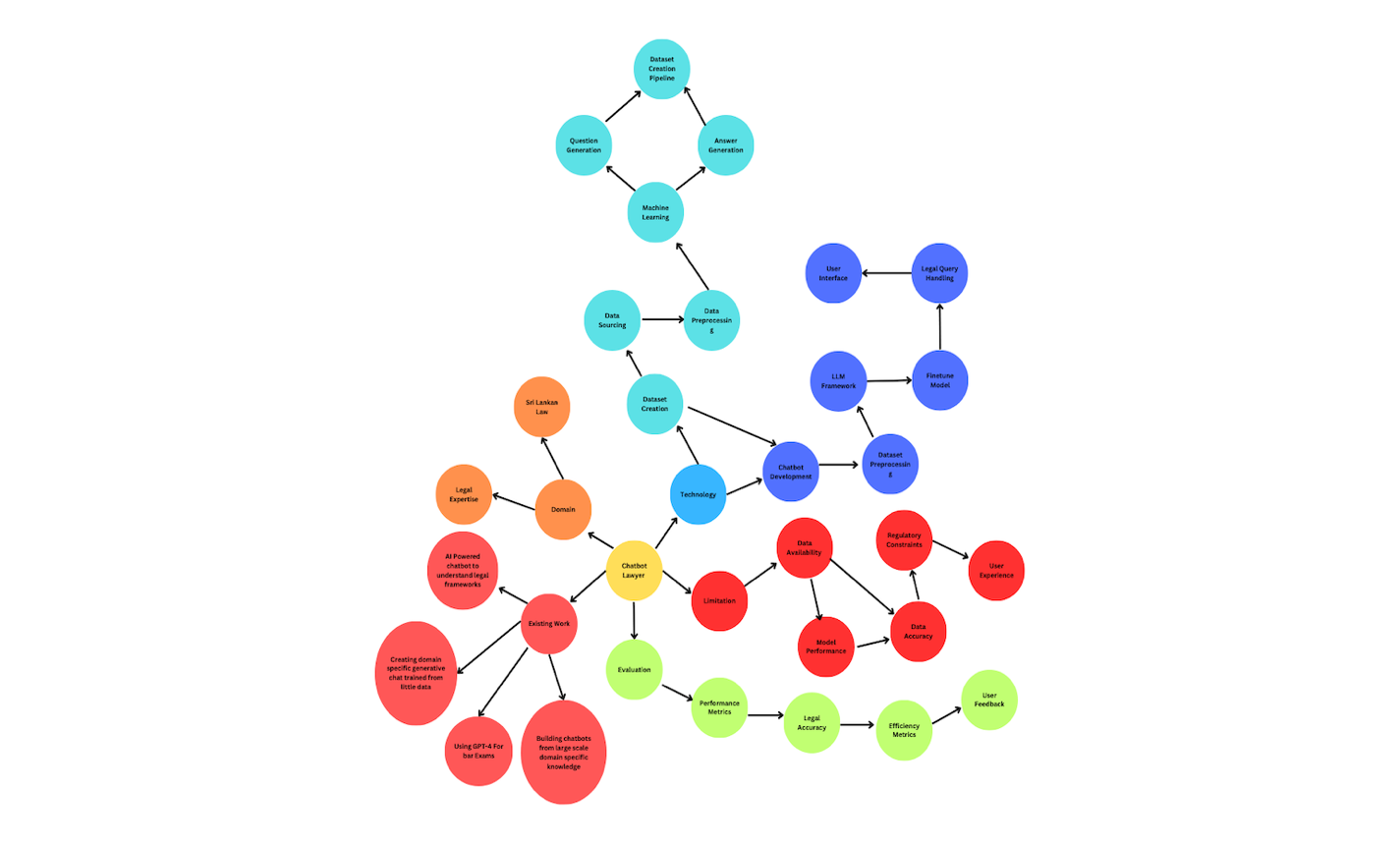
# **Chapter 2 - Literature Review**

## **2.1 Chapter overview**

Law, As indicated in the previous chapter is a very exclusive knowledge in order break this barrier. This chapter provides a comprehensive examination of the pivotal data preparation and transformation of Sri Lankan juridical documents. Additionally, this chapter discusses strategies for generating question-answer pairs from legal text, and the formatting required to construct a domain-specific dataset. Using this created dataset, the chapter will also focus on how the training and testing of the LLM will proceed.

## **2.2 Concept Map**



## **2.3 Problem Domain**

### 2.3.1 Legal Expertise

Developing a chatbot which is specialized in Sri Lankan law required a profound understanding about the legal domain to ensure reliable accuracy for the chatbot and should compliance with legal regulations. Legal expertise is the way of guiding into the development process. Legal professionals bring an invaluable depth of knowledge to the table, and with their involvement this significantly improves the chatbots performance and accuracy.

One of the most critical aspects of incorporating legal expertise in chatbot development is ensuring legal accuracy. Legal chatbots are often relied upon to provide legal advice and information. Legal experts, such as lawyers, are well-equipped to validate the accuracy of the chatbot's responses, thus reducing the risk of providing incorrect or misleading legal guidance to users. (Adamopoulou and Moussiades, 2020)

When considering Sri Lankan law there are various range of regulations and standards that may change or evolve over time. With this familiarity it ensures that the chatbot responses aligned with current legal standards, which minimizes the potential for legal issues and bolstering the chatbots credibility. (Sri Lanka Bar Association, 2020)

Additionally, legal experts possess a unique ability to interpret complex legal documents, such as statutes, case law, and legal opinions. Their expertise allows them to navigate intricate legal texts, extract pertinent information, and provide accurate explanations. (Sri Lanka Bar Association, 2020)

The trustworthiness of legal expertise and the unbiased law documents is a paramount for the legal chatbot. Legal expertise ensures this trust by engaging in the data preparations for the development cycle. This trust and unbiased law document will boost user confidence while using this legal chatbot which will enhance the chatbot’s utility.

Furthermore, legal professionals can offer valuable guidance on ethical considerations in chatbot development. This guidance is essential for upholding the integrity of the chatbot's interactions with users and preserving ethical standards within the legal community. (Courts, 2018)

In conclusion, the integration of legal expertise is foundation in the development of a chatbot which is specialized in Sri Lankan law. With legal expertise guidance this chatbot can uphold legal accuracy, regulatory compliance, user trust, and ethical considerations. Legal professionals contribute significantly to the chatbot's credibility, ensuring that it provides sound and dependable legal information to users.

### 2.3.2 Domain Specific Dataset Creation

When developing a chatbot using LLM the major problem is finding a dataset for the specific domain to eradicate this process domain specific dataset creation was bought into developing this chatbot. To create this dataset for the specific domain Large Language Model will be needing relatively high amount of data in order to fine tune the chatbot to work efficiently and get relatively high amount of accuracy in order to get any use of the chatbot. Initiating this approach is the collaborative effort between legal experts and data scientists in the careful sourcing of legal documents. The dataset draws from a diverse range, including statutes, case law, legal texts, and historical legal documents, mirroring the comprehensive nature of Sri Lankan law. (Baidoo-Anu and Ansah, no date)

The application of advanced Natural Language Processing (NLP) techniques is an essential process during the data extraction phase. Techniques like text parsing and document layout analysis are harnessed to transform the often intricate and unstructured legal text into a well-organized dataset in order to get high accuracy for a LLM model the data needs to be labeled and the LLM model needs to be guided using the dataset. This transformation is not merely a technical process but a strategic move to ensure that the dataset encapsulates the subtleties and complexities intrinsic to Sri Lankan legal discourse, setting the stage for nuanced machine learning applications. (Kapočiūtė-Dzikienė, 2020)

The collaboration between legal experts and data scientists reaches a zenith in the generation of question-answer pairs. This is not merely a data preparation exercise; it is a fusion of legal acumen and machine learning intricacies. Legal professionals work closely with data scientists to craft a rich set of questions and meticulously researched answers, infusing the dataset with a depth of legal understanding that is crucial for the chatbot and LLM to navigate and respond to a spectrum of legal inquiries with precision. After giving the model an idea of how to the pair needs to be generated, we just need to pass unstructured chunks of text through the model and generate pairs

The formatting of the dataset for machine learning compatibility is another crucial aspect where legal expertise shines. Through this phase, the dataset is tailored to meet the specific requirements of the LLM model, ensuring optimal performance in comprehending and responding to the diverse legal queries it is expected to encounter. (Schwarcz and Choi, 2023)

In conclusion this domain specific dataset creation creates question and answer pairs and context which will then be turned into the instruction column of the dataset along with input and output columns which are question and answer respectively. This will then be given to domain experts to cross verify.

## **2.4 Existing Work**

### 2.4.1 AI TOOLS FOR LAWYERS: A PRACTICAL GUIDE

This existing work specifically focuses how well does generative AI perform in real world bar exams which lawyers take to be a certified lawyer. This research specifically focuses on existing trained models like GPT-4, Bard and Bing chat giving them bar exam questions to do and evaluated by professionals on the topic. While these models already have the necessary knowledge to outperform average scores scored by students who take the bar exams every year. Whereas this proves that the model can perform as well as real world lawyers when the evaluation was done certain question was asked by GPT-4 and the responses was given to evaluate with GPT-4 excels at explaining legal rules governing well-known issues of federal law that feature prominently in the corpus of text on which it was trained. To illustrate, suppose a lawyer’s client is sued for defamation in connection with reporting they performed for a prominent newspaper, and they are looking to brush up on the basic rules governing this area. A good first step would be to ask GPT-4 about the rules governing this issue, as illustrated below. (Schwarcz and Choi, 2023)

What legal source material do these rules come from?

GPT-4’s response: “I am not a lawyer, but I can provide you with some general information on this topic. A reporter can be held liable for defamation if they publish or broadcast false information that causes harm to someone's reputation. There are some key elements to consider when determining if a reporter can be held liable for defamation:1. False statement: The reporter must have made a false statement of fact, not just an opinion or commentary. A true statement, even if harmful to someone's reputation, is generally not considered defamatory.2. Identification: The defamatory statement must identify the person being defamed, either directly or indirectly. This could include using the person's name, image, or providing enough information that they can be identified by others.3. Publication: The statement must be published or broadcast, meaning it has been communicated to a third party. This could be in a newspaper, online article, TV broadcast, or even social media.4. Fault: The reporter must be at fault for publishing the false statement. In the case of private individuals, the standard is usually negligence, meaning the reporter failed to exercise reasonable care in verifying the information. For public officials or public figures, the standard is typically "actual malice," which means the reporter either knew the statement was false or acted with reckless disregard for the truth.5. Damages: The person claiming defamation must prove that they suffered harm as a result of the false statement. This could include damage to their reputation, financial loss, or emotional distress. Laws regarding defamation can vary by jurisdiction, and there may be additional nuances and exceptions to consider. Anyone who believes they have been defamed should consult with a qualified attorney to understand their rights and potential remedies.”

GPT -4’s answer: “These rules for defamation liability stem from a combination of common law principles and statutory law, which can vary depending on the jurisdiction. In the United States, the key legal sources for defamation law include: (1) Common law: Defamation has its roots in English common law, which is a body of law developed through court decisions over time. Many U.S. states have adopted and adapted these common law principles in their own legal systems, and state courts continue to refine and develop defamation law through their decisions. (2) State statutes: Some states have enacted statutory laws regarding defamation, which may define elements of defamation, set forth defenses or exceptions, or specify remedies and damages. These statutes can vary from state to state and may modify or supplement common law principles. (3) U.S. Constitution: The First Amendment to the U.S. Constitution guarantees the right to freedom of speech and the. (Schwarcz and Choi, 2023)

The above explanation was given by the research papers this shows how GPT-4 is able to give an explanation for the general public who aren’t capable of understanding certain law terms but giving with this we can come to a conclusion that these LLM model are capable of giving an idea to general public who aren’t aware and further these chatbots aren’t specifically trained to do this task which is generally and yet is been reliable resource where the general public is capable of using this to be aware of certain rules and regulation which can save immense amount of fines and prison times for the general public as well as the society to live a peaceful life.

What are the limitations for this, this performing well but all these questions are based on US law which is giving convincing answers but when tried this on Sri Lankan law it gave not convincing answers and started to give wrong information about it. In Sri Lanka laws are static and when gave to professionals to evaluate and ask certain questions about Sri Lankan law the professionals were able to detect the outputs the chatbot gave aren’t correct and are not up to date law changes with time and certain laws come when new cases and new verdicts are being bought this includes government changes. With this it concludes the general chatbots can’t give domain specific answers as accurately as it gives general answers.

Comparatively, the proposed project diverges by tailoring the application of LLMs to the specific legal domain of Sri Lanka. While the existing work explores the broader potential of LLMs without geographical constraints, the project dedicates itself to extracting, transforming, and utilizing data from Sri Lanka's legal documents. This geographical specificity enhances the relevance and accuracy of the legal information provided by the chatbot, addressing the unique legal landscape of the country.

Notably, the project places a distinct focus on creating a domain-specific dataset, a facet not explicitly detailed in the existing work. Through meticulous crafting of a dataset tailored to the legal nuances of Sri Lanka, the project aims to ensure that the chatbot's responses are contextually relevant and aligned with the specific legal requirements of the country.

In summary, while the existing work provides a broad perspective on the application of LLMs in legal research and writing, the proposed project sets itself apart through its country-specific focus and the creation of a specialized dataset. This approach enhances the accuracy and relevance of the legal information delivered by the chatbot and addresses limitations related to the specificity of legal contexts, contributing to a more effective and contextually aware legal assistant.

**Limitations:**

**Dependency on Model Accuracy:** The efficacy of the AI large language models is contingent on their accuracy, and inaccuracies may persist, potentially leading to misinformation in legal analyses.

**Generalization Challenges:** LLMs, despite their advancements, may struggle with nuanced legal scenarios, leading to challenges in providing contextually relevant information for highly specialized legal queries.

**Ethical Considerations:** The ethical implications of relying extensively on AI models in legal practice need careful consideration, especially regarding issues of accountability, bias, and transparency.

**Limited Adaptability:** While LLMs can excel in specific tasks, their adaptability to the evolving legal landscape and the ability to handle unforeseen legal developments may be limited.

**Risk of Over-Reliance:** There is a risk of over-reliance on LLMs, potentially diminishing the development and exercise of critical legal thinking skills among lawyers and law students.

While AI large language models (LLMs) like GPT-4 offer transformative possibilities in legal research and writing, it is imperative to acknowledge their limitations. The dependency on model accuracy, challenges in nuanced legal scenarios, ethical considerations, limited adaptability, and the risk of over-reliance underscore the need for a cautious and balanced approach. Embracing LLMs should be complemented by the preservation of traditional legal skills and a vigilant awareness of ethical implications. Striking this balance is crucial for harnessing the benefits of AI in the legal domain while safeguarding the core tenets of legal practice.

### 2.4.2 Re-Evaluating GPT-4’s Bar Exam Performance

The analysis of GPT-4's performance on the Uniform Bar Exam (UBE) raises substantial concerns about the accuracy of its reported 90th-percentile score. The investigation, which considers diverse scenarios such as repeat test-takers and a broader UBE participant pool, suggests a potential overestimation of GPT-4's actual capabilities compared to the challenges encountered in real-world legal practice.

One pivotal aspect of the existing work that deserves attention is the scoring methodology employed in evaluating GPT-4's performance. The reported 90th-percentile score on the UBE, with an 80-percentile-points boost over its predecessor GPT-3.5, is a significant benchmark. However, this score may be subject to methodological scrutiny, especially when examined in the context of the specific populations considered, such as repeat test-takers.

The existing work's limitations become apparent when dissecting the geographical specificity of its focus, primarily centered on U.S. law. This limitation implies that the reported scores are grounded in the legal intricacies of the U.S. jurisdiction, potentially limiting the generalizability of GPT-4's performance to a global legal context.

Furthermore, the dynamic nature of legal landscapes poses a challenge for the existing work's scoring accuracy. Given the constant evolution of laws, new cases, and legislative changes, the reported score may be influenced by the static nature of the legal information upon which GPT-4 was trained. This could impact the model's ability to adapt to the evolving legal scenarios encountered in real-world practice.

In contrast, the Sri Lankan Chatbot Lawyer project's scoring methodology is strategically designed to account for these limitations. By tailoring responses to the specific nuances of Sri Lankan law, the project ensures that users receive contextually relevant and accurate legal information within the framework of Sri Lanka's legal system. The scoring system considers the intricacies of Sri Lankan law, reflecting a nuanced understanding that contributes to the chatbot's reliability and applicability in the local context.

Moreover, the project's approach to continuously updating its dataset from Sri Lankan legal documents is a crucial factor in enhancing the accuracy of its scores. This dynamic adaptation ensures that the chatbot remains informed about the latest legal developments, contributing to the precision of its scoring in providing up-to-date legal guidance.

**Limitation:**

**Geographical Specificity:** The focus on U.S. law restricts the generalizability of the reported scores to a global legal context, as the model's training is centered on the intricacies of a specific jurisdiction.

**Population Bias:** The scoring may be influenced by a bias towards repeat test-takers, potentially skewing the percentile performance assessment and limiting the applicability of the scores to a broader population.

**Static Legal Information:** The reported scores may be impacted by the static nature of the legal information upon which GPT-4 was trained. This static foundation could affect the model's adaptability to the dynamic and evolving legal scenarios encountered in real-world practice.

**Limited Adaptability:** The scoring methodology may not effectively capture the adaptability of GPT-4 to unforeseen legal developments, potentially impacting its performance in rapidly changing legal landscapes.

**Lack of Domain Specificity:** The existing work's scoring does not account for domain specificity, potentially leading to inaccuracies when applied to specialized legal contexts beyond the scope of its training.

In conclusion, while the existing work provides a score that showcases GPT-4's capabilities on the UBE, it is essential to critically examine the scoring methodology and its limitations. The Sri Lankan Chatbot Lawyer project, with its meticulously crafted scoring system, offers a more tailored, reliable, and context-aware resource for users seeking precise legal assistance in Sri Lanka.

### 2.4.2 Creating domain specific chatbot using IBM Watson

This paper unfolds the narrative of harnessing IBM Watson tools to fabricate a domain-specific chatbot, intricately navigating the complexities of tool accuracy assessment and presenting an in-depth exploration of their merits and demerits. Simultaneously, the discourse extends its tendrils into methodologies aimed at refining the overarching chatbot system, shedding light on indispensable architectural tools that form the bedrock of its developmental structure.

However, beneath the veneer of this research lies a terrain marked by certain notable limitations. The touted generic approach championed by the existing work exposes a misalignment with the nuanced demands of specific domains, especially those that demand a profound comprehension of intricate legal frameworks—a realm my project subtly inhabits. The absence of a tailored approach raises legitimate concerns regarding the precision and applicability of responses, particularly in fields as nuanced and context dependent as law.

Adding a layer of intricacy, the existing work overlooks the pragmatic aspect of implementation on a localized scale. The lack of consideration for local applicability and resource optimization renders it potentially less accessible and more resource intensive. Contrarily, a project of similar nature strategically embraces open-source tools and prioritizes the development of a domain-specific pipeline designed for local implementation. This not only amplifies accessibility but also positions it as a cost-effective solution, a critical consideration for developers operating within the constraints of budgetary parameters.

Furthermore, delving into the subtleties of language and context within the legal domain, the generic approach fails to encapsulate the nuanced intricacies of legal language, potentially compromising the accuracy and relevance of responses. In stark contrast, a project akin to mine implicitly aligns with the specificities of a designated jurisdiction. This approach ensures that responses are not merely accurate but also resonate with the contextual nuances of the local legal landscape. The conscious emphasis on local relevance and the strategic integration of open-source accessibility not only make this approach pragmatic but also positions it as a compelling choice for developers aspiring to tailor chatbot solutions to the idiosyncrasies of domains. (Bansal, 2021)

**Limitation:**

**Generic Approach:** The existing work adopts a generic approach that may not seamlessly align with the nuanced requirements of specific domains, particularly those involving intricate legal frameworks.

**Lack of Tailored Precision:** The absence of a tailored approach raises concerns about the precision and applicability of responses, especially in fields as context dependent as law.

**Overlooking Local Implementation:** The research overlooks the pragmatic aspect of localized implementation, potentially rendering it less accessible and more resource intensive.

**Resource-Intensive Nature:** The generic approach may require more resources, making it less feasible, particularly for developers operating within budgetary constraints.

**Insufficient Consideration for Local Nuances:** The generic approach fails to capture the subtleties of language and context within specific legal domains, compromising the accuracy and relevance of responses.

**Dependency on Proprietary Tools:** The reliance on proprietary tools, such as IBM Watson, poses potential financial barriers and security concerns, limiting accessibility for developers with budget constraints.

In concluding, while the existing work valuably explores the utilization of IBM Watson tools for domain-specific chatbots, its generic approach reveals limitations in precision, localized implementation, and resource efficiency. The subtle intricacies of legal language and context are often overlooked. Contrarily, a project akin to mine strategically embraces open-source tools, prioritizing local relevance, and cost-effectiveness. The nuanced approach, emphasizing tailored solutions within specific domains, positions itself as a pragmatic choice for developers navigating the complexities of specialized applications.

## **2.5 Technological Review**

In the technological review of the Chatbot Lawyer project, a pioneering approach is adopted to harness the power of Large Language Models (LLMs) in tandem with Questgen. The core technological underpinning revolves around training the LLM, a large neural network capable of understanding and generating human-like language, with a meticulously crafted dataset. This dataset, a product of innovative data generation techniques, is tailored specifically to the requirements of the LLM, ensuring a comprehensive grasp of the nuances inherent in Sri Lankan legal documents. Leveraging Questgen, a question generation library, the project introduces a dynamic element by generating questions and answers to construct a domain-specific dataset. This approach not only elevates the dataset's relevance to the legal context but also streamlines the training process for the LLM, paving the way for a sophisticated and contextually aware Chatbot Lawyer. The integration of cutting-edge technologies in this review underscores the project's commitment to advancing the capabilities of legal chatbots within the unique landscape of Sri Lankan law.

### 2.5.1 Dataset

In the absence of readily available datasets for Sri Lankan law, a groundbreaking approach is employed for dataset creation in the Chatbot Lawyer project. The dataset is meticulously crafted through collaboration with legal experts and practitioners, ensuring the inclusion of authentic and relevant legal content. Legal professionals, including lawyers specializing in Sri Lankan law, actively contribute to the dataset creation process, validating the accuracy and contextual appropriateness of the legal documents. This collaborative effort not only guarantees the authenticity of the dataset but also captures the nuances and subtleties that are inherent in legal language. The involvement of legal experts in the validation process ensures that the dataset accurately represents the complexities of Sri Lankan legal documents. This innovative methodology not only addresses the absence of existing datasets but also establishes a robust foundation for training the Large Language Model, reinforcing the project's commitment to creating a highly accurate and contextually aware Chatbot Lawyer.

### 2.5.2 Preprocessing of Collected Dataset

The preprocessing phase in the Chatbot Lawyer project involves a meticulous workflow tailored to the unique nature of legal documents. The process initiates with the collection of diverse legal materials, spanning constitutional texts and historical case verdicts, encapsulating the rich tapestry of Sri Lankan law. These documents, obtained in PDF format, undergo a sophisticated extraction process designed to remove extraneous elements, such as table of contents, headers, and footers, ensuring a streamlined and focused dataset.

Following this initial refinement, the legal documents traverse through a question-and-answer generation model, a pivotal step in transforming static content into an interactive and dynamic format. This model, powered by Questgen, dynamically generates pertinent questions and their corresponding answers, injecting an interactive element into the dataset. This iterative process not only enriches the dataset with contextually relevant queries but also aligns the content with potential user inquiries.

The culmination of these preprocessing steps results in a curated dataset, finely tuned to the intricacies of Sri Lankan law. This dataset, now imbued with dynamic question-answer pairs, serves as the cornerstone for training the Large Language Model, contributing to the development of a sophisticated and contextually aware Chatbot Lawyer. The innovative approach to preprocessing not only addresses the challenges of legal document structure but also enhances the dataset's depth and interactivity, reinforcing the project's commitment to delivering a comprehensive and responsive legal chatbot experience.

### 2.5.3 Feature Engineering, Extraction, and Selection

In developing the Chatbot Lawyer, we're making sure it really understands Sri Lankan law. We start by carefully picking features – things the chatbot will pay attention to. These features go beyond just words and include specific legal details. Then, we extract important information from our data, focusing on phrases and legal points. At the same time, we select the most critical features, using smart techniques to choose what really helps the chatbot understand and respond well. This approach ensures the chatbot knows the ins and outs of Sri Lankan law, balancing detailed understanding with efficiency. Importantly, the chatbot is already equipped with English proficiency and common sense, making legal knowledge just one part of its broader capabilities.

### 2.5.4 Algorithm Selection

In the algorithm selection phase of the Chatbot Lawyer project, we meticulously choose the computational methods that power the chatbot's understanding and responsiveness. The algorithms selected are tailored to process and analyze the nuanced features extracted from legal documents while aligning with the broader capabilities of the Large Language Model (LLM). These algorithms play a crucial role in decoding legal language intricacies specific to Sri Lankan law. The selection process prioritizes efficiency, accuracy, and adaptability to ensure the chatbot delivers contextually relevant and precise responses. The chosen algorithms synergize with the feature set, contributing to the chatbot's ability to navigate and interpret the complexities of legal language within the unique landscape of Sri Lankan legal documents.

### 2.5.5 Training and Testing

The training and testing phase in the Chatbot Lawyer project marks a crucial stage in refining the model's capabilities. Leveraging the selected algorithms, the chatbot undergoes a training process where it learns to understand the intricacies of Sri Lankan legal language, utilizing the curated dataset and feature set. This training ensures the model aligns its responses with the contextual nuances specific to the legal domain. Subsequently, rigorous testing is conducted to evaluate the chatbot's performance, assessing its ability to generate accurate and relevant responses to a diverse array of legal queries. The iterative nature of this process allows for fine-tuning, ensuring the chatbot continuously evolves and adapts to the complexities of real-world legal inquiries. The training and testing phase serves as a pivotal step in enhancing the chatbot's proficiency and responsiveness within the unique landscape of Sri Lankan law.

### 2.5.6 Training and Testing

In the Chatbot Lawyer project, evaluating the model involves a comprehensive analysis using various metrics to measure its performance and effectiveness. The chosen evaluation metrics go beyond mere correctness and delve into the nuanced aspects of the chatbot's responses. Precision, recall, and F1 score are employed to gauge the accuracy of legal information provided, ensuring a balance between correctness and completeness. Additionally, metrics such as perplexity and coherence are considered to assess the model's language understanding and its ability to generate coherent responses. The evaluation process is designed to capture the intricacies of legal language comprehension, reflecting the project's commitment to delivering a chatbot that not only provides correct answers but also aligns with the nuanced expectations of Sri Lankan legal discourse.

## **2.6 Literature Review Summary**

The literature review delves into the intricacies of developing a Chatbot Lawyer tailored to the legal landscape of Sri Lanka. The problem domain encompasses the creation of a domain-specific dataset, marrying machine learning with legal expertise. Addressing this challenge involves a meticulous pipeline for extracting, transforming, and generating questions and answers from legal documents. This not only contributes to dataset creation but also serves as a foundation for training a Large Language Model (LLM).

Existing work is explored, notably a study using GPT-4 for legal analysis, emphasizing its 90th-percentile performance on the Uniform Bar Exam. However, limitations are uncovered, questioning the validity of claims, and underscoring the importance of rigorous evaluations in AI model development.

The technological review unravels the intricacies of feature engineering, extraction, selection, algorithm selection, and model evaluation. These steps intricately shape the Chatbot Lawyer, ensuring it not only understands legal language but also leverages a feature-rich dataset and algorithms aligned with the nuances of Sri Lankan law.

A concept map illustrates the interconnected elements of the project, with the central node being the "LLM Project with Dataset Creation and Chatbot Integration." Domains, technologies, limitations, evaluations, and existing work orbit around this central theme, providing a visual representation of the project's comprehensive scope.

In conclusion, the chapter synthesizes insights from existing literature, articulating a methodical approach to address the project's unique challenges. The blend of legal expertise, machine learning, and technological innovations positions the Chatbot Lawyer project as a pioneering venture in the intersection of law and artificial intelligence, poised to navigate the complexities of Sri Lankan legal discourse.