A coat of arms with lions and a lion

Description automatically generated

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

**LLM-Backed Chatbot Lawyer for Enhanced Legal Services in**

**Sri Lanka**

A Project Proposal by

Mr. Omar Salman Shiraz

Supervised by

Mr. Prathieshna Vekneswaran

Submitted in partial fulfilment of the requirements for the BSc. in Computer Science degree at the University of Westminster.

**October 2023**

Table of Contents

[LIST OF TABLES iii](#_Toc147009927)

[LIST OF FIGURES iii](#_Toc147009928)

[LIST OF ABBREVIATIONS iv](#_Toc147009929)

[CHAPTER 01 – PROBLEM 1](#_Toc147009930)

[1.1 Chapter Overview 1](#_Toc147009931)

[1.2 Introduction Error! Bookmark not defined.](#_Toc147009932)

[1.3 Problem Domain 1](#_Toc147009933)

[1.3.1 Current State of Legal Services in Sri Lanka 1](#_Toc147009934)

[1.3.2 The Need for an LLM-Enhanced Chatbot Lawyer 1](#_Toc147009935)

[1.3.3 Role of Technology in Legal Services 1](#_Toc147009936)

[1.4 Problem Definition 2](#_Toc147009937)

[1.4.1 Problem Statement 2](#_Toc147009938)

[1.5 Motivation 2](#_Toc147009939)

[1.6 Existing Work 4](#_Toc147009940)

[1.7 Research Gap 5](#_Toc147009941)

[1.8 Contribution to the Body of Knowledge 6](#_Toc147009942)

[1.8.1 Contribution to Problem Domain 6](#_Toc147009943)

[1.8.2 Contribution to Research Domain 6](#_Toc147009944)

[1.9 Research Challenge 7](#_Toc147009945)

[1.10 Research Questions 7](#_Toc147009946)

[1.11 Research Aim 7](#_Toc147009947)

[1.12 Research Objectives 8](#_Toc147009948)

[1.13 Chapter Summary 9](#_Toc147009949)

[CHAPTER 02 – METHODOLOGY 11](#_Toc147009950)

[2.1 Chapter Overview 11](#_Toc147009951)

[2.2 Research Methodology 11](#_Toc147009952)

[2.3 Development Methodology 12](#_Toc147009953)

[2.3.1 Requirement Elicitation methodology 12](#_Toc147009954)

[2.3.2 Design Methodology 13](#_Toc147009955)

[2.3.3 Programming paradigm 13](#_Toc147009956)

[2.3.4 Evaluation Methodology 13](#_Toc147009957)

[2.3.4 Solution Methodology 14](#_Toc147009958)

[2.4 Project Management Methodology 15](#_Toc147009959)

[2.4.1 Project Scope 15](#_Toc147009960)

[2.4.2 Prototype Feature Diagram 17](#_Toc147009961)

[2.4.3 Schedule 18](#_Toc147009962)

[2.4.4 Resource Requirements 19](#_Toc147009963)

[2.4.5 Risks and Mitigation 22](#_Toc147009964)

[2.5 Chapter Summary 22](#_Toc147009965)

[References 23](#_Toc147009966)

# LIST OF TABLES

[Table 1: Existing Work 5](#_Toc147009596)

[Table 2 : Research Objectives 9](#_Toc147009597)

[Table 3: Research Methods 12](#_Toc147009598)

[Table 4 : Requirement Elicitation 13](#_Toc147009599)

[Table 5: Deliverables 18](#_Toc147009600)

[Table 6 : Risks and Mitigation 22](#_Toc147009601)

# LIST OF FIGURES

Figure 1: Prototype Diagram

Figure 2: Gantt Chart

# LIST OF ABBREVIATIONS

|  |  |
| --- | --- |
| **Acronym** | **Description** |
| LLM | Large Language Model |
| IT | Information Technology |
| AI | Artificial Intelligence |
| DL | Deep Learning |

# **CHAPTER 01 – PROBLEM**

## Chapter Overview

The topic "LLM-Backed Chatbot Lawyer for Enhanced Legal Services in Sri Lanka" is the cornerstone of our research project. This introductory section is crucial and provides a comprehensive overview of the scope, objectives, significance, and overall structural framework of the project. Its main purpose is to give readers a clear understanding of the challenge, the project goals, the importance of our company and the sequence of the following chapters. In addition, it is tailored to explain the motivations behind this project and give readers a clear road map to navigate through the following chapters, making it easier for them to commit to the next expedition.

## Problem Domain

### Current State of Legal Services in Sri Lanka

The accessibility and affordability of legal services in Sri Lanka have been persistent concerns. According to the World Justice Project's 2020 "Rule of Law Index," Sri Lanka ranked 65th out of 128 countries in terms of accessibility and affordability of legal services. (World Justice Project, 2020) The legal system's complexity, influenced by a mix of English common law and traditional practices, presents an additional hurdle for those seeking legal assistance (Ministry of Justice and Prison Reforms, Sri Lanka, 2018). Moreover, the distribution of legal professionals is uneven, with a concentration in urban areas, leaving rural regions underserved (Sri Lanka Bar Association, 2020).

These challenges underscore the need for innovative solutions to make legal services more accessible and comprehensible to a broader range of individuals and organizations in Sri Lanka. This project aims to contribute to this effort by exploring the integration of LLM expertise into a chatbot lawyer, with the goal of addressing these pressing issues.

### The Need for an LLM-Enhanced Chatbot Lawyer

In Lanka getting help for any legal issues or understanding the law can be a very big challenge especially for people who are in the rural areas. (Ministry of Justice and Prison Reforms, Sri Lanka, 2018) But the biggest barrier for people been the legal system been very complex to understand, due to this people get into all sort of trouble and specially people do get deceived by others. That’s where the chatbot comes into play with advanced legal knowledge the bot can offer quick affordable and reliable legal assistance specifically for Sri Lankan unique legal laws. The main goal is to make this accessible to everyone who needs it.

### Role of Technology in Legal Services

When seen the landscape of the legal service the research has noticed that there has been transformative shift driven by technological advancement. Technology has become an indispensable tool in the world of law and order. Legal professionals use this tool to streamline their process, enhance research capabilities, and improve client interactions. The integration of AI and other automation tools has revolutionized the legal sector, has the potential to offer more efficient and accessible legal services (Smith, 2022). Legal tech innovations, including chatbots and AI-powered research tools, have shown promise in simplifying legal procedures and reducing costs while maintaining the quality of legal counsel. In the context of Sri Lanka, embracing technology in legal services has the potential to bridge the justice gap and make legal assistance more readily available to diverse segments of the population.

## Problem Definition

Obtaining adequate legal aid in Sri Lanka presents considerable challenges. The country’s legal system is complex and often requires basic legal knowledge to travel efficiently. This complexity and misallocation of legal professionals has resulted in expensive legal services, especially in rural areas Globally, technological advances have transformed the legal profession, creating opportunities for access to legal aid has been increased (Smith, 2022). The project aims to solve these challenges by creating an "LLM-Backed Chatbot Lawyer" equipped with advanced legal skills. The intent of this chatbot is to provide fast, cost-effective, and customized legal guidance, and ultimately differentiate the legal services available to individuals and organizations in Sri Lanka and improve the overall legal experience.

### Problem Statement

Current lawyer chatbot are for other developed countries which has a significantly different laws compared to Sri Lanka and in order to train this chatbot using a LLM this would need an existing dataset but as per the research there no any specific dataset to train this model that’s where will take all existing books, pdfs, journal, case verdicts to create domain specific chatbot this research does not only contribute for legal advice but helps businesses and other orga nization to create domain specific chatbots

## Motivation

In recent times LLM models are having huge interest on the IT industry specifically on chatbot like GPT, Bing AI, Google Bard. These chatbot has revolutionized the industry in unimaginable ways. This opens new possibilities like academic exploration and innovation across diverse domains. The main aim in this research is to contribute to the academic discourse and reshape IT research and application. This research motivation signifies a transformative juncture in information technology.

## Existing Work

|  |  |  |  |
| --- | --- | --- | --- |
| Citation | Summary | Limitations | Contribution |
| (Martínez, 2023) | This project uses GPT-4 model to take bar exams and compares it to the previous version of GPT 3.5 and the average human | This limits to the knowledge which is given till 2021 and new verdicts and laws has been passed ever since | Through this we can conclude that GPT-4 is significantly better than an average human who take the bars and can be reliable |
| (Schwarcz and Choi, 2023) | This paper explains that how GPT-4, Bing chat, Bard can be turned into highly productive personal legal assistants. | This doesn’t give legal advice for 3rd world countries and law changes during time all these chatbots do not give the recent updates | These chatbots are capable of understanding what the user wants and can give relevant answer accordingly |
| (Firdaus, Saputra and Suprianto, 2020) | The paper presents the development of an AI-powered chatbot that assist users to understand legal frameworks so surrounding electronic information and transactions in Indonesia | This is very specific towards the one specific task and the chatbot might have poor language skills and common sense given to it | AI-powered chatbot that can assist users in understanding the complex legal framework surrounding electronic information and transactions in Indonesia. The chatbot is designed to provide information about applicable laws and can perform a search according to requests related to legal documents. |
| (Kapočiūtė-Dzikienė, 2020) | Discusses creating a domain-specific generative chatbot trained from little data | This reduces the number of datasets but does not eradicate this process | Makes a scientific contribution by exploring effective solutions to create chatbots from small datasets in English and Lithuanian languages |
| (Shalaby et al., 2020) | Building chatbots from large scale domain-specific knowledge bases, the document explores the challenges and opportunities of creating virtual assistants that can accurately understand and respond to user utterances, particularly when dealing with large volumes of domain-specific entities | These systems have size and entity limitations that can restrict the deployment of constructed knowledge bases on these platforms | Scalable framework for extracting knowledge and identifying entities in user utterances. The framework involves using a combination of manual tagging and syntactic rules to extract candidate entities, followed by a machine learning model to identify the correct entity |

Table : Existing Work

While looking at this existing work these collectives reflect on the dynamic landscape of chatbot research, showcasing innovations in AI-powered legal assistance, language understanding, and knowledge base construction. However, the limitation underscores for ongoing research to address evolving legal contexts, enhanced chatbot language skills, and have deep understanding about the domain specifics.

## Research Gap

The research gap for this project centers on the availability and maintenance of a comprehensive and dynamically updated legal knowledge base specifically tailored to the intricacies of Sri Lankan law. (Schwarcz and Choi, 2023) This challenge encompasses several critical facets: firstly, the need for legal data to be accessible in a suitable format, spanning statutes, regulations, case law, and legal precedents; secondly, addressing the dynamic nature of the legal landscape, (Martínez, 2023) requiring mechanisms for continuous updates and dynamic learning to ensure the chatbot's responses remain accurate and current; thirdly, achieving relevance and precision in responses by training the model using Sri Lankan legal content, preventing the chatbot from generating irrelevant information; and finally, incorporating verdicts and judgments from past Sri Lankan legal cases into the chatbot's training data, providing practical insights into the application of the law—a facet currently underrepresented in existing legal chatbot models. Addressing these gaps is crucial for the successful development of a precise, context-aware, and legally sound chatbot, ultimately enhancing access to justice and legal information in Sri Lanka.

## Contribution to the Body of Knowledge

### Contribution to Problem Domain

* **Continuous Updates Mechanism:** To address the dynamic nature of legal content, we devise and implement mechanisms for continuous updates and dynamic learning. This ensures that the chatbot remains current with the latest legal developments, enhancing its accuracy and relevance.
* **Contextual Relevance and Precision:** The research focuses on refining the chatbot's training process to ensure that responses are contextually relevant, precise, and legally sound. This prevents the chatbot from offering extraneous or inaccurate information and enhances its utility as a legal resource.
* **Incorporation of Verdicts:** It make a pioneering contribution by incorporating verdicts and judgments from past Sri Lankan legal cases into the chatbot's training data. This enriches the chatbot's knowledge base and provides users with practical insights into the application of the law.
* **Collaborative Interdisciplinary Effort:** The project encourages collaboration between legal experts, AI researchers, and data scientists, fostering a cross-disciplinary approach to building a reliable and effective legal chatbot for Sri Lanka.
* **Improved Access to Justice:** Ultimately, our contributions aims to enhance access to justice and legal information in Sri Lanka by providing a trustworthy and user-friendly resource for individuals seeking legal guidance.

### Contribution to Research Domain

* **Development of a Specialized Training Dataset:** The research makes a fundamental contribution by creating a specialized training dataset tailored to the unique characteristics and requirements of Specific Domain. This dataset fills a critical gap in the research domain, providing valuable resources for training chatbots to operate effectively within this specific domain. (Schwarcz and Choi, 2023)
* **Expansion of Chatbot Capabilities:** By offering a domain-specific training dataset, it can expand the capabilities of chatbots in the broader research domain. Researchers and developers can leverage this dataset to train chatbots that can comprehend, engage, and provide informed responses within Specific Domain, ultimately enhancing the utility of chatbot technology across various domains. (Shalaby et al., 2020)
* **Improved Chatbot Performance:** The availability of a domain-specific training dataset elevates chatbot performance within Specific Domain. This contribution extends to researchers and practitioners seeking to create chatbots for specific domains, providing a model for enhancing chatbot accuracy and context-awareness. (Hacker, Engel and Mauer, 2023)
* **Enhanced User Experiences:** Ultimately, the research seeks to enhance user experiences within Specific Domain by enabling the creation of chatbots that can provide more meaningful, relevant, and accurate responses. This improvement in user experiences is a valuable contribution to the broader domain of human-computer interaction. (Li, Zhang and He, no date)

## Research Challenge

The research project focused on developing the Chatbot Lawyer Sri Lanka within the realm of any domain, encountered a significant research challenge. This challenge centered around the creation of a specialized training dataset that truly reflects the intricacies and nuances of the legal landscape in Sri Lanka. (Adamopoulou and Moussiades, 2020) The grappled with the task of collecting a wide array of genuine legal conversations, questions, and interactions, each of which needed to be painstakingly annotated to serve as a reliable foundation for training chatbot. Additionally, faced complex task of ensuring that the chatbot's responses not only met the stringent legal standards of Sri Lanka but were also contextually relevant and tailored to the precise needs of the users. (Baidoo-Anu and Ansah, no date) This challenge, while demanding, was vital in not only shaping the development of the Chatbot Lawyer Sri Lanka but also in advancing the understanding of specialized language and context in the field of natural language processing, offering significant insights for the enhancement of chatbot technology in Sri Lankan law. (Firdaus, Saputra and Suprianto, 2020)

## 1.10 Research Questions

RQ 1: How to collect a diverse and representative dataset of legal conversations and queries specific to Sri Lankan law?

RQ 2: How to train the chatbot to understand and generate contextually relevant responses within the specialized domain of Sri Lankan law?

RQ 3: To what extent can domain-specific language models be beneficial in enhancing the chatbot's performance within Sri Lankan law, and how should such models be trained and fine-tuned?

RQ 4: How to preprocess Law data into a dataset which the LLM model can be trained on?

## 1.11 Research Aim

The aim of this project is to design and develop an advanced chatbot lawyer specialized for Sri Lankan law, with a focus on addressing the existing research gap domain specific dataset creation for LLM model in providing accessible and accurate legal information to the public.

## 1.12 Research Objectives

|  |  |  |  |
| --- | --- | --- | --- |
| Objective | Explanation | Learning Outcomes | Research Question |
| Problem Identification | **RO1 –** Aims to identify the specific challenges while developing a chatbot lawyer Sri Lankan law.  **RO2 –** Recognizing areas which user often seeks legal assistance, the challenges in finding reliable information. | LO4, LO5, LO2 | RQ1 |
| Literature Review | **RO3** – Extensive examination of the existing literature, research papers, legal document and chatbot development resources to gain knowledge about chatbot technology, NLP, and legal domain.  **RO4** – Systematically will review legal databases, case verdicts, journals to uncover valuable methodologies, best practices, case studies, and resources specific to Lankan law.  **RO5** – Build a strong theoretical foundation towards the project allowing it to be informed about decision-making, adoption of proven strategies and avoiding pitfalls on the development | LO1 | RQ1, RQ2 |
| Data Gathering and Analysis | **RO6** – Encompassing the collection, preprocessing noisy data analyzing the data according to the LLM acceptance dataset.  **RO7** – Develop technics to get the preprocess data from law books, case verdicts, legal journals.  **RO8** – The objective includes creation of dataset for the legal terminology using Sri Lankan law and legal documents. | LO1, LO2, LO5, LO7 | RQ4, RQ1 |
| Research Design | **RO9** – The research design cooperates for the tech stack used and chatbot architecture.  **RO10** – The objective is to create a structured plan, guiding the implementation process. | LO1, LO7 | RQ3, RQ4 |
| Implementation | **RO11** – The implementation divides the work into multiple sprints dividing the development process with continuous testing.  **RO12** – Deploying the API to messenger WhatsApp and other messaging applications | LO5, LO2 | RQ1, RQ3, RQ4 |
| Testing and Evaluation | **RO13** – To Evaluate the accuracy of the data which is represented by the chatbot accurate and reliable.  **RO14** – Doing extrusive testing to make sure the request to the server comes accordingly and gives a response according to the request | LO4 | RQ3, RQ4 |

Table : Research Objectives

## 1.13 Proposed Solution

This research will give a prototype which the user can use the bot in WhatsApp, messenger or any mainstream messaging application which support chatbots. The user will be able to message this chatbot and get the relevant legal information. Whilst training the chatbot in a domain specific steps and instructions will be given accordingly to train any chatbot for domain specific criteria.

## 1.14 Chapter Summary

This chapter explores the capabilities of Chatbot Lawyer Sri Lanka in deepening users' comprehension of Sri Lankan law. It highlights the adaptability of the LLM to tackle domain-specific issues, moving beyond generic problem-solving. Key areas of focus include dataset creation, LM training, continuous learning, user interaction design, and ethical compliance. This chapter underscores the chatbot's potential as an educational resource, enhancing users' understanding of Sri Lankan law within specific domains.

# **CHAPTER 02 – METHODOLOGY**

## 2.1 Chapter Overview

This chapter will concentrate on identifying optimal methodologies essential for fortifying research endeavors, offering a well-structured blueprint for orchestrating, and executing the Chatbot Lawyer Sri Lanka project with enhanced efficiency and productivity. The pursuit of ideal project methodologies is aimed at optimizing the utilization of limited resources, adhering to stringent project timelines, and aligning with industry best practices.

## 2.2 Research Methodology

|  |  |
| --- | --- |
| Philosophy | In the development of Chatbot Lawyer Sri Lanka, our philosophy leans towards **Pragmatism**. This philosophy underscores the commitment to adopting a balanced approach by synergizing quantitative data analysis (reflecting individual legal needs) and qualitative legal assessment (anchored in the intricacies of Sri Lankan law). This approach is meticulously crafted to harmonize with the multifaceted nature of the research, aiming to provide precise, context-aware legal assistance within the Sri Lankan legal framework. |
| Approach | The research project subscribes to a **Deductive** approach. It initiates with a well-defined theoretical framework and hypotheses, drawn from established knowledge in Sri Lankan law and the domain of conversational AI technology. Systematic data collection and analysis serve the purpose of not only validating but also enhancing this theoretical foundation, refining the capabilities of Chatbot Lawyer Sri Lanka. |
| Methodological Choice | For this project, it strategically employs a **Mixed Methodology**. This approach seamlessly blends quantitative methodologies, such as deep learning for legal query analysis, with qualitative methods tailored for interpretation and evaluation. This mixed approach is thoughtfully selected to ensure a comprehensive exploration of the research objectives, ultimately enabling Chatbot Lawyer Sri Lanka to excel in providing precise, contextually relevant legal guidance within the dynamic Sri Lankan legal landscape. |
| Strategy | The selected data collection strategy primarily revolves around **Experimentation**. This choice is steered by the project's imperative to design, develop, and deploy an advanced chatbot for legal assistance, necessitating rigorous experimentation to gather data and assess the chatbot's performance across various legal scenarios. Furthermore, it employs an Observation Strategy to gain insightful perspectives into user interactions and resource allocation dynamics. |
| Time Horizon | In this research endeavor, it embraces a **Cross-Sectional time horizon**, facilitating data collection and analysis over a defined period to effectively predict legal outcomes and evaluate the chatbot's performance within the Sri Lankan legal domain. |
| Data Collection and Analysis | **Legal Document** Analysis as the primary approach. This method systematically gathers and processes diverse legal documents, including statutes, case law, and regulations relevant to Chatbot Lawyer Sri Lanka's domain. This curated dataset informs the chatbot's knowledge base. Utilizing advanced Natural Language Processing (NLP) and Machine Learning (ML) techniques, NLP extracts insights and categorizes legal text, while ML builds predictive models, enhancing the chatbot's understanding of Sri Lankan law. This aligns with the pragmatic philosophy, using the rich legal documentation to empower the chatbot in navigating Sri Lankan law intricacies. |

Table : Research Methods

## 2.3 Development Methodology

### 2.3.1 Requirement Elicitation methodology

|  |  |
| --- | --- |
| Interviews and Surveys | Conducting user interviews and surveys is crucial because it helps to understand the specific needs and pain points of the target audience. Users often have valuable insights into their legal challenges, questions, and preferences. Gathering this information ensures that the chatbot addresses real-world user concerns and provides relevant assistance. |
| Expert Consultation | Legal experts possess in-depth knowledge of the intricacies of Sri Lankan law. Collaborating with them is essential to ensure that the chatbot provides accurate and reliable legal information. Legal experts can offer guidance on complex legal issues and help validate the quality of the chatbot's responses, increasing its credibility and usefulness. |
| Prioritization and Roadmap | Prioritizing requirements and creating a roadmap is essential for managing project resources effectively. It helps to decide which features or functionalities to tackle first, considering factors like user needs, project constraints, and dependencies. A well-planned roadmap ensures that the most critical features are delivered early, providing value sooner to users. |
| Prototyping | Prototyping allows to create a visual representation of the chatbot's user interface and functionality early in the development process. It's vital for gathering feedback from stakeholders and end-users. This feedback helps identify design flaws, user experience issues, or missing features before investing significant resources in development, reducing the risk of costly changes later. |

Table : Requirement Elicitation

### 2.3.2 Design Methodology

The design methodology involves selecting and integrating a suitable Language Model (LLM), preparing relevant legal data, fine-tuning the model with Sri Lankan legal knowledge, designing an intuitive user interface and conversational flow, ensuring data privacy and legal compliance, rigorous testing, optimizing for scalability and performance, providing comprehensive documentation and training, and finally, deploying the chatbot while continuously monitoring and maintaining its functionality and user experience.

### 2.3.3 Programming paradigm

The programming approach for the chatbot lawyer involves combining multiple methodologies to achieve its diverse functionalities. Primarily, it adopts a conversational approach to facilitate natural language interactions with users, where conversation flows are managed using state machines or dialogue trees. Additionally, an object-oriented approach is utilized to encapsulate and manage legal knowledge and user data through well-defined classes and objects. A data-driven approach plays a pivotal role in organizing and retrieving legal information from a structured knowledge base, ensuring accuracy and efficiency. Furthermore, an event-driven approach is employed to handle user inputs and system events, triggering appropriate responses or actions. These combined programming approaches enable the chatbot lawyer to provide users with seamless, context-aware legal assistance while efficiently managing and processing complex legal data and interactions.

### 2.3.4 Evaluation Methodology

**Legal Compliance Assessment**: Legal compliance assessment verifies that the chatbot's responses align with Sri Lankan legal regulations and ethical guidelines. It ensures that the chatbot does not provide legal advice beyond its capabilities, preventing potential legal liabilities, and infringement on the responsibilities of legal professionals. Compliance is essential to maintain the chatbot's credibility and legality in the context of Sri Lankan law.

**Accuracy Assessment**: Accuracy assessment evaluates how well the chatbot provides correct and reliable legal information and advice. This is crucial for ensuring that users receive trustworthy guidance. Comparing the chatbot's responses to established legal standards and collaborating with legal experts for validation are key components of this assessment.

**Privacy and Security Evaluation**: Privacy and security evaluation involves conducting a thorough security audit to identify and address vulnerabilities in data handling and storage. It ensures that user information is safeguarded and that the chatbot complies with data privacy regulations, enhancing user trust and protecting sensitive data.

**Error Handling and Resolution**: Evaluating error handling assesses how the chatbot deals with user queries it cannot answer accurately or completely. It's crucial for maintaining a positive user experience. The assessment also focuses on the chatbot's capability to appropriately escalate complex or critical queries to human experts for resolution.

### 2.3.4 Solution Methodology

This optimized methodology ensures the development of effective QoS prediction models while reducing complexity and emphasizing practicality and user engagement.

**Data Gathering and Preparation:** This step involves collecting relevant legal data and documents related to Sri Lankan law. Cleaning and preprocessing the data is crucial to ensure that it is accurate and structured appropriately. A well-organized and up-to-date dataset is essential for the chatbot to provide accurate legal information and advice to users.

**Knowledge Base Integration:** Integrating a knowledge base filled with legal information specific to Sri Lankan law is fundamental. The chatbot relies on this database to answer user queries accurately. Regular updates and maintenance are essential to keep the knowledge base aligned with changing legal regulations and precedents.

**User Interface and Experience Design:** Designing an intuitive and user-friendly interface ensures that users can interact with the chatbot easily, enhances accessibility and user satisfaction. User feedback and iterative design improvements are key to achieving an optimal user experience.

**Privacy and Security Measures:** Implementing robust data privacy and security measures is paramount to protect user information and maintain legal compliance. Regular security audits and vulnerability assessments help identify and address potential risks, ensuring the chatbot's trustworthiness.

**Testing and Quality Assurance:** Rigorous testing, including functional, performance, and security testing, is essential to verify that the chatbot meets quality standards. Engaging real users in beta testing provides valuable feedback and identifies any issues that need resolution. Comprehensive testing ensures a reliable and effective chatbot.

## 2.4 Project Management Methodology

For this research **Hybrid Agile-Waterfall Project Management Methodology** is used because this project will need to be switched from agile to waterfall when the need arises. Combining these two elements of both agile and waterfall methodologies can be effective due to the complexity of this project. They will be two parts to this project one been developing a program which creates the dataset according to the LLM model and the other been training the chatbot due to this Hybrid methodology will be used.

### 2.4.1 Project Scope

Project scope explains the project in scope what will be relevant and will be contributed to the prototype whereas the out scope will be the part where this prototype or research wont acknowledged or contributed. This will ensure the realistic boundaries of the project.

#### 2.4.1.1 In Scope

* Development of Chatbot Lawyer: The primary scope of the project is the design, development, and implementation of a chatbot lawyer specialized in providing legal information and assistance related to Sri Lankan law.
* Legal Information and Advice: The chatbot will provide accurate and reliable legal information, answer legal queries, and offer guidance on various aspects of Sri Lankan law, including statutes, regulations, case law, and legal procedures.
* Data Collection for Legal Knowledge Base: The project includes the gathering of data related to Sri Lankan law, such as statutes, regulations, court decisions, legal dictionaries, and legal glossaries. This data is essential for building a comprehensive knowledge base.
* Data Preprocessing: Preprocessing activities like data cleaning, structuring, and formatting to ensure the data is suitable for integration into the chatbot's knowledge base are within scope.
* Data Management: The project encompasses the management of legal datasets, including data gathering, organization, preprocessing, and integration into the chatbot's knowledge base.
* User Training and Documentation: User training materials and documentation will be provided to help users effectively interact with the chatbot and understand its capabilities.

#### 2.4.1.2 Out Scope

* Legal Representation: The chatbot will not provide legal representation or advice in specific legal cases. Users will be directed to consult legal professionals for individual legal matters.
* Language Translation: While the chatbot may support multiple languages, it will not provide professional translation services for legal documents or texts. It will focus on providing legal information in the selected language.
* Non-Legal Data: Gathering data unrelated to Sri Lankan law, such as general user data, non-legal documents, or non-legal glossaries, is out of scope.
* Legal Changes and Compliance: While the project aims to provide accurate legal information, it cannot guarantee real-time updates or immediate compliance with legal changes. Users should verify information with legal authorities for critical legal matters.

### 2.4.2 Prototype Feature DiagramA diagram of a system Description automatically generated

Figure 1: Prototype Diagram

### 2.4.3 Schedule

#### 2.4.3.1 Gantt Chart

Figure 2: Gantt Chart

#### 2.4.3.2 Date of Deliverables

|  |  |
| --- | --- |
| **Final Project Tentative Timeline** | |
| 1st August 2023 | Fill the google form for Tentative Idea and finalized supervisor |
| 5th October 2023 | Project Proposal Submission |
| 31st October 2023 | LR Submission |
| 27th November 2023 | SRS Submission |
| 1st February 2024 | PSPD Submission |
| 7th March 2024 | Prototype |
| 4th April 2024 | Thesis |

Table : Deliverables

### 2.4.4 Resource Requirements

#### 2.4.4.1 Hardware Requirements

* Processor: A high-performance CPU with multiple cores.
* Graphics Processing Unit (GPU) or Tensor Processing Unit (TPU): Multiple powerful GPUs or TPUs, especially for large-scale LLMs.
* RAM (Random Access Memory): Several hundred gigabytes to terabytes of RAM, depending on model size and dataset.
* Storage Drives: High-speed SSD or NVMe storage drives for quick data access.
* Distributed Computing Infrastructure: For extremely large LLMs and datasets, a cluster with multiple GPUs, CPUs, and distributed storage.
* Cooling Solutions: Robust cooling systems to prevent overheating.
* Power Supply: A stable and sufficient power supply for uninterrupted training.
* Network Connectivity: High-speed internet for data transfer.
* Rack or Data Center Space: For large-scale training and hardware housing.
* Backup and Redundancy: Data backup and component redundancy for reliability and data protection.

#### 2.4.4.2 Software Requirements

* Operating System: A Linux-based operating system (e.g., Ubuntu, CentOS) is commonly used for training LLMs due to its performance and compatibility with deep learning frameworks.
* Python: Python is the primary programming language for machine learning and natural language processing. Python 3.6 or higher is recommended.
* Deep Learning Frameworks: Install deep learning frameworks such as TensorFlow or PyTorch to build and train LLMs efficiently. These frameworks provide GPU support for accelerated training.
* CUDA and cuDNN: If using Nvidia GPUs, install CUDA and cuDNN libraries to harness the full potential of GPU acceleration.
* Version Control (Git): Git is essential for version control and collaborative development. It helps track changes to code, configurations, and datasets. Git is a valuable tool for managing LLM training projects.
* Development Environment: Set up a virtual environment (e.g., Anaconda or virtualenv) to manage dependencies and isolate the project environment.
* Text Editors or Integrated Development Environments (IDEs): Choosing a text editor or IDE of preference (e.g., Visual Studio Code, PyCharm) for coding and development.
* Data Management Tools: Use data management tools such as pandas for data preprocessing and manipulation.
* Jupyter Notebooks: Jupyter notebooks are useful for interactive development, experimentation, and documentation of LLM training workflows.
* BibGuru: Install and use BibGuru for managing references and citations, which can be essential for academic research and documenting sources.
* Package Managers: Use package managers like pip or conda to install and manage Python libraries and packages required for the project.

#### 2.4.4.2 Data Requirements

There’s no dataset for the project as of now the dataset will be created using the following documents.

* Legal Texts and Documents: Access to a wide range of Sri Lankan legal texts, including statutes, regulations, court decisions, and legal precedents, is essential. These documents serve as the foundation for the chatbot's knowledge base and legal information.
* Case Law Databases: A comprehensive database of Sri Lankan case law is crucial for providing legal advice and references to previous court decisions. This data helps users understand how legal principles are applied in real-world situations.
* Legal Dictionaries and Glossaries: A collection of legal terminology, definitions, and glossaries specific to Sri Lankan law is necessary for interpreting legal jargon and providing clear explanations to users.
* Legal Texts and Documents: Access to a wide range of Sri Lankan legal texts, including statutes, regulations, court decisions, and legal precedents, is essential. These documents serve as the foundation for the chatbot's knowledge base and legal information.
* Case Law Databases: A comprehensive database of Sri Lankan case law is crucial for providing legal advice and references to previous court decisions. This data helps users understand how legal principles are applied in real-world situations.
* Legal Dictionaries and Glossaries: A collection of legal terminology, definitions, and glossaries specific to Sri Lankan law is necessary for interpreting legal jargon and providing clear explanations to users.

#### 2.4.4.3 Skill Requirements

* Legal Texts and Documents: Access to a wide range of Sri Lankan legal texts, including statutes, regulations, court decisions, and legal precedents, is essential. These documents serve as the foundation for the chatbot's knowledge base and legal information.
* Case Law Databases: A comprehensive database of Sri Lankan case law is crucial for providing legal advice and references to previous court decisions. This data helps users understand how legal principles are applied in real-world situations.
* Legal Dictionaries and Glossaries: A collection of legal terminology, definitions, and glossaries specific to Sri Lankan law is necessary for interpreting legal jargon and providing clear explanations to users.
* Programming Languages: Proficiency in Python, as it's commonly used for chatbot development and natural language processing (NLP). Knowledge of JavaScript, whilst developing a web-based user interface.
* Natural Language Processing (NLP): Understanding of NLP concepts and libraries like NLTK or spaCy for text processing and analysis.
* Machine Learning and Deep Learning: Familiarity with machine learning algorithms and frameworks such as TensorFlow or PyTorch for building and fine-tuning language models.
* Chatbot Development: Experience in developing chatbots using frameworks like Rasa, Dialogflow, or custom solutions.
* Data Management: Ability to manage and preprocess legal data, often in structured and unstructured formats.
* Database Management: Knowledge of database systems like MySQL or MongoDB for storing and retrieving legal information.
* Version Control: Proficiency in Git for managing code versions and collaboration.

### 2.4.5 Risks and Mitigation

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Probability of Occurrence** | **Magnitude of Loss** | **Mitigation Plan** |
| Data Privacy Breach | Moderate | High | Implement robust data encryption and access controls. Regularly audit and update security measures. |
| Inaccurate Legal Information | Moderate | High | Continuously update the knowledge base with the latest legal information and engage legal experts for validation. |
| Hardware requirements | High | High | Making use of remote and cloud computers like google colab Microsoft azure to train model. |
| Data Inaccuracies in Knowledge Base | Moderate | Moderate | Implement automated data validation processes and user feedback mechanisms for data accuracy checks. |
| Ethical Concerns | Low | Moderate | Establish strict ethical guidelines and consult legal experts to ensure adherence to ethical standards. |

Table : Risks and Mitigation

## 2.5 Chapter Summary

In this chapter, it evaluates on developing a Sri Lankan law focused chatbot. The research introduces the project to emphasizes its potential benefits. The research explores various requirement and methodologies to follow before proceeding in development. Project management technics to make sure an efficient and proper documented approach is taken whilst using these technics. Aligning them with constraints, deadlines, and expertise, with plans to use the most suitable approach in the coming months.

# References

Adamopoulou, E. and Moussiades, L. (2020) “An overview of chatbot technology,” in IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing, pp. 373–383.

Baidoo-Anu, D. and Ansah, L. O. (no date) Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning, Org.tr. Available at: https://dergipark.org.tr/en/download/article-file/3307311.

Botero, J. C. et al. (no date) The world justice project, Worldjusticeproject.org. Available at: https://worldjusticeproject.org/sites/default/files/documents/WJP-ROLI-2020-Online\_0.pdf.

Budzianowski, P. and Vulic, I. (no date) Towards the use of pretrained language models for task-oriented dialogue systems, Arxiv.org. Available at: http://arxiv.org/abs/1907.05774.

Courts, P. R. S. (2018) MINISTRYOFJUSTICEAND, Gov.lk. Available at: https://www.moj.gov.lk/images/pdf/progress\_report/2018/English.pdf.

Firdaus, V. A. H., Saputra, P. Y. and Suprianto, D. (2020) “Intelligence chatbot for Indonesian law on electronic information and transaction,” IOP conference series. Materials science and engineering, 830(2), p. 022089. doi: 10.1088/1757-899x/830/2/022089.

Hacker, P., Engel, A. and Mauer, M. (2023) “Regulating ChatGPT and other Large Generative AI Models,” in 2023 ACM Conference on Fairness, Accountability, and Transparency. New York, NY, USA: ACM.

Haristiani, N. (2019) “Artificial intelligence (AI) chatbot as language learning medium: An inquiry,” Journal of physics. Conference series, 1387(1), p. 012020. doi: 10.1088/1742-6596/1387/1/012020.

Kapočiūtė-Dzikienė, J. (2020) “A domain-specific generative chatbot trained from little data,” Applied sciences (Basel, Switzerland), 10(7), p. 2221. doi: 10.3390/app10072221.

Li, C., Zhang, M. and He, Y. (no date) The stability-efficiency dilemma: Investigating sequence length warmup for training GPT models, Neurips.cc. Available at: https://proceedings.neurips.cc/paper\_files/paper/2022/file/aac02401755a65904cf977a33136af4a-Paper-Conference.pdf (Accessed: October 3, 2023).

Martínez, E. (2023) “Re-evaluating GPT-4’s bar exam performance,” SSRN Electronic Journal. doi: 10.2139/ssrn.4441311.

Schwarcz, D. B. and Choi, J. H. (2023) “AI tools for lawyers: A practical guide,” SSRN Electronic Journal. doi: 10.2139/ssrn.4404017.

Shalaby, W. et al. (2020) Ieee.org. Available at: https://ieeexplore.ieee.org/abstract/document/9187036#citations.

Smith, M. (2022) Researchgate.net. Available at: https://www.researchgate.net/publication/360344582\_Technology\_Law\_in\_Legal\_Education\_Recognising\_the\_Importance\_of\_the\_Field.

Sri Lanka Bar Association (2020) Basl.lk. Available at: https://basl.lk/assets/pdf/BASL%20Annual%20Report%202019-2020.pdf.