SMALL LANGUAGE MODELS FOR SRI LANKAN LEGAL APPLICATIONS

25-26-J-240

Project Proposal Report

MATHUSIGAN E S

THUVARAGA A

NIRUTHTHIKA E

ABIRAMY T

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

July 2025

LAW RECOMMENDATION SYSTEM FOR USERS FOCUS ON EMPLOYEE / LABOR LAW

25-26-240

Project Proposal Report

NIRUTHTHIKA E IT22322326

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka July

2025

DECLARATION

I declare that this is our own work, and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or Institute of higher learning, and to the best of our knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgment is made in the text. Also, I hereby grant to the Sri Lanka Institute of Information Technology, the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other mediums. I retain the right to use this content in whole or part in future works (such as articles or books).

Name	Student ID	Signature
Niruthtthika E	IT22322326	

The supervisor/s should certify the proposal report with the following declaration. The above candidates are carrying out research for the undergraduate Dissertation under my Supervision.

_____ _____

Signature of the co -supervisor **Signature of the supervisor**

_____ _____

Date **Date**

ABSTRACT

In Sri Lanka, labor and employment law mainly talks about workplace problems, employee rights, and employer duties. Many people who face issues at work need to know the exact law, the section, and the year that match their case. Today, legal information is easier to find than before, but for most people who are not lawyers, it is still hard to pick the correct law for a specific problem.

This project tries to fix that by building a law recommendation tool for Sri Lankan labor and employment law. The idea is simple: a user can type their question in normal language, and the system will understand it. Then it will show the correct legal section and year. It will also give examples of other times the same law was used, so the person can see the bigger picture.

The system will use Natural Language Processing (NLP) together with a fine-tuned Transformer model, more exactly a Small Language Model. To make the answers more accurate, it will also use something called Retrieval-Augmented Generation (RAG). By putting these together, the tool can give answers that are both reliable and easier to understand.

The goal is to make legal knowledge more reachable for lawyers and also for normal people. It can save time in legal research, reduce extra work for lawyers, and help people take better legal decisions. Over time, it will also support a legal system that is fairer and more efficient in Sri Lanka.

Keywords — Law Recommendation System, Natural Language Processing (NLP), Transformer-based (T5) Small Language Model (SLM), Retrieval-Augmented Generation (RAG)

TABLE OF CONTENTS

DECLARATION i

ABSTRACT ii

LIST OF FIGURES v

LIST OF TABLES vi

LIST OF ABBREVIATIONS vii

1. INTRODUCTION 1

1.1. BACKGROUND AND LITERATURE SURVEY 2

1.2. RESEARCH GAP 9

1.2.1. IDENTIFIED RESEARCH GAPS 10

1.2.2. PROPOSED SOLUTION 12

2. RESEARCH PROBLEM 13

3. OBJECTIVES 15

3.1. MAIN OBJECTIVES 15

3.2. SPECIFIC OBJECTIVES 16

4. METHODOLOGY 17

4.1. SYSTEM ARCHITECTURE DIAGRAMS 21

4.2. COMPONENT SPECIFIC SYSTEM DIAGRAM 22

4.3. DATA COLLECTION 25

4.4. SOFTWARE SOLUTION 27

5. PROJECT REQUIREMENTS 29

5.1. FUNCTIONAL REQUIREMENTS 29

5.2. NONFUNCTIONAL REQUIREMENTS 30

5.3. EXPECTED TEST CASES 31

5.4. FEASIBILITY STUDY 34

6. TOOLS AND TECHNOLOGIES SELECTION 36

6.1. TOOLS 36

6.2. TECHNOLOGIES 37

6.3. IMPLEMENTATION 38

6.4. DEPLOYMENT 39

7. PERSONNEL AND FACILITIES 40

8. WORK BREAKDOWN STRUCTURE 42

9. GANTT CHART 43

10. BUDGET AND BUDGET JUSTIFICATION 44

11. COMMERCIALIZATION OF THE PRODUCT 45

12. CONCLUSION 47

REFERENCES 48

List of Figures

Figure 1:Hierarchy of Courts 3
Figure 2:: Survey of existing legal information 3
Figure 3: Survey of existing legal information 13
Figure 4:system architecture 21
Figure 5:Component specific diagram 22
Figure 6:Agile development life cycle 28
Figure 7:Work break down structure 42
Figure 8:GANTT Chart 43

List of tables

Table 1:Large Language Models (LLM) 4
Table 2:Hybrid Models (LLM + RAG / Embeddings / Extra Layers)..... 5
Table 3:Small language models (SLM) 6
Table 4:Comparative Analysis of Research Features 10
Table 5:Existing system comparison with our system..... 11
Table 6:Expected Test cases 31
Table 7:Budget..... 44

LIST OF ABBREVIATIONS

Abbreviations	Descriptions
ML	Machine Learning
NLP	Natural Language Processing
SLM	Small Language Model
RAG	Retrieval-Augmented Generation
T5	Text-to-Text Transfer Transformer
NER	Named Entity Recognition
API	Application Programming Interface
AI	Artificial Intelligence
RoI	Return on Investment
LoRA	Low-Rank Adaptation

1. INTRODUCTION

There are many different laws in Sri Lanka. Some laws pertain to property, family, labor, and criminal law, among other areas. They attempt to know all of them, though as a matter of fact, they are not easy to read. The terms are highly formal and legalistic. The legislation also is diffused in many sections. Therefore, ordinary individuals find them hard, and even lawyers find it difficult to take advantage of them at times.

With AI, and namely NLP, one can now train a small language model only for Sri Lankan law. The emphasis of this study is on how SLMs can help in real use. The four key concepts here are:

1. Optimizing an SLM to offer step-by-step end-to-end instruction to resolve the user. Family Law and Property Law issues in Sri Lanka.
2. Review and analysis of risks in Sri Lankan deed documents by fine-tuning an SLM.
3. Building a system of law recommendations to users, and in particular to the labor and employment law.
4. Predicting probable legal decisions based on past cases' information to build up an SLM, thereby enabling users to be familiar with applicable precedents and possibilities.

On my side, I am undertaking the labor and employment law recommendation system. The plan is simple: if a person types a query in plain words, the system displays the name of the law, section, and year, as well as other scenarios under that law. This will be a SLM with Sri Lankan laws fine-tuning. On top of that, I will include RAG, thus making the answers more accurate. This way, the system will save time and will make the work of lawyers less complex and will make average people aware of which law applies to their problem.

1.1. Background and Literature Survey

Too much legal information has been a problem for a long time, and not only in law but in other fields too. As demonstrated by Hemp [1], it becomes difficult when people encounter excessive information to take fast and quality decisions. The same applies to lawyers in Sri Lanka who are required to deal with a large number of laws, which are divided into portions. The Sri Lankan court was described by Tambimuttu [2] as a system, and he described its complexity through the fact that there is a multiplicity of legal traditions simultaneously. His work did not, however, discuss utilizing new computer techniques in order to ease the process of locating laws. The ministry of justice [3] also indicated that the management of legal documents should be in a more orderly way. Nevertheless, the majority of the processes they propose are to be manual, and not much is automated.

Wijesinghe [4] presented a good illustration of the organization of courts in Sri Lanka and how legal power travels between the tiers. However, once again, the research failed to describe how digital. Judgments could be made more easily accessible to people through the help of systems. At a later stage, Fernando [5] discussed the challenges that people face in locating legal information in Sri Lanka. He demonstrated how inadequate systems and lack of links prevent people from obtaining justice in a timely way. In a similar way, according to Jayawardena [6], legal literacy is quite significant. He stated that even the professionals (not just the general population) tend to struggle with navigating the legal system.

All of these studies point to the same issue: it is difficult to manage and access legal information. However, none of them proposes a suitable tool of AI that can be used to suggest laws, particularly in the field of labor and employment law in Sri Lanka. This leaves no doubt that an actual requirement exists regarding the need to have a system that employs NLP and RAG. Such legal information can be more accessible, more precise, and more equal to all people with the help of the system.

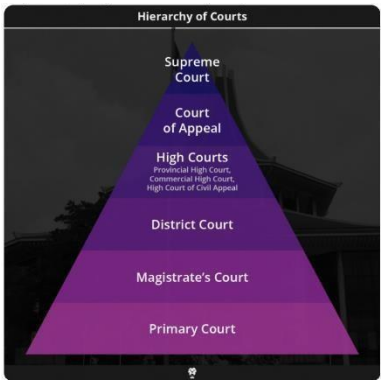

Figure 1:Hierarchy of Courts

We asked 40 people in Sri Lanka about legal info systems. The group had law students, some lawyers, and a few normal users too. The point was just to check if these systems really help them. Most were not happy. Around 1 in 5 people said the systems are kind of okay, not great but not the worst. More than half said they don't really work — they can't find the right laws, no proper guidance for real cases, and labor law is almost missing. Only a very small group, maybe 1 out of 10, said they were satisfied. So overall, people feel the current platforms are not enough for real legal research.

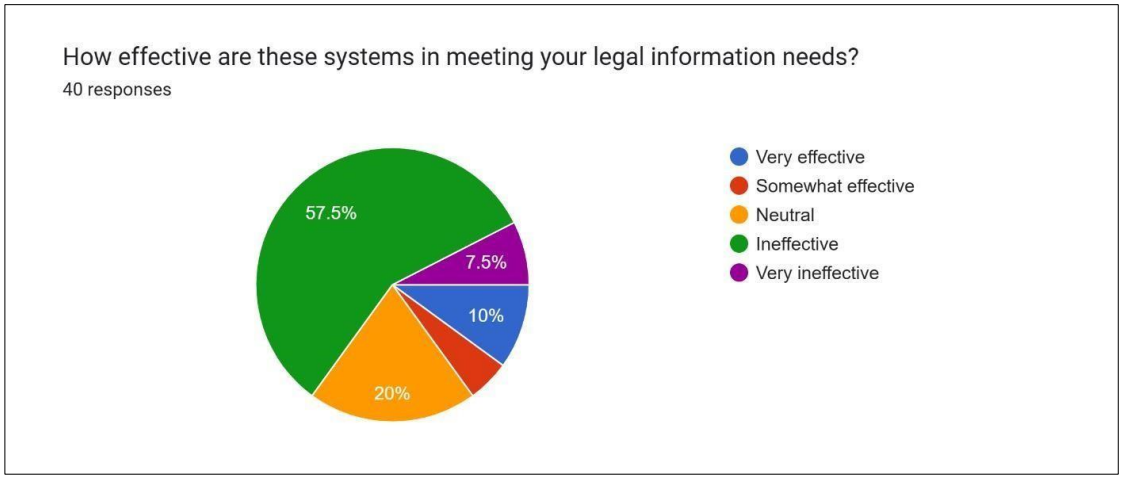

Figure 2:: Survey of existing legal information

These results highlight a significant research gap: current systems do not fully support user needs, particularly in delivering accessible, accurate, and scenario-based legal information.

Table 1:Large Language Models (LLM)

Paper	Technologies Used	Strengths	Limitations
Shu et al. (2024) – LawLLM	Large LLM fine-tuned on U.S. statutes + case law	Outperforms GPT4 baseline on U.S. tasks	U.S.-only, costly annotations
Sun et al. (2024) – Logic Rules for Legal Case Retrieval	Logic rules + embeddings (no finetuning)	Improves interpretability, explainability	Limited scalability, small datasets
Ma et al. (2023) – CaseEncoder	Knowledge enhanced encoder + pretrained embeddings	Outperforms BERT baselines, captures legal semantics	Still experimental, heavy infra

The other researchers, in turn, have concentrated on large language models (LLMs) that are fine-tuned to legal settings. The LawLLM [7] developed by Shu et al. is a domain-specific LLM trained on the U.S. statutes and case law and performed better than GPT-4 baselines in both reasoning and retrieval tasks. It is, however, less generalizable across the world because of its dependence on U.S. datasets and because of its annotation cost. The article by Sun et al. (2024) [8] investigated logic rule-based legal case retrieval, which integrates logical rule-based reasoning with embeddings to enhance the level of transparency and explainability. Although this method afforded easier interpretation of outputs, it did not scale well when using it for large problem datasets. CaseEncoder (Ma et al., 2023) [9], a knowledge-enhanced encoder using legal knowledge, was reported to be better than BERT baselines on a variety of retrieval benchmarks, but it was computationally expensive and still at the experimental phase.

Such studies show the strength of LLMs, which capture a broader semantics and reasoning ability. Nevertheless, they also point at such limitations as jurisdictional bias, high computational expenses and the inability to scale to a variety of legal systems.

Table 2:Hybrid Models (LLM + RAG / Embeddings / Extra Layers)

Paper	Technologies Used	Strengths	Limitations
Rahman et al. (2024) – LQ-RAG	RAG + fine-tuned embeddings + feedback loop	Reduces hallucinations, boosts relevance	Proprietary eval agent, no human loop
ICISS (2024) – AI Legal Companion	RAG + ChatGPT	Improves literacy, enhances access	Prone to misinformation, accuracy gaps
Magrani & Fernandes (2024) – Ethical Critique	Policy, human-rights analysis of AI RS	Strong ethical depth	No technical solutions
Thomas et al. (2021) – Quick Check (Thomson Reuters)	Ranking models + citation networks + annotations	Proven industry tool	Proprietary, no public benchmarks
Zeng et al. (2025) – RoSiLC-RS	LLM with abstraction layers	Improves conceptual similarity	Early-stage, no quantitative metrics
Mentzingen et al. (2025) – Precedent Retrieval	Summarization + LLM embeddings	Efficient retrieval, cost-effective	May lose nuance, dataset limited to Brazil
Su et al. (2023) – Caseformer	Pre-training for intercase distinctions	Robust, SOTA in zero/few-shot retrieval	Retrieval-only, no ranking/recommendation

Ethical criticisms of AI-based legal recommender systems were given by Magrani and Fernandes [7], considering fairness, transparency, and ethical aspects of human rights, giving the critiques an intense ethical focus but no concrete implementations. Following this, ICISS researchers [8] built an AI Legal Companion based on RAG and ChatGPT that made legal literacy more accessible and provided people with more access to legal information but had a weakness in misinformation and was susceptible to poor accuracy. Rahman et al. [10] suggested LQ-RAG, which integrates RAG with fine-tuned embeddings and a feedback loop to minimize hallucinations and enhance relevance, but it used a proprietary evaluation agent to reduce transparency, and a human-in-the-loop mechanism was not included. The tool Quick Check, nomenclature of Thomas et al. [11], is

a proprietary Thomson Reuters tool that utilizes ranking models, citation networks, and annotations; has been shown to be effective in industry application but does not have any publicly available benchmarks; and uses closed methodologies. Zeng et al. [12] introduced RoSiLC-RS, an LLM-based architecture that adds abstraction layers to the architecture to enhance similarity in the concepts but is at an early stage without quantitative validation. Mentzingen et al. [13] built a precedent retrieval system based on summarization and LLM embeddings, which proved cost-effective and efficient in terms of retrieval quality but may lack the detail and only support Brazilian legal data. Likewise, Su et al. [14] introduced Caseformer, a pre-trained inter-case model, which demonstrated state-of-the-art results on both zero- and few-shot retrieval, and Ma et al. [15] followed up by introducing CaseEncoder, a case retrieval model that uses external knowledge graphs to boost its performance.

Table 3:Small language models (SLM)

Paper	Technologies Used	Strengths	Limitations
Dhanani et al. (2021) – Dictionary-based LDRS	Doc2Vec, dictionary preprocessing	Lightweight, efficient retrieval	No transformers, noisy corpus, frequent dictionary updates
Dhanani et al. (2021) – Cluster-based LDRS	Doc2Vec + Louvain clustering	Scalable, reduced $O(n^2)$, good accuracy	Limited to India, no natural language queries
Dhanani et al. (2021) – P-LDRS	Pre-learned embeddings + Doc2Vec + Spark	Accuracy 0.88, scalable infra	Infra-dependent, still Doc2Vec-based
Zheng et al. (2022) – LawRec	BERT + Skip-RNN	High accuracy (92%), strong baseline	Not global, text-only
Zhou et al. (2018) – Legal Decision Rec.	TextCNN + BiGRU + Attention	Good results across tasks (crime, law, sentencing)	Limited outside China
Nithya et al. (2024) – AI-driven automation	NLP + ML pipeline	Automates summarization, drafting, Q&A	Still early, needs global validation

A number of works have studied the application of SLMs and lightweight. strategies of retrieving legal information. The dictionary-based Legal Document Recommendation System (LDRS) is recommended [16] by Dhanani et al., who operated on the idea of Doc2Vec, with the highest efficiency level but required frequently updated with dictionaries and was not as semantically rich as transformer-based models. In an effort to solve the scalability issue, Cluster-based LDRS [17] was written using Louvain clustering and was based on the Indian legal datasets only. They subsequently followed this up with P-LDRS [18], which added pre-trained embeddings and distributed processing on Spark, and obtained better accuracy (0.88) at the expense of infrastructure dependency.

Moving towards transformer models, Zheng et al. (2022) [19] created LawRec, combining BERT and Skip-RNN, which obtained 92% accuracy in datasets related to Chinese law and demonstrated the benefit of fine-tuned transformer models. Similarly, Zhou et al. (2018) [20] employed TextCNN with BiGRU and Attention, demonstrating strong performance across tasks such as crime prediction, law recommendation, and sentencing, although it was limited outside China. More recently, Nithya et al. (2024) [21] developed an AI-based NLP + ML summarization and drafting and Q&A model that could potentially automate more legal tasks but needs validation across jurisdictions. Together, these publications indicate that SLMs are efficient and performance-focused but cannot scale, adapt, and comprehend context

In summary, the current state of research reveals significant yet fragmented progress. While large language models (LLMs) like LawLLM [7] have demonstrated superior performance on U.S. legal tasks, they are jurisdictionally biased, computationally expensive, and not designed for the nuances of Sri Lankan law. Similarly, hybrid and RAG-based approaches like LQ-RAG [10] show promise in reducing hallucinations and improving relevance but remain untested and unadapted for the Sri Lankan context. Furthermore, existing small language model implementations, such as the cluster-based LDRS for India [17] or LawRec for Chinese law [19], though efficient, are limited by their geographic focus, lack of natural language query support, and absence of RAG integration for factual grounding.

Therefore, a clear and compelling gap exists: there is no system that combines the efficiency of a fine-tuned Small Language Model with the factual accuracy of Retrieval-Augmented Generation (RAG) to provide precise, context-aware, and accessible legal recommendations specifically for Sri Lankan Labor and Employment Law.

Proposed solution

Our solution is a combination of SLMs and RAG, which are fast, accurate, and grounded in law to answer labor and employment law in Sri Lanka. We do not use expensive Large Language Models, but rather our lightweight SLMs are fine-tuned to structured outputs with the help of LoRA, and the FAISS retrieval strategy is used to retrieve the most relevant acts, sections, and precedents. This makes the responses context-sensitive, section-wise, and clear with easy citations. The strategy is cost-effective, efficient, and scalable and minimizes hallucinations and makes legal information accessible and reliable to lawyers and other members of society.

1.2. Research Gap

Legal documents are normally lengthy and unstructured and contain complicated terms. Due to this, individuals usually struggle to find the specific information that they require at any given time in a short period of time. In Sri Lanka, the problem is even bigger since laws are spread across many statutes, case reports, and regulations. All these are highly formal language, making it hard to tell which law or section of the law is applicable to a particular problem for both the lawyers and ordinary people.

some past studies through legal information retrieval or knowledge graphs. However, none of them refer to the special needs of the labor and employment law in Sri Lanka. Most of the previous systems do not apply modern transformer-based NLP models, and they do not add retrieval procedures in this field.

Limitations in Past Research :

- Quick Check [14] – A case law recommender that uses citation networks and ranking. It works fine for U.S. legal data, but it does not fit Sri Lanka's system.
- LawRec [9] – Combines BERT with Skip-RNN and got about 92% accuracy on Chinese legal data. Impressive, but it doesn't handle Sri Lankan law or employment cases.
- Cluster-Based LDRS [10] – Uses graph clustering (Louvain) with Doc2Vec for Indian Supreme Court judgments. Scales well, but is very domain-specific and doesn't allow natural language queries.
- P-LDRS [11] – Adds pre-learned word embeddings and distributed Doc2Vec training. Accuracy and scalability improve, but it doesn't focus on employment law.
- Legal Query RAG [8] – A recursive RAG framework that reduces hallucinations and looks promising, but it has never been tuned for Sri Lankan labor law.

1.2.1. Identified Research Gaps

- No fine-tuned transformer-based models (e.g., T5) for Sri Lankan Labor and Employment Law.
- Lack of integration of RAG to improve legal recommendation accuracy.
- Absence of systems dedicated to Sri Lankan legal documents and their linguistic characteristics.
- Inability to process natural language queries for legal advice and recommendations effectively.
- No provision of detailed legal sections, enactment years, and related scenarios in an accessible format.
- Limited or no user interfaces tailored for both legal professionals and non-expert users.

Table 4:Comparative Analysis of Research Features

Feature	Quick Check	LawRec	Cluster-Based LDRS	P-LDRS	Legal Query RAG	Proposed Project
Transformer Models	No	Yes (BERT)	No	No	Yes	Yes
RAG Integration	No	No	No	No	Yes	Yes
Sri Lankan Focus	No	No	No	No	No	Yes
Labor/Employment Focus	No	No	No	No	No	Yes
Natural Language Queries	Partial	Partial	No	No	Yes	Yes
Legal Sections & Scenarios	No	No	No	No	Partial	Yes
Scalability	Medium	Medium	High	High	Medium	High

Existing Similar Systems and Feature Comparison with my component

Several AI-based tools provide legal information, but their information accuracy, user reliability, and guidance vary greatly. In Sri Lanka, AYCA and AIPAZZ are the two known systems, while OpenAI/ChatGPT represents global large-scale models. None of these fully address the requirements of Sri Lankan labor and employment law.

Table 5:Existing system comparison with our system

Feature	AYCA	AIPAZZ	OpenAI	Proposed system
Information Accuracy	Provides only basic definitions and legal references. Correct but shallow, does not map to law sections, enactment years, or scenarios.	Uses searchable legal databases, giving moderately accurate results. However, answers are general and often lack specificity for labor law. And no recommendation	provide information from referring all available online ,social media because of wrong information also given to user .	Trained on authorized Sri Lankan labor law documents. Outputs detailed sections, enactment years, and scenario-based explanations, ensuring high accuracy.
User Reliability	Low reliability. Users must interpret results themselves, which risks misinterpretation by non-lawyers.	Medium reliability. Provides answers, but without reasoning or context. Users still need external legal help.	Low reliability for Sri Lanka. While confident, it often provides foreign laws or unverified information, which is risky.	High reliability. Uses agentic RAG with confidence scoring, ensuring outputs are verified, consistent, and trustworthy
Technology Level	Minimal NLP. No use of transformers or RAG. Functions as a static glossary.	Basic NLP with keyword matching. No transformers or RAG integration.	Uses large transformerbased LLMs. Advanced in language ability but not domain-specific, and resource-heavy.	Built on a fine-tuned Small Language Model with Agentic RAG, optimized for lowresource environments while maintaining high accuracy.

Practicality (Sri Lanka)	Free and accessible, but too limited in scope and not scalable.	Subscription-based, creating barriers for ordinary users. Not specialized for labor law.	Free to try, but not localized. Provides quick answers but with high risk of wrong or irrelevant results.	Free, inclusive, specialized for Sri Lankan labor law. Designed for workers, lawyers, and the general public.
Scenario-Based Guidance	No	No	No	Yes

1.2.2. Proposed Solution

To address these gaps, this research proposes the development of a fine-tuned small language model integrated with RAG, trained on curated Sri Lankan Labor and Employment Law documents. The system will process natural language queries and return accurate, context-aware legal recommendations, including relevant law sections, enactment years, and scenario-based explanations. It will feature a user-friendly interface accessible to both legal professionals and laypersons, ensuring efficiency, accuracy, and inclusivity in accessing Sri Lankan Labor and Employment Law.

2. Research Problem

The legal system in Sri Lanka is quite complicated since its laws are not gathered in one clear framework but are spread across different acts, sections, and amendments passed at different times. This situation makes it hard to identify the exact law that fits a particular problem. One important area where this difficulty can be seen is Labor and Employment Law. Here, not only the general public but even lawyers sometimes find it confusing to trace the correct section or the right year of enactment. This clearly shows the scattered nature of the system. In other countries, especially developed ones, digital legal assistants and advanced decision-making tools help to solve these problems. In Sri Lanka, however, people still mostly rely on manual searches, checking libraries, or using very simple keyword searches online. These ways may give some information, but often they fail to capture the whole meaning of a legal question, and as a result, the answers may be incomplete or misleading.

At present, there are a few digital systems such as AYCA and AIPAZZ. While they do provide some help, their services are very limited. They mostly work as search tools and do not make use of technologies such as NLP or transformer-based models that can process complex or detailed queries. It is important to note that they also do not give special attention to Labor and Employment Law, which is one of the areas where such tools are actually needed. Another issue is that these resources are mainly built for legal experts, so ordinary citizens cannot easily make use of them. This situation leaves a clear gap in Sri Lanka, as people still lack simple, accessible, and reliable digital legal support.

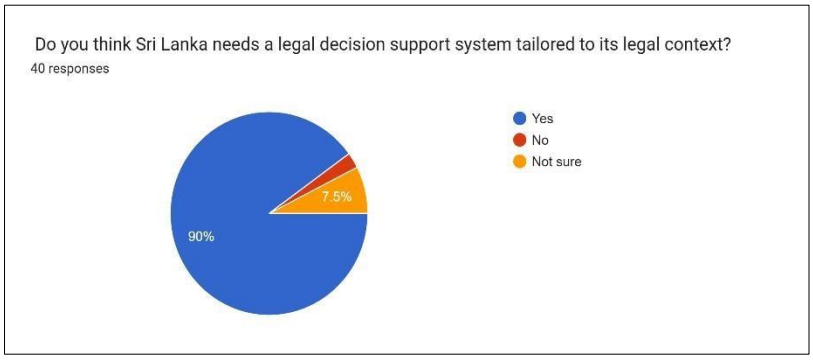

Figure 3: Survey of existing legal information

A survey of 40 participants, including lawyers, law students, and members of the general public, revealed that over 90% agreed there is a pressing need for a legal decision-support system tailored to the Sri Lankan context. The survey responses highlighted three major issues: delays in accessing

relevant legal documents, difficulty in interpreting fragmented and technical laws, and lack of user-friendly systems that allow natural language interaction. Without a dedicated system, users risk misinterpretation, wasted time, and in some cases, denial of justice.

From a technological standpoint, current systems lack fine-tuned transformer-based models such as T5, DisitilBERT , MiniLM, Legal-BERT, which are proven effective in other domains for semantic understanding and text generation. They also fail to incorporate Retrieval-Augmented Generation (RAG) frameworks, which combine document retrieval with generative models to improve relevance and reduce hallucinations. This absence results in poor handling of domainspecific legal reasoning, low adaptability to the Sri Lankan legal framework, and limited usability for both professionals and non-experts.

Thus, the research problem can be summarized as follows:
Despite the urgent need for accurate, efficient, and context-aware legal decision support, Sri Lanka currently lacks an AI-driven legal recommendation system that integrates NLP, transformer-based models, and RAG, tailored specifically for Labor and Employment Law. This gap creates barriers to justice, reduces efficiency in legal practice, and widens inequality in access to legal knowledge.

3. Objectives

3.1. Main Objectives

The primary goal of this project is to create a Law Recommendation System of the Sri Lankan labor and employment law that will be able to interpret the query made by the user in a natural language and give a response with the legal term and section, the year of enactment, and similar cases. SLMs fine-tuned with Low-Rank Adaptation (LoRA) should be used to provide the system with domain-specific knowledge in a cost-efficient and resource-efficient way, and finally, RAG with the help of FAISS-based retrieval will provide the responses with the grounding in context and the support of citation and reliability.

3.2. Specific Objectives

The proposed law recommendation system aims to achieve the following objectives:

- To create a centralized and organized legal corpus by way of compilation, pre-training, and labeling of Sri Lankan labor and employment law by retrieving, examining, and labeling Sri Lankan law acts and pertinent case precedents by authoritative sources.

- To design a robust NLP pipeline by training a Transformer-based SLM on natural language queries to structure outputs with the relevant name of the law, section and year of enactment, and a situational scenario.

- To be factual and minimize hallucinations by means of integrating a RAG model, which has the answers given by the model based on retrieved passages of the authoritative legal corpus.

- To launch a scalable and easy-to-use web application that includes an intuitively designed interface that allows end-users to place queries and receive clear and well-structured recommendations, it is essential to make it accessible to the end-users with different levels of legal expertise.

- To perform a systematic assessment of the system performance in terms of quantitative measures on a test set that is expert-evaluated and qualitative feedback on the usability of the system and its accuracy and practical utility provided by legal practitioners and the students.

- Establish a scalable framework that can be expanded to cover other areas of Sri Lankan law in the future.

4. Methodology

This part explains how the research and development will be done to reach the goals of the proposed system. The main idea is to make a web-based legal recommendation tool that can understand questions written in normal language and give the most relevant Sri Lankan laws. This includes the right sections, years, and situations where the law might apply. Honestly, it's not always easy for people to find or understand the correct law, so this system could really help. Given that the legal sector is a corporate and formal industry where professionals primarily work using desktop or laptop computers, the system will be implemented as a web application. The development process will follow iterative software development principles, where incremental improvements are made until the final, fully functional product is achieved and research goals are met.

The proposed system comprises four interdependent development objectives, each critical to building a comprehensive legal recommendation system. The objectives are:

1. Collection, Extraction, and Structuring
2. Preprocessing and Model Fine-tuning
3. Retrieval-Augmented Generation Integration
4. Legal Decision Support System Deployment

These four features are deeply interconnected, where the output of one stage serves as the foundation for the next. Together, they will produce the final system the Sri Lankan Law Recommendation System.

A. Legal Data Collection, Extraction, and Structuring

The first stage focuses on building a reliable dataset by collecting statutes, labor laws, and case precedents from authoritative Sri Lankan sources such as government gazettes, official digital archives, and legal databases.

1. Extraction and Cleaning – Collected texts will be cleaned to eliminate inconsistencies, noise in formatting and redundancy

2. Entity Identification – Techniques such as Named Entity Recognition (NER) and Rule-based parsing will be employed to extract law titles, section names, enactment years, and contextual scenarios.
3. Data Structuring – Extracted data will be organized into a structured format (JSON/CSV), creating a uniform schema suitable for model fine-tuning and retrieval.

B. Preprocessing and Transformer Model Fine-tuning

To capture legal linguistic patterns, the small language model will be fine-tuned using the structured dataset.

1. Text Preprocessing – Includes tokenization, lemmatization, stop-word removal, and embedding generation.
2. Dataset Splitting – The dataset will be divided into training, validation, and testing sets to ensure robust evaluation.
3. Fine-tuning – The small language model will be trained to map natural language queries to legal recommendations, enabling contextual understanding and generation of structured outputs (law, section, year, scenario).

C. Retrieval-Augmented Generation (RAG) Integration

To enhance factual grounding, the small language model is integrated with a RAG pipeline.

1. Embedding Generation – Legal texts will be embedded into a vector space using domain-specific embeddings and stored in a vector database (FAISS/Pinecone).
2. Document Retrieval – At query time, relevant documents will be retrieved and combined with small language model outputs to reduce hallucinations.
3. Hybrid Response – The integration ensures that generated responses are both context-sensitive and legally grounded, improving reliability and trustworthiness.

D. Legal Decision Support System Deployment

The final phase involves deployment of the system as a web-based application tailored for legal professionals and the public.

1. Backend – Implemented using FastAPI, containerized via Docker for scalability.
2. Frontend – Provides a user-friendly interface for query input and structured recommendation output.
3. Core Functions –
   a. Relevant Document Retrieval: RAG retrieves top-ranked law sections.
   b. Recommendation Generation: Fine-tuned small language model provides structured answers including law name, section, year, and scenario.

E. Evaluation and Validation

In making sure that the proposed Law Recommendation System is reliable, accurate, and usable, the assessment plan shall be as follows with three important aspects; the development of labeled test set, quantitative evaluation based on predetermined metrics, and qualitative evaluation, based on structured user studies.

1. Labeled Test Set Creation

- Upon the development of a gold-standard test set, the collaboration between two practitioners in labour-law in Sri Lanka and a legal academic supervisor will be employed.
- The test set will involve user queries in natural language and represents a wide variety of situations relating to labour and employment-law (e.g. EPF/ETF issues, wrongful termination, salary disputes, workplace harassment, maternity benefits and contract violation).
- The ground truth will be made up of annotating each query with the appropriate law name, section number, year it was enacted, and a short scenario description to the query.
- The dataset will be split into training, validation, and testing subsets, with the test set kept unseen during model development to ensure unbiased evaluation.

2. User Study Methodology

The evaluation will be user-based in order to confirm the usability, interpretability, and trustworthiness:

- Participants: law students, practicing lawyers, and HR/legal officers (20 total users).
- Procedure: Each participant will complete 10 real-world tasks (legal queries) using the system and a baseline tool (e.g., AIPAZZ) for comparison.
- Data Collected:
    - Task completion time
    - Perceived accuracy and clarity (Likert 1–5 scale)
    - Confidence in recommendations
    - Feedback on UI/UX
- Success Criteria:
    - Average usability score ≥ 4/5
    - Significant reduction in time compared to baseline
    - Positive qualitative feedback on clarity and trust

All feedback will be analyzed to guide iterative improvements to both the model and the interface

F. Ethical and Privacy Considerations

Privacy: Only publicly available legal PDF texts and verified data from lawyers are used.
Transparency: The system makes clear that it only explains laws and gives guidance, not legal advice.

4.1. System Architecture Diagrams

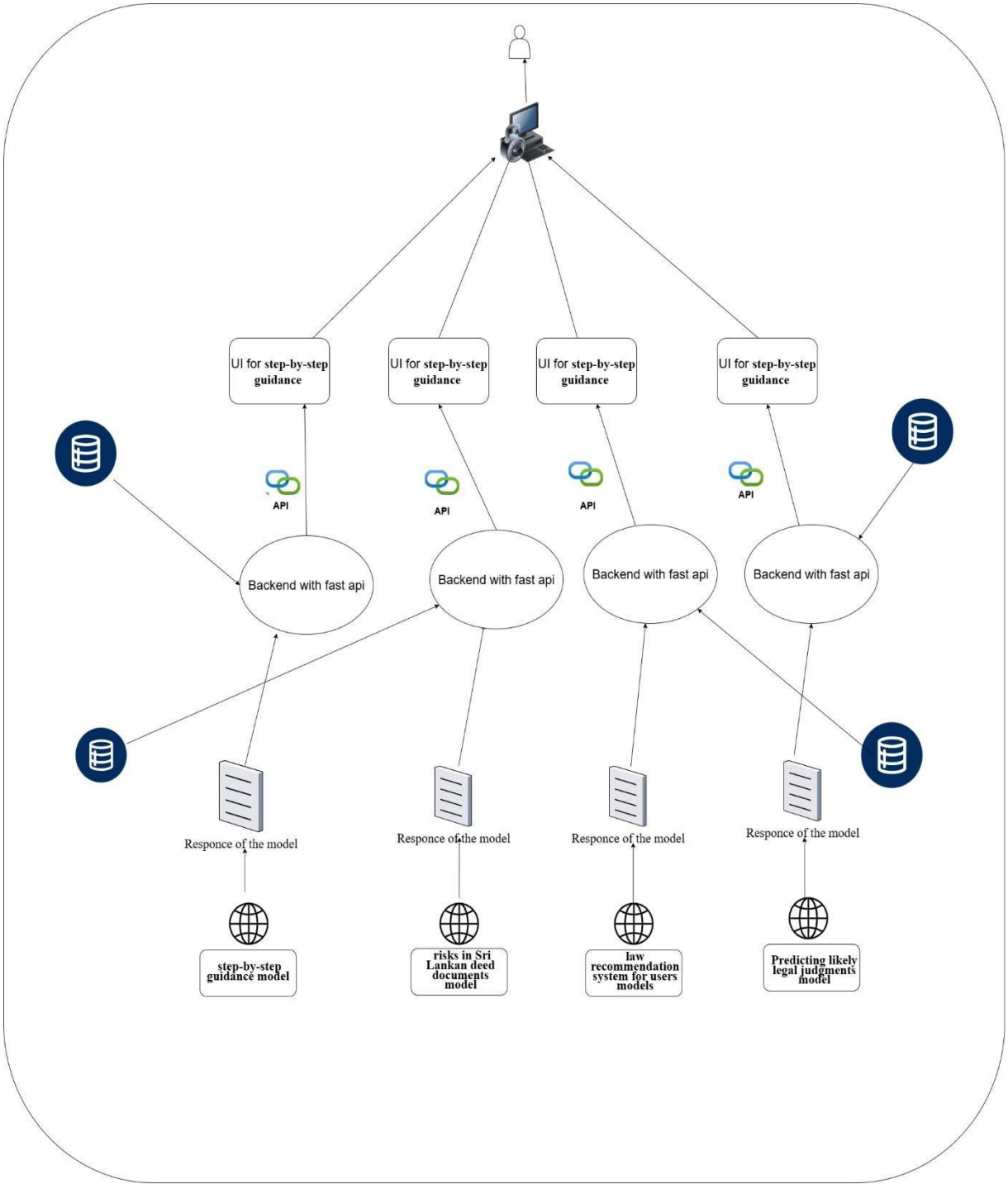

Figure 4:system architecture

4.2. Component Specific System Diagram

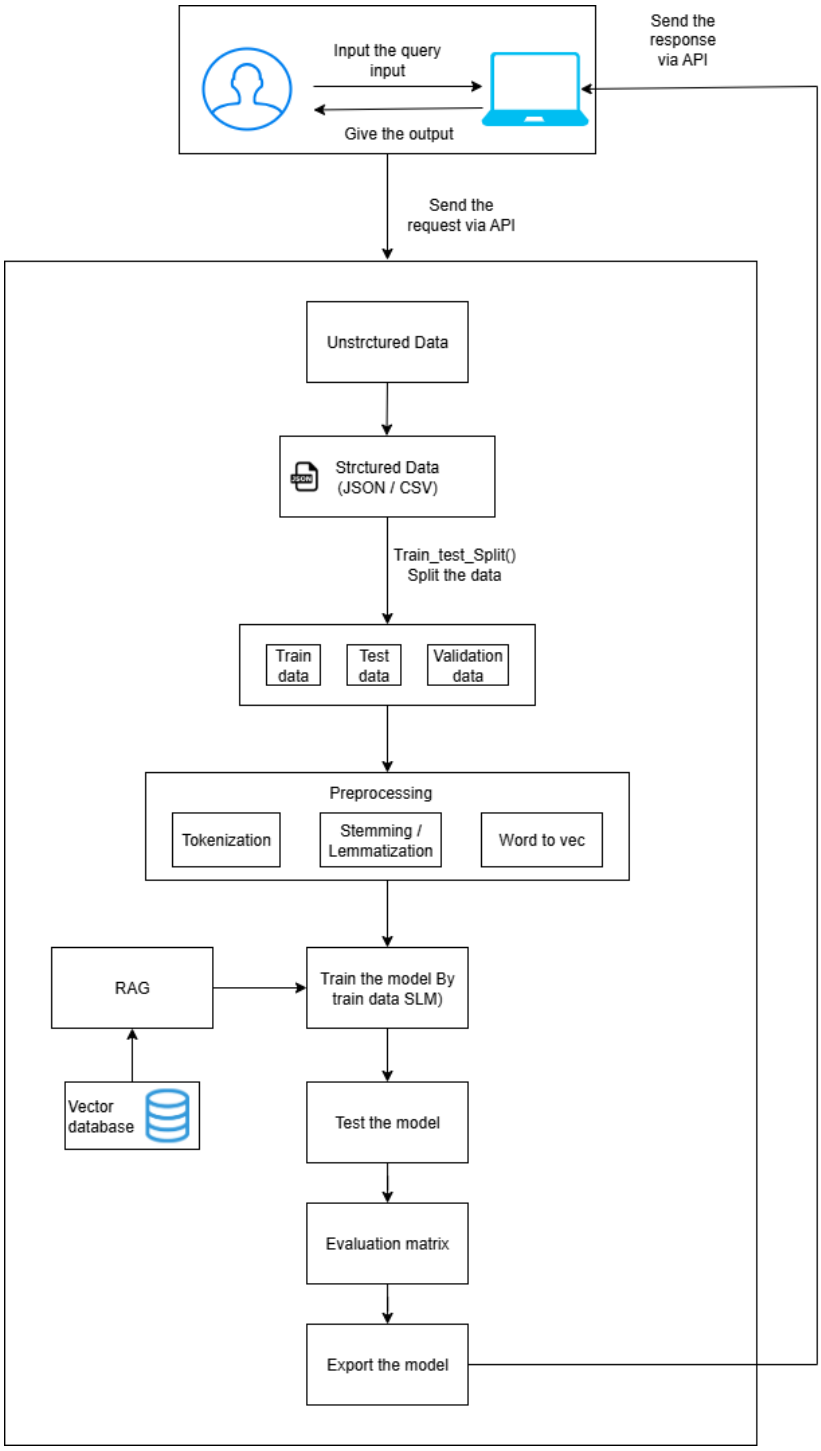

Figure 5:Component specific diagram

The development of the proposed Law Recommendation System will commence with legal-information acquisition from authoritative sources, including LawNet, the Ministry of Justice, the Supreme Court of Sri Lanka, and other official archives. When required materials are not available online, primary documents will be collected from practicing lawyers; these will be scanned, digitized, and archived as PDFs to ensure traceability.

The raw corpus will then be transformed into a structured dataset (JSON/CSV) using automated information-extraction methods (e.g., LLM-assisted parsing with rule-based validation). The dataset will be partitioned into training, validation, and test sets to support reproducible experimentation. Prior to modeling, standard preprocessing will be applied—text cleaning, tokenization, lemmatization, stop-word removal—and document embeddings will be generated to support retrieval.

Model development will focus on fine-tuning Small Language Models (SLMs) specifically Qwen-2.5, GPT-2 Small, and T5-base—to identify the model that achieves the best task performance. Fine-tuning will be conducted with Unsloth and LoRA to reduce GPU memory requirements and accelerate training, enabling efficient experimentation in resource-constrained settings. SLMs are selected over larger models due to their lower computational cost, faster inference, and strong performance after domain-specific fine-tuning.

To improve factual grounding, a RAG pipeline will be integrated prior to deployment. All acts, sections, and relevant case materials will be embedded and indexed in a FAISS vector database. At inference time, the system retrieves the most relevant passages and supplies them to the SLM via a schema-guided prompt, yielding citation-backed, context-aware outputs.

Finally, the integrated SLM+RAG system will be evaluated on the held-out test set using standard metrics (e.g., F1-score, accuracy, and mean reciprocal rank), alongside qualitative assessment by legal experts. The best-performing configuration will be deployed as the production recommendation engine.

A rigorous, reviewer-oriented evaluation will use a lawyer-verified gold set of 200–400 queries labeled with *act_name, section, year,* and supporting excerpts; data are split 70/15/15 into train/dev/test with the test set frozen, and all models are run under identical prompts/decoding with results reported as mean ± SD over three seeds. Retrieval quality is assessed by Recall@5

(target ≥ 0.85), MRR@10 (≥ 0.70), and Precision@5 (≥ 0.70); answer correctness for the SLM+RAG pipeline is measured by section-level Exact Match (≥ 0.75), field-level F1 for {act, section, year} (≥ 0.80), Citation-Support Rate (≥ 0.90), and JSON-Validity (100% via schema checks). Operational SLOs track end-to-end p95 latency ≤ 5 s, retrieval p95 ≤ 500 ms, and SLM inference p95 ≤ 2.5 s at 20–50 RPS. Robustness is examined through ablations (no-RAG vs RAG; single- vs multi-vector; k∈{3,5,8}) and a prompt-injection canary suite with zero critical leaks. Monthly human evaluation by practicing lawyers on a blinded sample (≈20 queries) rates legal adequacy and clarity (targets ≥ 4.0/5) and logs an error taxonomy (wrong section, unsupported citation, ambiguity). Metrics are recomputed nightly on dev/test and pushed to dashboards; 5% of anonymized live queries are sampled for expert review; deployment occurs only when all targets are met, otherwise the system rolls back to the last green model.

4.3. Data Collection

Data collection is a crucial phase in the achievement of law recommendation system efficacy and precision. To provide the relevant legal advice answering the questions posed by the users, the system requires a complete body of the Sri Lankan legal document

The data collection process will focus on the accumulation of the following types of legal issuance:

- Acts of parliament and acts of section (with date of enactment)
- Interpretative studies and legal commentaries
- Applicable legal cases and precedents.
- Both Sri Lanka Reports (SLR) and New Law Reports (NLR) which contain decisions of the judiciary authority

Since the number of Sri Lankan legal materials exceeds by far, the undertaking would be guided by specialists in the domain and they would determine the most relevant legislation, cases, and provisions that are common when a user undertakes a query. This is a specific approach that will create a representative data set that is appropriately customized to the recommendation system.

The relevant legal sources are publicly available in accordance with the Right to Information Act, No. 12 of 2016 that provides the open access to the information of the government. These are mostly found in the official government websites in PDF and HTML version.

In order to extract the bulk of data necessary in automation, a web scraper will be implemented in Python. The scraper is going to address the following official repositories:

- Ministry of Justice (www.lawnet.gov.lk)
- Court of Appeal of Sri Lanka (courtofappeal.lk)
- Supreme Court of Sri Lanka (supremecourt.lk)
- Department of Government Printing (documents.gov.lk)
- Sri Lanka Reports (SLR) and New Law Reports (NLR) archives

Along with the automated scraping, manual data collection will be done with the help of the working lawyers and law experts. Such manual work will help assure that the recent and

unpublished cases or other hard-to-find cases and interpretations are included that otherwise might not be accessible via other public channels. Qualification of the quality and relevance of the aggregated data will also be facilitated by the involvement of the legal professionals.

The gathered data will be preprocessed and structured in a way that allows it to be inputted into the fine-tuning of a specialized language model and as input to the RAG architecture, therefore, increasing the accuracy of law recommendation outputs.

4.4. Software Solution

The process of creating the law recommendation system will be carried out according to the Agile and Scrum approaches to make sure that the process of software development is efficient, adaptive, and user-centered. The methodologies would be suitable in the project that needs flexibility and often incorporation of stakeholder feedback to develop a system that appropriately responds to the complex and dynamic needs of legal information retrieval.

Agile methodology facilitates the development of iterative and incremental development in short cycles referred to as sprints, which are normally shortened to two to four weeks. Sprints are aimed at providing a functional part of the system, as a result of which the development team can receive feedback from its stakeholders, such as domain experts, legal professionals, and end users, as quickly as possible. Such feedback can be very helpful in streamlining the capabilities of the system, making it easier to use, and making the recommendations the system makes legally sound and pertinent.

Sprint planning will be done periodically to prioritize the tasks and to have realistic goals that can be achieved at every iteration. Sprint reviews will be used after each sprint, giving stakeholders an opportunity to assess progress and propose changes or improvements. To analyze the retrospectives will be also held performance, communication, and workflows of team, and promoting culture of continuous improvement

The Scrum framework is a complementary framework to Agile principles that have structured roles and ceremonies that have improved coordination of teams and transparency of projects. The Scrum team will be composed of developers, data scientists, experts in the legal domain, and a Scrum Master to facilitate the process

Moreover, Scrum emphasizes collaboration with stakeholders through frequent demonstrations of working software, ensuring that the development aligns with user expectations and regulatory standards. This is particularly important in the legal domain where accuracy and compliance are critical.

In addition to Agile and Scrum, the project will integrate continuous integration and continuous deployment (CI/CD) pipelines to automate testing and deployment processes. This approach will ensure that new features and improvements are quickly and safely incorporated into the system, maintaining high software quality and reliability.

By adopting Agile and Scrum methodologies, the development process will be transparent, flexible, and collaborative, ultimately leading to the delivery of a robust law recommendation system. This system will not only meet functional requirements but will also adapt to emerging legal information and user feedback, making it a valuable tool for users seeking precise and trustworthy legal guidance.

Figure 6:Agile development life cycle

5. Project requirements

5.1. Functional requirements

- User Query Input:

  - The system will also have a text input area where the users will input their queries using natural language.

  - The system will be able to handle and analyze queries with normal grammatical errors, synonyms, and paraphrases.

- Law Retrieval:

  - The system will extract and suggest the most applicable labor laws in Sri Lanka depending on the same query's semantic meaning.

  - Each recommendation will list the law name, section number(s), the law enactment year, and a verbatim statute fragment

- Scenario Generation:

  - The system shall generate related legal scenarios or interpretations under the retrieved laws to enhance contextual understanding.

- NLP Model Integration:

  - The system shall employ a fine-tuned small language transformer model to accurately interpret and process user queries.

- RAG Architecture:

  - The system shall implement Retrieval-Augmented Generation (RAG) to combine factual retrieval with generative responses, minimizing hallucinations.

- Data Management:

  -The system will have a versioned database of legal documents (statutes, acts, and case law from SLR/NLR).

- Feedback Collection:

  - The system shall allow users and legal experts to provide feedback on accuracy and relevance, enabling continuous improvement.

- User Interface:

  - The system should provide a user-friendly web interface for query entry and structured recommendation output.

5.2. Nonfunctional requirements

- Performance: The responses would be generated within 5 seconds so as not to jeopardize the user experience
- Scalability: The system shall scale horizontally to handle increasing query volumes and corpus expansion.
- Reliability: The system will be very stable in the sense that it will not take a long time to be down or provide wrong responses.
- Security: Data entry and legal data sets will be kept safe and will abide by the data protection laws.
- Usability: The interface will be user-friendly to both lawyers and ordinary people.
- Maintainability: The architecture will ensure that datasets and NLP models can be updated easily without involving any major downtime
- Adherence: The system will be in conformity with the laws and regulations of data privacy and the rules of general use and sharing of legal documents.

5.3. Expected Test cases

Table 6:Expected Test cases

Test Case ID	Test Objective	Pre-conditions	Test Steps	Expected Result
TC-001	Verify system returns correct law for valid query	System is running and connected to the backend & vector database	1. Open web application. 2. Enter query: "My employer has not paid EPF for 10 years". 3. Click "Search".	Output includes correct law name, section, year, and applicable scenario from dataset.
TC-002	Verify system handles spelling errors in queries	Same as above	1. Enter query: "My employr didnt pay EPF". 2. Click "Search".	System corrects or tolerates spelling errors and returns relevant details.
TC-003	Verify response for query outside dataset coverage	Same as above	1. Enter query: "How to apply for a driving license". 2. Click "Search".	System displays "No relevant legal section found" without error.
TC-004	Verify RAG retrieves most relevant law documents	System has vector database populated	1. Enter a query linked to a unique law section. 2. Inspect retrieved documents.	Top retrieved documents are the correct legal sections matching query context.

TC-005	Verify empty query validation	Same as above	1. Leave query box empty. 2. Click “Search”.	System displays “Please enter a query” and does not send API request.
TC-006	Verify dataset coverage	Same as above	1. Choose a known law from dataset. 2. Enter matching query.	Output matches law name, section, year exactly from dataset.
TC-007	Verify API endpoint returns correct JSON	API server running	1. Send POST request to /predict	API returns { law name, section, year, scenario }
TC-008	Verify multiple sentence query handling	Same as above	1. Enter: "I worked for 5 years without EPF. The company also refused to give my salary slip". 2. Click “Search”.	System identifies both issues and output laws related to EPF and salary slips.
TC-009	Verify model output format consistency	Same as above	1. Enter multiple queries (min. 5). 2. Check API responses.	All outputs follow consistent JSON format: { "law name": "", "section": "", "year": "", "scenario": "" }.

TC-010	Verify correct handling of irrelevant but legal-domain query	Same as above	1. Enter query: "How many judges are in the Supreme Court of Sri Lanka?".	System responds with “No relevant law section found” or provides reference if available.

5.4. Feasibility Study

A feasibility study is done, where the proposed Sri Lankan Legal Information Retrieval and Recommendation System is analyzed to determine its practicability, functionality, and usefulness in its development. The research is conducted within the following aspects:

- Technological Feasibility

This tool is technically reasonable since it relies on lightweight and open-source solutions like Small Language Models (T5, DistilBERT, Qwen2.5, GPT-2 Small), fine-tuned with PyTorch and Unsloth. To retrieve, FAISS and RAG have guaranteed effective and precise results. It will be built with FastAPI on the backend and React.js on the front, with MongoDB for data storage. As every technology is affordable and scalable, the implementation can take place without problems, and the demands in regard to the GPUs can be handled through Google Colab

- Economic Feasibility

The project is economical in terms of the economy since it relies on free or open-source frameworks such as PyTorch, FAISS, and React.js. Preparation of dataset and the use of codlab in Google are the primary expense. The advantages are however, high-faster access to law, time saved by the legal professionals and high potential of commercialization since there is no such system in Sri Lanka. This implies that the system is of great benefit to its cost.

- Operational Feasibility

Operationally, the system will be useful to lawyers, HR managers, and employees in that it will offer the answers to the natural language queries in structured form (Act, Section, Year, Scenario). It simplifies the access to legal information. The only issue here is that non lawyers may either abuse or misinterpret the output but this will be governed by placing clear disclaimers that the tool is not legal advice but merely a guide.

- Legal Feasibility

  The project is a legal one since they utilize only the government PDFs publicly available and lawyer-verified materials. It complies with the law of the Right to Information Act and does not use personal or exclusive material. There will also be transparency since the system will display sources and disclaimers and it is evident that outputs are not binding legal advice but only explanatory. This would make the system operate ethically and legally.

- Schedule Feasibility

  The project schedule from May 2025 to May 2026 is realistic, giving time for dataset collection, preprocessing, model fine-tuning, backend and frontend development, integration, and evaluation. The main risk is delays in dataset collection, but this will be managed by working with lawyers in parallel while other technical parts continue. This ensures timely completion within one year.

6. Tools and technologies selection

6.1. Tools

- Visual Studio Code

  A lightweight, cross-platform source code editor with extensive language and tooling support. Used for coding the backend and frontend components.

- GitLab

  Manages source control, CI/CD pipelines, and code reviews to streamline collaborative development.

- Google Colab

  For training and fine-tuning NLP models (e.g., T5 , DISTILBERT) in a cloud environment with access to GPUs and TPUs.

- Trello

  Organizes tasks, sprints, and workflow management to maintain Agile development cycles.

- Microsoft Teams

  Facilitates communication and collaboration among developers and stakeholders.

6.2. Technologies

- FastAPI

  FastAPI is a modern web application framework, which is faster and lightweight in comparison to Flask or Django. It is asynchronous, automatically generates API documentation and is also high performing.

- React & Vite

  React provides the front-end user interface, whereas Vite represents a fast build tool and development server to React applications, which can be used to quickly repeat a development cycle and generate optimized build outputs.

- FIASS

  It is an effective similarity-search library based on embeddings. It is much faster than ordinary database consultations and has been streamlined to artificial intelligence workloads. FAISS is used to rely on Retrieval-Augmented Generation (RAG) pipelines that retrieve the most relevant sections of the law in a database in a timely way.

- NLP Libraries

  Hugging Face Transformers make it easy to make use of and customize smaller language models. spaCy and NLTK offer text-processing functions, including tokenization, named-entity recognition and stop-word removal. Regular expressions are used to purify legal text and are used to extract structured information.

- PyTorch

  This is a highly adaptable and popular deep-learning platform, which is efficient in terms of model fine-tuning and can be easily connected to the Hugging Face ecosystem.

- Unsloth

  Unsloth uses up to about thirty times less operation cost and encores more effectively by speeding up the refinements of language models, optimizing the process and reducing the memory usage..

- RAG

  It enhances predictive accuracy by grounding model outputs to the retrieved legal documents to reduce hallucinations. hallucinations.

6.3. Implementation

The system is implemented as a web application with the following components:

- Frontend:
  React/Vite one-page application. This can be combined to create fast and effective development of an interactive and responsive user interface through which users can inquire with natural language and see visible and organized legal advice.

- Backend:
  A high - performance RESTful API using FastAPI. This back-end server will be dealing with the entire business logic, which includes getting and authorizing user requests. Arranging the NLP inference and RAG retrieval process. Controlling every communication with the databases. Sending back organized JSON data to the front.

- NLP Model Integration:
  The main artificial intelligence engine will be made up of: A T5 Transformer model which is fine-tuned and used to interpret queries and produce responses. A Retrieval-Augmented Generation (RAG) pipeline, with FAISS as the vector database, to do an effective similarity search of the embeddings of the legal corpus. This is to make sure that all the answers generated are based on the real Sri Lankan law.

- Database:
  The system will use a two-database approach to taking advantage of the better sides of various databases paradigms:
  MongoDB: Will be the main working database. MongoDB is a document-based NoSQL database and hence suitable due to its ability to be flexible and also due to its ability to query the database in a powerful way.
  FAISS: Will be the special high-performance vector database that is only used to store and recall the numerical embeddings (vector representations) of the legal text chunks. This is the separation which makes the semantic search operations as fast as possible.

6.4. Deployment

- Amazon Web Services (AWS):
  The complete application is deployed on AWS, leveraging services such as Elastic Beanstalk or AWS Lambda for backend hosting, S3 and CloudFront for serving static frontend files, and AWS RDS or other services if relational data is needed alongside Firebase.
- CI/CD Pipeline**:**
  GitLab pipelines automate testing, building, and deployment processes to AWS, ensuring rapid delivery and high system availability.

7. PERSONNEL AND FACILITIES

Since the proposed Sri Lankan Law Recommendation System is directly connected to the highly professional field of law practice, it is necessary to involve the expert views in the process to make sure that the recommendations made by it are correct, pertinent, and do not violate ethical standards. As such, an interpretation of the law needs expertise that is not simply limited to the technical skills of the development team; the services of a legal expert have been integrated into the project.

Legal Expert Contribution

The legal professionals that will bring domain-specific knowledge to the project will include the following professionals:

Mr. Pirabakaran, Sri Lanka Law College—specialist in Sri Lankan statutes, case law, and law drafting.

Mr. Pirabakaran will be instrumental in simplifying complicated legal requirements, maintaining the application of relevant legal and ethical requirements in the system, and the authenticity of generated recommendations before the deployment of the system. His participation ensures that the system is in correspondence with the practice of law in the real world and the peculiarities of the Sri Lankan labor and employment law

Technical Guidance

To complement the legal expertise, technical supervision will be provided by academic mentors specializing in Artificial Intelligence (AI) and Knowledge Systems:

- Dr. Prasanna Sumathipala (Supervisor) – Center of Excellence for Artificial Intelligence, responsible for guiding the fine-tuning of the small language model and the integration of NLP techniques.
- Dr. [Karthiga Rajendran] (Co-supervisor)

Their combined expertise will ensure the seamless integration of advanced AI methodologies into the legal decision support framework.

Facilities and Resources

The success of this project depends on access to both legal and technical resources. Essential facilities include:

- Legal Data Access: Official government legal portals (e.g., Ministry of Justice website, LawNet), online legal databases, and published legal reports (e.g., SLR and NLR).
- Computing Infrastructure: GPU-enabled systems for model training, cloud platforms for scalable data storage and retrieval, and secure servers for hosting the deployed system.
- Software Tools: Python, FastAPI, React.js for web development, Hugging Face Transformers for NLP model fine-tuning, FAISS/Pinecone for vector-based retrieval, and Docker for containerized deployment.

Together, these expert contributions and facilities will ensure that the proposed system is both technically robust and legally reliable, meeting the dual objectives of advancing AI research while addressing critical gaps in Sri Lankan Labor and Employment Law accessibility.

8. Work breakdown structure

A Work Breakdown Structure (WBS) is a project's hierarchical division into smaller, more manageable components or work packages. It is a graphical representation of the project scope that divides it into deliverables and sub-deliverables. The work breakdown structure (WBS) enables effective task organization and planning, ensuring that all project components are accounted for and completed on time. The structure can be presented graphically or tabularly, as shown below:

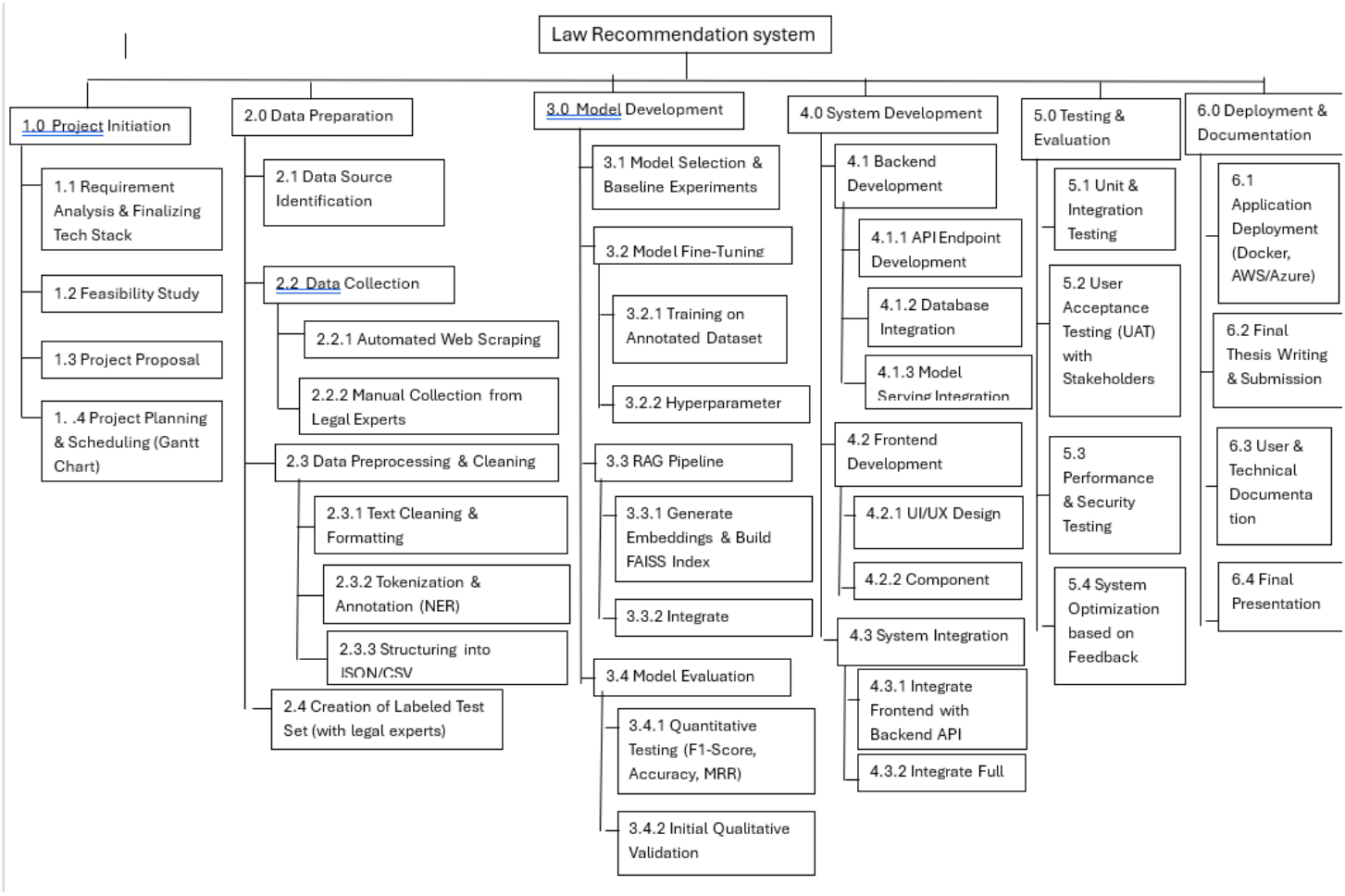

Figure 7:Work break down structure

9. Gantt Chart

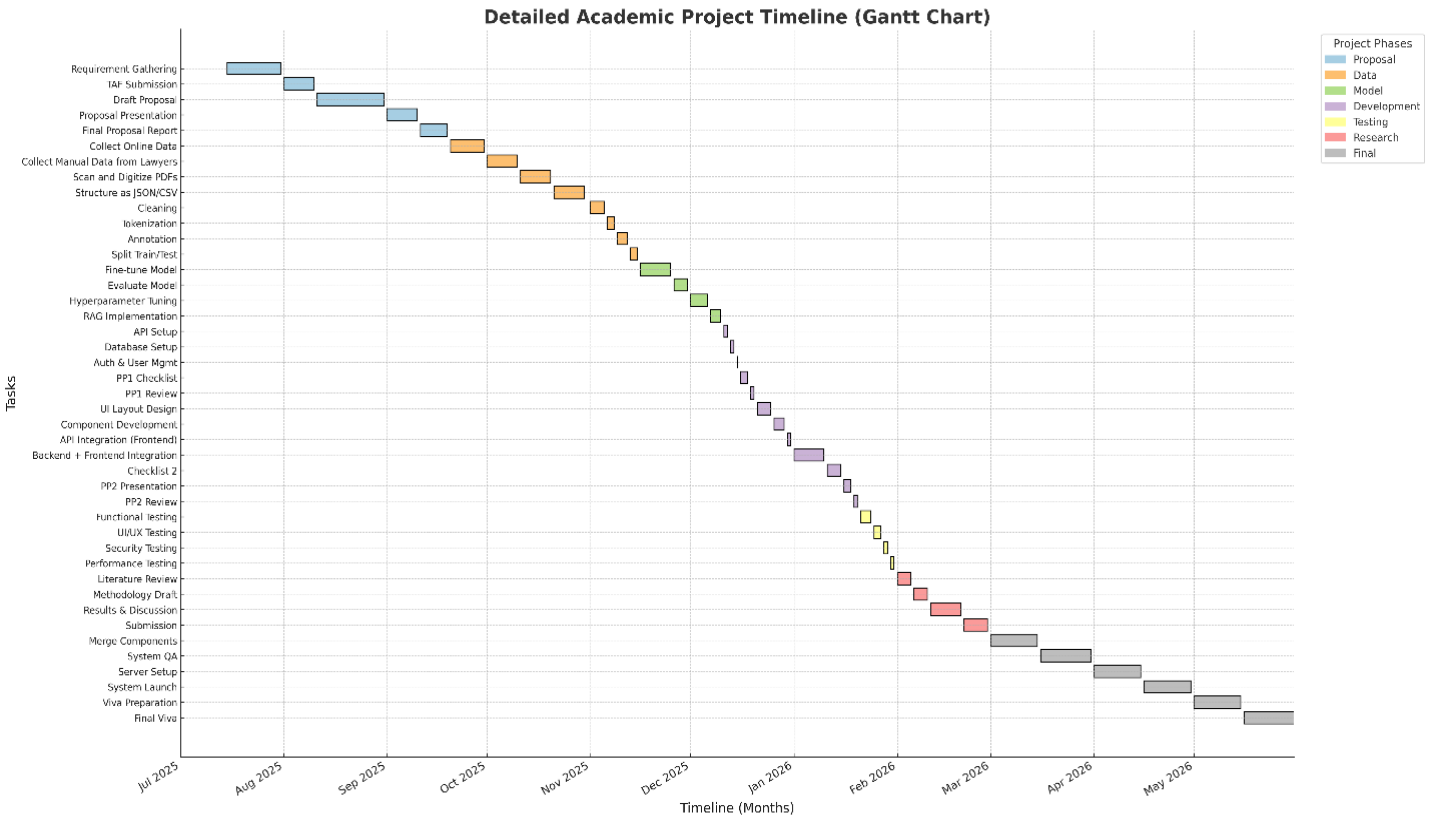

Figure 8:GANTT Chart

10.BUDGET AND BUDGET JUSTIFICATION

Table 7:Budget

Item	Description	Cost (LKR)
Google Colab Pro Subscription	Cloud-based GPU resources for finetuning the small language model and running NLP tasks	3500
Advertisement	Online and offline promotions to raise awareness of the system	10000
Manual Document Scanning	Scanning physical legal documents for dataset creation	5000
Documentation Printing	Printing research reports, manuals, and related documents	2000
Internet Cost	Internet access for research, cloud usage, and collaboration	5000
Total		25500

The total estimated cost for the Sri Lankan Law Recommendation System is LKR 25,500. This budget covers essential expenses for cloud computing resources, promotional activities, data digitization, printing documentation, and internet connectivity. These costs ensure the smooth development, testing, and deployment of the system while supporting its outreach and adoption among intended users.

11.COMMERCIALIZATION OF THE PRODUCT

Market Opportunity

Sri Lanka is a developing country, similar to other developing countries that have difficulties in dealing with complicated and constantly evolving legal systems, especially Labor and Employment Law. The proposed system seals this divide by using Natural Language Processing with Retrieval-Augmented Generation (RAG) to provide wide-reaching relevant recommendations, such as law titles, law sections, the year they were enacted, and situations to apply those law sections.

Potential Users include:

- Lawyers & Law Firms: Reduce research time and access precise legal precedents.
- Businesses and HR: Get immediate advice about the compliance of employee issues.
- Government Departments & Legal Aid Services: Aids in the correct and quick response to the people.
- Courts and Labor Tribunals: Rapidly confirm the laws and candidates of related case situations.
- Universities & Law Students: Enhance learning with structured examples and direct citations.
- General Public: Easy and straightforward access to legal information without the services of a professional.

Unique Selling Points:

- Combines small language model and RAG for context-sensitive and grounded responses.
- Provides exact law references (section, year, scenario) instead of vague keyword matches.
- Specially fine-tuned for Sri Lankan law, with adaptability for other jurisdictions.

Revenue Model

1. Business Subscriptions (B2B):
    a. *Small Plan*: Solo lawyers and small firms (limited queries/month).
    b. *Professional Plan*: Medium-sized firms with extended features and higher limits.
    c. *Enterprise Plan*: Large firms, government institutions, and universities with unlimited queries, custom dashboards, and enhanced security.

2. Government & Institutions:
    a. Annual licenses for ministries, labor departments, universities, and tribunals.
3. Freemium (Public Access):
    a. *Free Tier*: Limited queries with simplified answers.
    b. *Paid Upgrade*: Unlimited access, detailed citations, downloadable reports.
4. API Access:
    a. Licensing fees for integration with third-party software (HR systems, case management tools).
5. Custom Services:
    a. Domain-specific AI fine-tuning, private legal databases, and compliance audits for large clients.
6. Data Reports (Optional & Privacy-Safe):
    a. Aggregated insights (e.g., most-searched laws) for research organizations and policymakers.
7. White-Label Offering:
    a. Licensing technology for organizations to rebrand and deploy under their own systems.

Pricing

- Flexible pricing model: Monthly/annual subscriptions + pay-per-use for occasional users.
- Discounts for early adopters, universities, and government institutions.
- Service-level guarantees (e.g., uptime, response time) for enterprise customers.

12.Conclusion

The law profession in Sri Lanka, especially in the area of Labor and Employment Law is typified by disjointed laws, overlapping, laws and multi-faceted interpretation of the procedure. This poses substantial problems to both the legal professionals and the general population to be able to access the right and timely legal information. The current legal information systems though providing partial solutions are basically based on the search of keywords and generalized legal glossaries which do not provide the contextual subtlety of the user queries. This often creates delays, inaccuracies and limited access to good legal advice on the part of the users.

The current study attempts to overcome those difficulties by proposing a Legal Recommendation System, which combines an optimized Small Language Model (SLM) and Retrieval-Augmented Generation (RAG). The system uses structured legal information gathering, preprocessing, and retrieval by embedding to decipher natural language queries by the user and provide specific legal advice, such as names of laws, section numbers, year of enactment, and other associated cases. This mixed methodology will help make sure that the responses produced are contextually correct and have a solid foundation in authoritative legal sources, which will enhance the factor of trustworthiness as well as usability.

Other than technical innovation, the suggested system focuses on accessibility and inclusivity. It is meant to cater to a large user base who include lawyers, law firms, government agencies, universities and the general population. Its easy web-based interface means that the people with a different degree of legal knowledge may receive the correct and situation guidance of law. Additionally, the system also has a commercialization plan, which emphasizes its possible sustainability and practical implementation in the professional, academic, and population spheres.

Finally, the suggested solution does not only fill the gap between the complex legal documentation and the accessibility of the former but also introduces a scalable and adaptable legal decision support systems framework in the emerging markets. This project shows that it is possible to implement modern NLP methods in the localized legal context by first restricting to the Sri Lankan Labor and Employment Law. As refined and extended over time, this method can be applied to wider scopes of law and changed to other jurisdictions, eventually resulting in increased efficiency, equity and accessibility in the law making process.

REFERENCES

[1] P. Hemp, “Information overload,” Harvard Business Review, vol. 79, no. 5, pp. 70–76, 2001.

[2] A. V. Tambimuttu, “The judicial system of Sri Lanka,” Sri Lanka Law Reports, 2018; “An overview of the judicial system of Sri Lanka,” Sri Lanka Law Journal, 2015.

[3] Ministry of Justice, Sri Lanka, Legal documents and their management, Government Publication, 2019.

[4] S. Wijesinghe, “Sri Lankan court hierarchy and structure,” Journal of Law and Society, vol. 5, no. 1, pp. 15–27, 2020.

[5] R. Fernando, “Challenges in legal information access in Sri Lanka,” Sri Lanka Law Review, vol. 10, no. 2, pp. 100–110, 2021.

[6] R. Jayawardena, “Legal literacy and access to justice in Sri Lanka,” Law and Society Journal, vol. 10, no. 1, pp. 27–38, 2021.

[7] E. Magrani and P. G. F. da Silva, “The ethical and legal challenges of recommender systems driven by artificial intelligence,” in *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, H. Sousa Antunes, P. M. Freitas, A. L. Oliveira, C. M. Pereira, E. V. de Sequeira, and L. B. Xavier, Eds. Cham, Switzerland: Springer, 2024, pp. 231–250. DOI: 10.1007/978-3-03141264-6.

[8] A. I. Companion Study (Anonymous metadata), “Integrating RAG and ChatGPT for legal literacy and access to justice,” in *Proc. Int. Conf. on ICT for Smart Society (ICISS)*, Bandung, Indonesia, Sept. 2024. DOI: 10.1109/ICISS62896.2024.10751371.

[9] Y. Zeng, X. Xiao, Z. Xu, and S. Liu, “RoSiLC-RS: A robust similar legal case recommender system,” *Neurocomputing*, vol. 616, pp. 127–137, 2025. doi: 10.1016/j.neucom.2025.126983.

[10] S. M. W. Rahman, S. Kim, H. Choi, D. S. Bhatti, and H.-N. Lee, "Legal Query RAG: Enhancing legal recommendation with retrieval-augmented generation and recursive feedback," *IEEE Access*, vol. 12, pp. 115631–115645, 2024. DOI: 10.1109/ACCESS.2024.3456789.

[11] J. Thomas, T. Vacek, X. Shuai, W. Liao, G. Sanchez, P. Sethia, D. Teo, K. Madan, and T. Custis, "Quick Check: A legal research recommendation system," in *Proceedings of [Conference not specified]*, Thomson Reuters Research, pp. 1–10

[12] R. Mentzingen, D. Campiolo, and A. Silva, "Effectiveness in retrieving legal precedents: Exploring text summarization and language model embeddings," *Artif. Intell. Law*, pp. 1–27, 2025. doi: 10.1007/s10506-025-09440-2.

[13] W. Su, L. Zhang, and Y. Ma, "Caseformer: Pre-training for legal case retrieval based on inter-case distinctions," *arXiv preprint* arXiv:2311.00333, 2023.

[14] D. Shu, H. Zhao, X. Liu, D. Demeter, M. Du, and Y. Zhang, "LawLLM: Law Large Language Model for the US Legal System," in *Proc. 33rd ACM Int. Conf. Information and Knowledge Management (CIKM '24)*, Boise, ID, USA, Oct. 2024, pp. –. doi: 10.1145/3627673.3680020.

[15] Z. Sun, K. Zhang, W. Yu, H. Wang, and J. Xu, "Logic Rules as Explanations for Legal Case Retrieval," *arXiv preprint* arXiv:2403.01457, Mar. 2024.

[16] J. Dhanani, R. Mehta, and D. P. Rana, "Legal document recommendation system: a dictionary based approach," *International Journal of Web Information Systems*, vol. 17, no. 3, pp. 187–203, 2021. DOI: 10.1108/IJWIS-02-2021-0015.

[17] J. Dhanani, R. Mehta, and D. Rana, "Legal document recommendation system: A cluster based pairwise similarity computation," *International Journal of Web Information Systems*, vol. 17, no. 3, pp. 187–203, 2021. DOI: 10.1108/IJWIS-02-2021-0015.

[18] J. Dhanani, R. Mehta, and D. Rana, "Effective and scalable legal judgment recommendation using pre-learned word embedding," *Journal of Information and Knowledge Management*, vol. 20, no. 4, pp. 2150031-1–2150031-15, 2021. DOI: 10.1142/S0219649221500311.

[19] M. Zheng, B. Liu, and L. Sun, "LawRec: Automatic recommendation of legal provisions based on legal text analysis," *Computational Intelligence and*

Neuroscience, vol. 2022, Article ID 6313161, 10 pages, 2022. DOI: 10.1155/2022/6313161.

[20] J. Zhou, Y. Li, B. Zhang, X. Chen, X. Wei, and X. Sun, "A fast method of legal decision recommendation system," in *Proceedings of CAIL Workshop*, 2018.

[21] M. Nithya, H. S., K. S., and K. Srinidhi, "AI-Driven legal automation to enhance legal processes with natural language processing," in *Proc. Int. Conf. on Intelligent Computing and Networking Information Systems (ICICNIS)*, Bengaluru, India, Dec. 2024. DOI: 10.1109/ICICNIS64247.2024.10823316.

[22] Y. Ma, Y. Wu, W. Su, Q. Ai, and Y. Liu, "CaseEncoder: A Knowledge-enhanced Pretrained Model for Legal Case Encoding," *arXiv preprint* arXiv:2305.05393, May 2023.