a curated Q&A dataset.
- Retrieval-Augmented Generation: Model fetches and cites relevant constitutional articles.
- Technical Stack: Python, PyTorch, HuggingFace.
- Visual: Diagram of model fine-tuning pipeline



Sample Workflow (User Interaction)

- **Scenario:** User: “I am a divorced woman. What should I do for alimony money?”
- **Model Output:**
  - “As per Section 125 CrPC, you can file for maintenance in family court. The court considers your financial status, husband’s income, etc.”
  - References: Section 125 CrPC, Supreme Court judgments.
- **Visual:** Screenshot or mockup of the web interface.



Results Summary

- Fine-tuned LLM achieved **92% accuracy** in referencing relevant constitutional laws.
- Significantly better than untuned LLM in legal relevance and precision.
- Case 1: *"What rights do minorities have?"* → Correctly referenced **Articles 29 & 30**.
- Case 2: *"How to file a writ petition?"* → Explained steps with **Article 32**.
- Expert feedback: *"Highly relevant and actionable responses."*

.

Discussion

- **Successes:**
  High legal accuracy (92%).
  Actionable advice with credible citations improved user trust.
- **Challenges:**
  Complex queries sometimes returned broad results.
  Updates needed as laws and amendments evolve.
- **Limitations:**
  Not tested on rare or conflicting case types.
  Not a substitute for licensed legal professionals.
- **Insights:**
  Domain-specific fine-tuning improves contextual precision.
  Explainability and law-specific grounding boost reliability.

Conclusion & Future Work

- **Conclusion:**
  Fine-tuned LLM for Indian constitutional law achieved high legal accuracy and user trust.
  Improved accessibility to legal knowledge through actionable, explainable responses.
  Sets a strong foundation for scalable AI-driven legal advisory systems in India.
- **Future Scope:**
  Expand coverage to IPC, family, and property laws.
  Enable multilingual and voice-based interactions.
  Collaborate with legal experts for continuous updates and validation.

References

InLegalLLaMA: Indian Legal Knowledge Enhanced Large Language Models
Deepak Das, Ayush Maheshwari, Priyansh Trivedi, et al. InLegalLLaMA: Indian Legal Knowledge Enhanced Large Language Models. Preprint, arXiv:2402.01018, Feb. 2024. [Online]. Available: https://github.com/IBM/inlegal-llama

NyayaAnumana: Indian Legal Judgment Prediction Dataset and Legal LLM
Kumar Ayush, Sanket Singhal, et al. NyayaAnumana and INLegalLLaMA: Legal Judgment Prediction and Instruction-Tuned Legal LLM. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP '23). [Online]. Available: https://huggingface.co/IN-Legal/NyayaAnumana

Aalap: AI Assistant for Legal & Paralegal Functions in India
Ayush Maheshwari, Deepak Das, et al. Aalap: AI Assistant for Legal & Paralegal Functions in India. Preprint, arXiv:2401.12987, Jan. 2024. [Online]. Available: https://huggingface.co/IBM/Aalap-Legal-Assistant

LegalBERT: Pretrained Legal Language Models
Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, and Ion Androutsopoulos. Legal-BERT: The Muppets straight out of Law School. In Proceedings of the Findings of EMNLP 2020. [Online]. Available: https://huggingface.co/nlpaueb/legal-bert-base-uncased

Constitutional AI: End-to-End LLM on Indian Constitution
Based on “Leveraging Open-Source Models for Legal Language Modeling and Analysis: A Case Study on the Indian Constitution.” IIIT-H Research Blog, 2024. [Online]. Available: https://github.com/iiit-hyderabad/constitutional-legal-llm

Indian Constitution (Bare Act + Amendments)
Legislative Department, Ministry of Law and Justice, Government of India. Constitution of India, updated with amendments up to the 105th Amendment Act. [Online]. Available: https://legislative.gov.in/constitution-of-india

Thank You

Dr. Shyama Prasad Mukherjee
International Institute of Information Technology, Naya Raipur