NyaAI - An AI Powered Assistance for Legal Queries About Women & Children in Bangladesh

Final Year Design Project

By
Ummay Mahjabeen
213-15-4579

Samurtha Jahan Ayshe
213-15-4407

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements forthe **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Md. Jahidul Alam

Lecturer
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Mr. Md. Sadekur Rahman

Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 16, 2025

APPROVAL

This Project titled **"NyaAI - An AI Powered Assistance for Legal Queries About Women & Children in Bangladesh"**, submitted by Ummay Mahjabeen, ID No: **213-15-4579** and Samurtha Jahan Ayshe, ID No: **213-15-4407** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **16 September, 2025**.

BOARD OF EXAMINERS

Dr. Bimal Chandra Das (BCD) — **Chairman**
Professor and Dean (in-charge)
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Most. Hasna Hena (HH) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Mayen Uddin Mojumdar (MUM) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Nazibur Rahman — **External Examiner**
Technical Lead, Database Administrator
Telenor – Grameen Phone Account

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md Jahidul Alam, Lecturer,** Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Jahid 15.09.25

Md. Jahidul Alam
Designation
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Mr. Md. Sadekur Rahman
Designation
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Mahjabeen 15.09.25

Ummay Mahjabeen
213-15-4579
Department of Computer Science and Engineering
Daffodil International University

Ayshe 15.09.2025

Samurtha Jahan Ayshe
213-15-4407
Department of Computer Science and Engineering
Daffodil University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has as- sisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Md. Jahidul Alam**, **Lecturer**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Artificial Intellegence and Natural Language Processing to** carry out this project. His endless pa- tience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Com- puter Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Access to legal knowledge in Bangladesh remains highly unequal, with women and children being among the most vulnerable groups who struggle to obtain affordable, understandable, and timely legal support. Complex legal language, financial barriers, and limited professional guidance further widen this gap. This thesis proposes NyaAI, an AI-powered legal assistance system designed to provide reliable and accessible legal information for women and children in Bangladesh. The system was implemented using both traditional retrieval methods (TF-IDF, BM25) and semantic transformer-based approaches (MiniLM embeddings) on a dataset of 3,131 legal queries and documents. After preprocessing and splitting into training, validation, and testing sets, the models were evaluated using Precision@5, Recall@5, F1@5, and Success Rate. Results showed that BM25 achieved the best overall performance with Precision@5 = 0.8133, Recall@5 = 0.9667, F1@5 = 0.8834, and Success Rate = 96.7%. TF-IDF followed closely with an F1@5 of 0.8674 and the same success rate of 96.7%. In contrast, the neural models performed significantly worse, with F1@5 scores of 0.6302 (MiniLM-L6) and 0.5853 (MiniLM-L12), and success rates dropping to 83.3% and 76.7% respectively. The research contributes both academically and practically. It demonstrates that traditional IR models can outperform transformer-based methods in domain-specific, resource-constrained environments. Furthermore, the selected BM25 model was integrated into a web-based application, enabling users to submit queries and receive relevant legal information quickly and efficiently. These findings highlight the feasibility of building a practical, AI-driven legal assistant that supports marginalized groups in understanding and exercising their legal rights in Bangladesh.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, the author presents the research background that describes the problem of women and children accessing legal information in Bangladesh. It talks about the reason why an AI-based assistance system needs to be built, the purposes of the research, and the area of the work. The chapter also presents the key research questions and the contributions of the study that are expected. It establishes the background thus leading the reader to why such a system is required and how this thesis will fill the gap.

1.1 Introduction

The law sphere has a leading position in the provision of justice, safeguarding of rights and social order. But in such countries as Bangladesh, the access to legal resources is extremely unbalanced, especially to the most vulnerable population groups such as women and children. Some of the obstacles that many people encounter are lack of awareness, having to pay a lot of money in court, inaccessibility of professional help and legal language. This gap has been previously addressed by developing mobile apps that allow reporting on abuse and AIs to predict violence [4] and [3], respectively, but there is still an absence of legal support that is accessible to all marginalized groups. Consequently, the necessity to find solutions that would allow democratizing the knowledge of the law and offer culturally sensitive advice to the neediest people arises [1], [5].

There has been recent development in Artificial Intelligence (AI) and Natural Language Processing (NLP) which have created new opportunities for solving these problems. TF-IDF and BM25 have also found common law application in information retrieval via various applications to order and locate statutes and case documents [18], [25]. More advanced systems like the transformer-based-modelling and retrieval-augmented-generation [6] (RAG) systems have been seen to offer promising opportunities in answering lawful or law-related queries, locating precedents and document categorising. [6], [11], [23]. Examples of AI systems include Legal Assist AI [6] and SM-BERT-CR [13] which are more correct and context-aware than other AI systems, and culturally-specific systems such as Socheton [5] and JusticeNetBD [1] among others. This type of technology can provide resources with the strength to produce smart legal assistants that can help many people.

One possible integration is the marriage of Bangladeshi women and children legal needs and these AI technologies. This research study attempts to construct a helpful and contextual legal assistant by merging both data questioning, data preprocessing and model design, both using non-contemporary strategies of TF-IDF and BM25 and non-contemporary strategies of MiniLM embeddings. The created system NyaAI will be based on the system of methodological cycle of the collection and analysis of

legal information, models of testing, and using them to a web based system to be offered help in the real time. In this way, it will not only implement the power of modern artificial intelligence methods, and at the same time will also solve the social problem of legal knowledge being more empowering, accessible and inclusive to marginalized groups in Bangladesh.

1.2 Motivation

Women and children find it quite difficult to receive any legal help in Bangladesh. Some cannot afford lawyers, and some cannot grasp complicated legal terms and processes. Online tools, such as child protection systems through the Safe-Guard [4] and the violence prediction systems through machine learning [3], have been helpful, but they are not designed to provide direct user instructions on the legal aspect of the issue. Similarly, the efforts of research organizations, including JusticeNetBD, which purport to provide women with rights support by implementing RAG and Llama 3 [1], indicate that the work of AI may potentially provide certain legal support. The majority of such efforts are, however, either specific to their domains of interest or small, and there is still much room to fill the wide gap in the provision of a more comprehensive platform that could help both women and children in Bangladesh in handling more general legal questions.

Conversely, progress in AI and NLP has demonstrated good performance in legal text retrieval, classification and question answering in various countries [6], [11], [13]. And indeed, the case law retrieval systems that have proved the most accurate and efficient on complex legal problems include Legal Assist AI in India [6] and SM-BERT-CR [13]. However, the vast majority of these works continue to treat resource-rich legal systems and fail to focus on the special needs of marginalized users in low-resource settings. This inspires the creation of NyaAI, which combines both the old techniques (TF-IDF, BM25) and the semantic embeddings (MiniLM) to make legal help reliable, accessible, and culturally appropriate. Not only does it focus on expanding access to the knowledge of the law, but it also seeks to ensure that women and children who are left behind most often are able to move to justice with self-confidence and understanding.

1.3 Objectives

The primary goal of this thesis is to create and develop an AI-based assistant that can give women and children in Bangladesh access to and reliable legal advice. Law knowledge is often complex and ununderstood, especially in the low resource context, and hence, the current research will assist in bridging the gap in knowledge between the law and the day-to-day needs of the discriminated populations. The goals to be attained and the means to ensure that the system is not only technically correct but also culturally and user-friendly are determined by both social needs and technical possibilities. The particular aims of this study are the following:

- To retrieve and process pertinent legal information - Find laws, policies, and other legal

materials concerning women and children rights in Bangladesh and filter them by cleaning the text, normalizing and filtering the content by quality to allow the use of the data.

- To create a hybrid legal information retrieval system - Integrate the classical methodology (TF-IDF and BM25) with a semantic system (multilingual MiniLM) to enhance the quality and context of user queries.
- To study and compare various models - Experiment with various retrieval and embedding methods, evaluate their performance by statistical analysis, and choose the best model to be used.
- To create an easy-to-use web-based application - Integrate the selected AI models into a web application with JavaScript so that users (particularly women and children) can easily communicate with the system and obtain legal advice in a way that is easy to understand.
- To achieve inclusivity and cultural relevance - Develop the system so it will meet the specific needs of Bangladeshi women and children and enable them to know more about the law, make it more affordable and accessible.

1.4 Methodology

This approach of conducting research is premised on a structured and planned procedure that is supposed to foster both technical and practical respect. Data collection or the collection of legal information with reference to women and children in Bangladesh i.e. law, policy, judicial literature is the first one. This raw-material is then subjected to a preprocessing pipeline that comprises of text cleaning, text normalization, language-detection and quality-filtering. Having done this preparation, the data can be broken down into training, validating, and testing parts and freely and objective results of developing and testing the models can be obtained.

The model creation and the experimental design is an integration of both the traditional and the semantic approach. Document ranking and query matching are based on the traditional information retrieval systems such as TF-IDF and BM25. In parallel to this highly generalized types of semantics which do include multilingual MiniLM and lightweight MiniLM embeddings are also used which are aimed to determine the contextual meaning and improve performance over legal question answering. The models would be tested with the help of special assessment framework containing performance measure, the model comparison and the results statistical analysis in order to examine which approach is the most efficient.

The next step is to complete the preparation of the deployment because convenience, scalability, and accessibility are the variables that require optimization as the models are chosen to be used in the real-life situation. Finally, but not the least, the outputs are integrated into a proposed web application that is developed in JavaScript, where the user can interact with the AI assistant in the clean and clean and user friendly environment. Research direction will also contribute towards ensuring that the project will not simply be technical and progressive, but the key purpose that is, to offer women and children

in Bangladesh a stable and inclusive AI-based legal assistant, will be met.

1.5 Project Outcome

Hopefully, the final output of this thesis will be academic and social in the sense reflecting a phased approach in the form of a systematic intervention in improvement of access to legal services in Bangladesh. The project is significant because it will be addressing the most vulnerable classes of women and children who in the majority of cases are the ones who are poorly represented by the legal process. The proposed system will allow justice to be cheaper and more conveniently available to break the doors of legal complexity, lack of awareness, and inaccessibility to professionals. Along with the practical component, the project also underlines how Artificial Intelligence (AI) may be adapted to a low-resource environment to create a foundation of research and development within such an environment that, in the future, could be adopted again in the headquarters.

Concerning the methodology, the project demonstrates the synthesis of both the traditional and the semantic method of retrieval of the legal information. To see that the system can offer a good base Daniel performance, TF-IDF and BM25 are called upon, and semantic embeddings provided via MiniLM are called upon to learn more about queries. This methodological tradeoff can in itself be considered very successful in its classification and to an ironic extent, it shows that even lightweight models can yield as precise predictions without any intense processes of computation. It also provides a rational roadmap of the future legal artificial intelligence projects in developing regions, where funding is likely to be limited.

In practice, the project will offer a web-based application, in such a way that the users of this application will be able to ask legal questions afterwards receiving corresponding and situation-specific answers. Through this medium, individuals (particularly women and children) will be made aware of their rights, the relevant laws to them, and the existing legal redress. The system is user friendly which implies that it is not difficult to be used by non technical individuals. This acts as a resource in aiding the disconnection of law knowledge with normal citizens, and serves as the source that would allow individuals to proceed through to actually know how to make informed decisions in a situation where the legal help may not be as readily available.

The collective impact of this work of writing is that it might alter how the stigmatized communities approach the legal system. The project assists in curbing disparities in the disposition of the justice system through availing the legal information readily reachable, traceable and culturally possible. Moreover, it gives an example of how AI-based solutions may be introduced in a responsible way in a highly sensitive field, such as law, where the accuracy and morality issues are crucial matters. Comprehensively, the result of this thesis will be a structural change in AI help along with one of the steps toward a older, open, and making legal ecosystem in Bangladesh.

1.6 Organization of the Report

This thesis is chaptered into six which dotted on a particular aspect of the research work. The layout has been created to put forth the problem, background, methodology, implementation, and outcomes logically and coherently.

Chapter 1: Introduction

This chapter explains the subject matter of the research and presents the reason behind this study. It establishes the purpose of the research, explains the methodology proposed, describes the anticipated project results, and gives a summary of the structure of the report.

Chapter 2: Background and Literature Review

This chapter includes the background information needed and the analysis of the existing literature on this issue in legal assistance. It presents both conventional and current methods of legal information retrieval, semantic search and AI-directed legal chatbots, specifically those focusing on women and children in Bangladesh. The chapter also measures the research gaps that are being addressed by this study.

Chapter 3: Research Methodology

This chapter explains the suggested methodology. It describes how the data are going to be collected, preprocessed and split. The experimental design (the use of traditional models (TF-IDF, BM25) and semantic models (MiniLM), the evaluation framework, the statistical analysis and the preparation of the deployment) are also described in the chapter.

Chapter 4: Implementation and Results

The chapter is dedicated to the application of the proposed framework and to the technical setup. It introduces the metrics of evaluation, testing, and performance comparison of the chosen models. Findings are also explained, with a focus on the most efficient methods of answering and retrieving legal queries.

Chapter 5: Project Standards and Design Challenges

In this chapter, the norms and issues encountered in system development are discussed. It addresses elements of ethical accountability, social influence, cultural awareness and AI-based legal system sustainability. It also describes the multifaceted difficulties of adapting AI to a low-resource setting such as Bangladesh and the steps undertaken to deal with them.

Chapter 6: Conclusion and Future Work

The concluding chapter is a summary of all the research, where the most important findings and results are denoted. It also introduces the limitations that were faced in the course of the study and gives an insight on the way forward and how the system could be expanded to cover larger legal areas, enhance the system by adding multilingual support and also to increase the size of the proposed web application.

Chapter 2

Background

The chapter is a review of previous literature and more recent methods in the information retrieval, natural language processing, and artificial intelligence-run legal systems. It compares not only classical models, i.e. TF-IDF and BM25, but also modern neural models, their merits as well as their flaws. The review also discusses related works in the area of legal technology and how they can be applied to the situation in Bangladesh. Based on this review, the chapter will come up with what research is missing and situate this thesis within a broader scholarly and practice context.

2.1 Introduction

Artificial Intelligence (AI) has been on a perfect growth pace and has touched many domains of professionals, the law industry being no exception. The use and access of legal information is evolutionarily transforming, whether by retrieving documents and classifying them by cases, or by providing questions and responses to cases as well as decision support, through AI technologies. These are mostly important in the societies where facilities of justice are generally limited by lack of resource, knowledge or professional advice. In this case, AI-based tools could be highly useful in the solution of this issue connected with the availability of the legal knowledge to ordinary people and decrease the gap between legal institutions and citizens.

Legal assistance is particularly important and necessary in Bangladesh, in terms of vulnerable communities: women and children. These organizations tend to seek legal advice too late and never get to know of their rights because of social, cultural and economic restrictions. This means that despite the laws which safeguard women and children in the country, there are lots of people who cannot easily exercise their right due to the complication of the legal texts and the impossibility to receive the help of an advocate at a decent price. Technology thus offers us a new leading edge in dealing with these issues through an inexpensive, hassle free and dependable legal service.

The recent advance of natural language processing (NLP) and machine learning (ML) offered a chance of developing such systems. Semantic models enable efficient access to relevant legal documents and allow retrieval methods to be used in efficient access to relevant legal documents. Combined, these methods enable building conversational agents and Web-based systems capable of directly helping users. The combination of retrieval and transformer-based models has already proven useful in several international studies, and some initiatives have already begun to investigate AI in relation to law in Bangladesh. Nevertheless, such attempts are small-scale, language-enhancing, and implementation-oriented.

This chapter grounds the current work by examining the history and previous research in the field of

legal assistance with AI. It begins with the overview of the general role of AI in the legal industry, then proceeds to describe the use of AI in Bangladesh and elsewhere, and concludes with an indication of the gaps that prompted this thesis. After learning about these changes, it emerges why a platform such as NyaAI, an AI-powered assistant for legal queries about women and children in Bangladesh, is much-needed and timely.

2.2 Literature Review

The field of law has encountered the rise of Artificial Intelligence (AI) in research, where its application is mainly applied in areas of legal text classification, document searching, case finding, and answering questions. The field has attracted different researchers such as computer science, law and information retrieval with these critical researchers showing the significance of the field that is of interest to both the professionals in the field of law and the general population as well. It is demonstrated by the broad spectrum of research that AI has a potential to democratize the knowledge of the law and enhance access to justice. Nevertheless, difficulties are still encountered in scaling these methods to resource-limited contexts and also meeting the legal demands of the societal discriminated groups, including women and children in Bangladesh.

Hasan [1] has proposed a Retrieval-Augmented Generation (RAG)-based JusticeNetBD, markedly designed as a legal help catalyst especially to target women in Bangladesh. The system showed high accuracy at retrieval (Recall at k = 0.90) and a high response quality (BERTScore F1 = 0.896) under the umbrella of the Llama 3 (8B) model and utilizing national laws. The added value is that it offers culturally sensitive and traceable law assistance, even though the special focus on the rights of women exists at present, not touching upon other aspects of the law. Wasi et al. [2] studied how artificial intelligence could be used to offer legal services to Bangladesh and fine-tuned GPT-2 on UKIL-DB-EN, an English corpus of Bangladeshi law. The resulting model, GPT2-UKIL-EN, showed promising results in semantic assessments. Their contribution marks the first structured attempt at a legal AI for Bangladesh, though the model requires further refinement to ensure credibility and accuracy, and it remains constrained by its English-only focus.

Reza et al. [3] developed a machine learning–based support system aimed at mitigating violence against women and children. Using a web-based platform and crime forecasting, they tested several ML algorithms, with XGBoost achieving superior performance ($R^2 = 0.99$). The system's strength lies in predictive analytics for violence prevention, but it remains largely data-dependent and limited to forecasting rather than direct legal assistance. Chowdhury and Noor [4] designed Safe-Guard, a smartphone application for child abuse detection and legal aid. The system integrates multiple features such as abuse reporting, lawyer selection, and police assistance. Its novelty lies in a holistic approach to immediate child assistance, with over 58% of users giving positive evaluations. However, the study did not address deeper integration with judicial processes or long-term support mechanisms.

Sultana et al. [5] introduced Socheton, an AI-mediated tool for reproductive well-being in Bangladesh, designed through ethnographic research and guided by distributive justice principles. The system combats misinformation using culturally appropriate language and professional moderation. The key contribution is cultural alignment in sensitive healthcare AI, though the gap remains in its scalability to broader legal and social domains. Gupta et al. [6] presented Legal Assist AI, a transformer-based model fine-tuned on Indian legal datasets. By specializing in legal question answering, it outperformed GPT-3.5 Turbo and Mistral 7B, achieving 60.08% on the AIBE benchmark. Its strength lies in domain-specific accuracy and reduced hallucination, yet the model is restricted to Indian law and does not address marginalized communities specifically.

Bhatnagar and Huchhanavar [7] advanced legal AI by enhancing intent classification and domain-specific model tuning. Using BERTopic, GPT-3.5, and fine-tuned Mistral 7B, they developed a domain-specific language model for legal conversational agents. Their pipeline contributed new datasets and improved human alignment of chatbot responses, though the reliance on large-scale resources raises challenges for low-resource legal systems. Ujwal et al. [8] tackled the problem of legal long-form question answering (LFQA) by introducing the Legal-LFQA (L2FQA) dataset and incorporating legal reasoning into response generation. Their results showed that interpretability and reasoning improved LFQA performance. While this represents a significant contribution toward explainability in legal AI, the dataset's semi-synthetic nature limits its direct applicability to real-world legal cases.

In reconstructing legal question answering reliance on datasets that are annotated, Italiani et al. [9] suggested Ace-Attorney, a knowledge URL knowledge distillation framework. They obtained red model level performance at a much reduced computation cost by producing synthetic data based on frozen LLM models and training small models. The approach is useful in terms of sustainability and scalability but not explicitly tested concerning low-resource legal situations such as adolescence in Bangladesh. A BERT-based system [10] of legal question answering was introduced by Wang and Luo that was adapted to Chinese law. Pursuing Milvus a similarity search on vectors, their methodology obtained the highest recall rate of 86 percent and the method has shown to be effective in queries arising in a given domain. Nevertheless, its emphasis on the Chinese language and law does not allow its generalization across the world, especially when applied to the South Asian context.

Garlapati et al. [11] suggested an Indian law legal chatbot using Legal BERT, GPT-2, and Retrieval-Augmented Generation (RAG). They exhibited context effects and effectiveness in response to legal queries their system, trained on the wholly Indian Constitution and associated text culminated in, and promoted their authenticity relative to comparable contemporary legal systems trainable by state or parliamentary authorities. The primary contribution can be found in the integration of profound knowledge of the legal terms and human-like retrieval-enhanced responses. Nevertheless, the strategy is also confined to the Indian jurisdiction, and it is not applicable to low-resource settings. The chatbot proposed by Singh and colleagues [12] is an AI-based document search and is used to classify cases

based on TF-IDF, cosine similarity, and the Elasticsearch engine. They also have an option that is truly low in complexity i.e. fast in response time and stressed on scalability; their orientation is on accessing legal information. Although, the system is much easier to use in navigating the legal resources, it does not match the reasoning abilities of the modern LLM-based assistants.

Vuong et al. found Mildred is SM-BERT-CR, a deep learning case law retrieval system that proposes a retrieval and entailment approach that includes paragraph and paragraph matching combined with case law in ways [13]. They also suggested weakly labelled datasets to address the issue of data scarcity that they obtained them with state-of-the-art results in case law retrieval and entailment. Given the good performance, the use of the model with the associated complexities and its dependence on the long-text representation is still difficult because of the low resource jurisdictions. Visions Kumar et al. [14] have taken the next step in the integration of AI in legal activities by utilizing transformer-primed models to perform summarization, multilingual translational, and conversational heurist tasks. Their system was used in the combination of LED-base-16384, M2M-100, and Mistral-7B to simplify legal services and this reduced time and fiscal expenditure of professionals to a considerable extent. Its shortcoming is that it focuses on a professionally efficient approach as opposed to access to the vulnerable groups. Nithya et al. [15] proposed a framework for AI-driven legal automation using NLP and ML to draft documents, summarize legal texts, and answer complex queries. Their comparative analysis showed higher accuracy and operational efficiency than existing methods, with a strong emphasis on democratizing access for underserved communities. However, the study remained at the framework level without real-world deployment or evaluation in resource-constrained settings. Jacob et al. [16] introduced LegalLink, a chatbot designed to provide the Indian public with accessible legal information. Using Nomic AI encoders, RAG, and Mixtral-8x7B-Instruct, the system delivers contextually accurate responses. While it offers a modern architecture for public access, its evaluation lacks empirical evidence of effectiveness among marginalized users.

Judijanto and Vandika [17] performed a bibliometric analysis of multilingual NLP research trends between 2013–2023, highlighting the growth of transformer-based models such as mBERT and XLM-R. They identified challenges such as unequal language representation and bias in model training. Their findings emphasize the importance of inclusive multilingual AI, providing strategic insights for applying NLP to underrepresented languages like Bangla. Kayalvizhi et al. [18] applied word embeddings (Word2Vec, GloVe, TF-IDF) and similarity measures (Jaccard, cosine) for case and statute retrieval using the AILA@FIRE-2019 dataset. Their contribution demonstrates effective ranking of legal documents through vectorization, though the approach lacks the contextual reasoning provided by transformer-based models. Basu et al. [19] proposed automating legal petition generation using TF-IDF, NLTK, and ML classification. By incorporating ChatGPT to enrich training data, their system streamlined legal documentation and improved efficiency. The novelty lies in petition categorization automation, though it remains largely experimental.

Similarly, Panda and Mohapatra [20] applied TF-IDF and ML algorithms to classify court cases into

petition types. Their model demonstrated accuracy in automating repetitive legal documentation tasks, though the paper lacked detailed empirical evaluation and broader contextual applications. Bhattacharya et al. [21] organized the FIRE 2019 AILA track, which benchmarked legal AI approaches in precedent retrieval and statute retrieval using Indian Supreme Court cases. This initiative provided a shared dataset and evaluation framework, establishing a foundation for future legal IR research. While valuable, the focus on structured tasks limits its scope for conversational legal assistance.

Aydemir et al. [22] contributed to COLIEE competitions by applying BERT and TF-IDF for entailment prediction in law-based queries. Although their models did not achieve state-of-the-art results, the work showed pathways for improvement and contributed to advancing entailment tasks in legal AI. Arora et al. [23] explored AI as a legal research assistant within FIRE 2020, employing BM25, topic embeddings, and Law2Vec for precedent and statute retrieval, alongside BERT for rhetorical role segmentation of legal documents. Their system ranked among the top performers, demonstrating the potential of embeddings and segmentation in enhancing legal research efficiency.

Kim et al. [24] reported on the University of Alberta's performance in COLIEE 2021, applying BM25, transformers, and semantic knowledge integration for retrieval and entailment tasks. Their approaches achieved top-five rankings across multiple subtasks, showing the effectiveness of hybrid IR and NLI methods. However, high computational costs remain a challenge. KS and Aravindan [25] employed BM25 for statute retrieval in the AILA@FIRE-2020 dataset, demonstrating improved MAP scores (0.2975). Their contribution reinforces the importance of ranking-based approaches in legal information retrieval, though limited by the reliance on traditional IR rather than contextual AI.

In summary, the reviewed studies highlight a clear progression in legal AI, evolving from traditional machine learning approaches using TF-IDF and embeddings [18, 19, 20, 25] to advanced transformer-based and retrieval-augmented generation (RAG) architectures [11, 13, 14, 16]. Systems such as JusticeNetBD [1] and GPT2-UKIL-EN [2] address specific Bangladeshi legal needs, while others like Legal Assist AI [6] and Ace-Attorney [9] focus on scalability, accuracy, and sustainability in broader contexts. Benchmarks such as FIRE and COLIEE [21, 22, 23, 24] have helped shape evaluation standards, and bibliometric analyses [17] have highlighted gaps in multilingual inclusivity. Nevertheless, persistent limitations remain, including insufficient support for low-resource languages, limited attention to gender- and child-sensitive legal applications, and minimal integration into real-world legal frameworks. These gaps underscore the necessity for solutions like NyaAI, a culturally relevant AI assistant specifically designed to provide accessible, inclusive, and context-aware legal guidance for women and children in Bangladesh. The below table 2.1 shows the summary of the related studies.

Table 2.1: Summary of Literature Reviewed.

Author(s)	Model / Method Used	Results	Contributions
Hasan [1]	RAG with Llama 3 (8B)	Recall@k = 0.90, BERTScore F1 = 0.896	Built JusticeNetBD for women's rights; culturally sensitive and traceable legal AI.
Wasi et al. [2]	GPT-2 fine-tuned on UKIL-DB-EN	Promising semantic performance	First structured attempt at Bangladeshi legal AI; limited to English corpus.
Reza et al. [3]	ML models (XGBoost, others)	XGBoost $R^2 = 0.99$	Forecasted crimes against women & children; strong predictive tool but no direct legal aid.
Chowdhury & Noor [4]	Mobile app with ML features	58% users gave positive feedback	Safe-Guard app for child abuse detection and legal aid; practical but limited judicial integration.
Sultana et al. [5]	AI with ethnographic design	Effective in combating misinformation	Socheton supports reproductive well-being; culturally aligned but not scaled to wider legal issues.
Gupta et al. [6]	Transformer-based QA	60.08% on AIBE benchmark	Legal Assist AI, domain-specific and accurate; focused on Indian law.
Bhatnagar & Huchhanavar [7]	BERTopic, GPT-3.5, Mistral 7B	Improved chatbot alignment	Enhanced legal intent classification; requires large resources.
Ujwal et al. [8]	Legal-LFQA dataset + reasoning models	Better interpretability in LFQA	Focused on explainable long-form QA; dataset semi-synthetic.
Italiani et al. [9]	Knowledge distillation from LLMs	Smaller models with LLM-level output	Reduced costs, scalable QA; not tested in low-resource countries.
Wang & Luo [10]	BERT + Milvus vector search	Recall = 86%	Effective Chinese legal QA system; limited to Chinese law.
Garlapati et al. [11]	Legal BERT + GPT-2 + RAG	High contextual accuracy	Indian legal chatbot; retrieval + generation hybrid.
Singh et al. [12]	TF-IDF + cosine similarity + Elasticsearch	Fast and scalable retrieval	Simple and efficient legal assistant; lacks deep reasoning.
Vuong et al. [13]	SM-BERT-CR	State-of-the-art retrieval	Strong retrieval and entailment; complex and data-heavy.
Kumar et al. [14]	LED-base, M2M-100, Mistral-7B	Reduced time & cost in workflows	Advanced AI for summarization & translation; focused on professionals.
Nithya et al. [15]	NLP + ML automation framework	Higher accuracy than older methods	Proposed automation for underserved groups; no real deployment yet.

2.2.1 Similar Applications

Some have tried to develop applications that would offer legal help or some other kind of help that are still related to the approach of this thesis. To illustrate this, Hasan [1] has proposed JusticeNetBD, a RAG-based system, Retrieval-Augmented Generation, that generates responses to legal questions asked by women in Bangladesh with the help of a large language model. The paper shows that retrieval and generation can be integrated to achieve frequency and cultural sensitivity, which is associated with the scope of the given thesis that refers to the quality of retrieval and reliability of control of the response. In the same spirit, Wasi et al. [2] have constructed GPT2-UKIL-EN, an initial prototype that is trained on Bangladeshi legal texts. It is written in English only, but emphasizes the role of preprocessing and language adaptation in data preparation, which also belongs to the workflow of this thesis.

There are other applications, which are more general in their scope, but which have similarities with this study in the area of technical design and deployment. Among the simplest examples of how digital tools could be used to provide access to legal and social assistance, once should mention Safe-Guard by Chowdhury and Noor [4], a child protection mobile application. An example of how statistical analysis can inform prevention is described by Reza et al. [3], who developed a web-based system that predicts crimes committed against women and children based on machine learning. Similarly, Gupta et al. [6] and Garlapati et al. [11] developed legal chatbots in India using transformer-based models and retrieval systems, and these studies have proved that, after the evaluation and model selection are done, such systems can be scaled to working applications.

The methodological flow is similar in these works: data is collected, preprocessed, both traditional retrieval models (e.g., TF-IDF, BM25) and semantic models (e.g., BERT variants, transformers) are used, and the results are evaluated and deployed to the user-facing system (chatbots, mobile applications, or web-based services). But not one of these applications deal directly with the integrated legal requirements of women and children in Bangladesh under one, culturally-sensitive system. This thesis, thus, expands on these examples as it customizes its approach and implementation to an inclusive, web-based assistant - NyaAI. The below table 2.2 shows the summary of the similar studies.

Table 2.2: Similar Applications table.

Author	Model	Accuracy / Result	Contribution
Hasan [1]	RAG with Llama 3 (8B)	Recall@k = 0.90, BERTScore F1 = 0.896	Built JusticeNetBD, legal AI for women in Bangladesh.
Wasi et al. [2]	GPT-2 fine-tuned on UKIL-DB-EN	Promising semantic results	First AI model for Bangladeshi legal texts (English only).
Reza et al. [3]	ML models (XGBoost best)	$R^2 = 0.99$	Web system to forecast crimes against women & children.
Chowdhury & Noor [4]	Mobile app with ML	58% positive user feedback	Safe-Guard app for child abuse detection & legal help.
Gupta et al. [6]	Transformer-based QA	60.08% on AIBE benchmark	Indian legal assistant with better domain accuracy.
Garlapati et al. [11]	Legal BERT + GPT-2 + RAG	High contextual accuracy	Indian chatbot with retrieval + generation design.

2.3 Gap Analysis

The analyzed literature shows that legal AI systems have been gradually evolving, with more modern types of systems, such as transformer-based network and Retrieval-Augmented Generation (RAG) systems, replacing the traditional information retrieval techniques, i.e., TF-IDF and BM25 [18, 25]. As much as these developments reflect technical maturity, there are still a number of glaring gaps when it comes to the particular context of Bangladesh and specifically, women and children. In Bangladesh, the first steps towards localized legal AI have already been taken, including JusticeNetBD [1] and GPT2-UKIL-EN [2]. Yet, they are quite limited, JusticeNetBD is concerned only with the rights of women, and GPT2-UKIL-EN is restricted to the use of English texts, leaving out no fewer than 50 percent of the Bangladeshi population who feel more at home in Bangla. This is the gap, which makes the multilingual and more inclusive systems, which turned out to be more reflective of the country legal and linguistic diversity, important.

The second significant weakness is that it is lacking an equal balance between retrieval-based methods and semantic comprehension. The classical techniques and those considered by Singh et al. [12] and Kayalvizhi et al. [18], are effective in retrieving but lack any form of reasoning on the context. More complex models On the other hand, more semantically accurate models of transformers, such as those in Gupta et al., [6] and Italiani et al., [9], can be adapted to resource-intensive in no time, hence will never be easily transferred to low-resource contexts such as Bangladesh. The challenge of figuring out hybrid structures that could fuse both traditional retrieval and semantic models, that are both efficient and contextually accurate, are only one methodological challenge and are also computationally expressible.

The breaches also occur in deployment-based studies. Though the system(s) of children protection and prevention of crimes elaborated by both Chowdhury and Noor [4] as well as by Reza et al. [3] are

incredibly useful during that stage of practically applicable system(s) advancement, they are not yet implemented fully into the legal system; they focus on either prevention or direct support in terms of additional legal framework. This applies equally to such articles as Safe-Guard [4] and Socheton [5] that justify the utility of tools on offer to users but only to some but not broad legal frames of social issues. This is to mean that there is a lack of co-connection between end to end solutions created after prototypes to viable web based systems that can possibly give holistic legal guidance. Last but not least, another challenge that appears repeatedly is the lack of attention to marginalized groups regarding the research on legal AI. Although the global systems like Legal Assist AI [6], SM-BERT-CR [13], and LegalLink [16] prove to be highly effective in the field of retrieval and question answering, they only address legal issues faced by the Bangladeshi women and children in regard to Indian law or international law. This thesis fills in these gaps through a proposed solution, NyaAI, which is an AI-based assistant that is a combination of retrieval and semantic approaches, tested to be accurate and usable, and implemented as a web-based platform with a specific goal to help women and children in Bangladesh. The below table 2.3 shows the Gap Analysis Summary.

Table 2.3 Related Research description table

Research Gap	**Evidence from Literature**	**How This Study Addresses the Gap**
Limited coverage of Bangladeshi legal domains	JusticeNetBD [1] focused only on women's rights; GPT2-UKIL-EN [2] restricted to English legal texts.	NyaAI integrates both women's and children's legal issues, using Bangla and English resources for inclusivity.
Imbalance between retrieval efficiency and semantic reasoning	Traditional models (TF-IDF, BM25) [12, 18, 25] provide fast retrieval but weak reasoning; transformers [6, 9, 13] improve accuracy but are resource-heavy.	NyaAI combines traditional retrieval with semantic models, ensuring contextual accuracy while remaining computationally feasible.
Lack of deployment-ready systems in Bangladesh	Safe-Guard [4] and ML-based forecasting [3] remain siloed prototypes without deeper legal integration.	NyaAI is deployed as a web-based assistant, designed for practical accessibility and real-world usability.
Minimal focus on marginalized groups	Global tools like Legal Assist AI [6], LegalLink [16], and SM-BERT-CR [13] achieve strong results but do not address Bangladeshi women and children.	NyaAI specifically targets women and children in Bangladesh, ensuring cultural relevance and inclusive access to legal guidance.
Weak evaluation frameworks for local contexts	Many studies [5, 11, 14] focus on technical accuracy but lack user-centered evaluation in low-resource settings.	NyaAI includes both technical evaluation (retrieval/semantic accuracy) and user-oriented testing for relevance and clarity.

2.4 Summary

This chapter has discussed the background that is required to comprehend the creation of an AI-based legal assistant trained to support women and children in Bangladesh. It started by outlining the more general application of Artificial Intelligence in the legal field and the ways the technologies of retrieval models, semantic understanding, and conversational agents are transforming access to justice across the globe. The developments indicate how AI can streamline law-related processes, minimize obstacles, and offer prompt assistance to users who do not always have access to professional assistance.

A literature review was then conducted of various other related works both in Bangladesh and elsewhere. The work of JusticeNetBD [1], GPT2-UKIL-EN [2], and other studies is an initial move towards localized law-related AI, and programs like Safe-Guard [4], and Socheton [5] demonstrate how AI applications can provide specific support in such sensitive fields as in the child protection profession or in reproductive health. Simultaneously, international systems as well as transformer-based model [6, 9, 13] and retrieval-augmented model [11, 16] demonstrate the technical directions towards creating correct and scalable legal assistants. These articles combined show how AI has become increasingly mature in the field of law, as well as highlight some of the critical limitations of AI in application to resource-limited settings, such as Bangladesh.

The gap analysis revealed some of the issues that have been persisting. Current systems are mostly limited in scope, based on English-only sources, or still at the prototype stage and not fully implemented. Further, global studies are highly technical, but rarely discuss marginalized communities, or fit the cultural and linguistic context of Bangladesh. The results highlight the importance of a system, which combines retrieval efficiency with semantic reasoning, is multilingual, and is implemented as an operational web application.

Based on these findings, this thesis describes NyaAI, a web-based legal assistant tailored to women and children in Bangladesh. The system will also ensure that users have access to reliable legal advice, which is context-related and not technical and which offers adequate user experience through a combination of old and new semantic models. The following chapter identifies the research methodology leading to the integration of data collection, preprocessing, model assessment, and deployment to fill the gaps identified and reach the goals of this study.

Chapter 3

Research Methodology

This chapter is an explanation of the methodological skeleton used in the thesis. It describes the data collection process, preprocessing and the division of the data into training, validation and testing sets. The next section of the chapter describes the design of the experimental setup, in which both traditional and semantic retrieval models were used. The analysis model is introduced, including the model selection and the statistical analysis. Lastly, the chapter explains how the model with the highest performance was ready to be deployed into a web-based application. A flowchart is shown to explain the step-by-step process in a better manner.

3.1 Methodology/Requirement Analysis & Design Specification

3.1.1 Overview

This paper adopts a systematic approach to designing and evaluating NyaAI, an AI-powered legal information assistance system that provides answers to legal questions for women and children in Bangladesh. The research process begins with the systematic gathering of more than 3,000 legal documents and query-document pairs, ensuring that relevant acts, sections, and categories of victims for the target domain are included. For improved usability, the data is carefully preprocessed in terms of text cleaning, normalisation, language detection, and quality filtering before being divided into training, validation, and test sets for rigorous experimentation. Both conventional retrieval techniques (TF-IDF, BM25) and semantic transformer-based models (Multilingual MiniLM, Lightweight MiniLM) are used and compared within a common evaluation framework. Performance is quantified using conservative automated metrics, including Precision@5, Recall@5, F1@5 and Success Rate, with statistical validation provided through bootstrap sampling and pairwise tests. The model ultimately chosen is the one that is both frequently interpretable and reliable and statistically significant. Models are packaged into deployable configurations, combined with client-side JavaScript, and optimised for accessibility using responsive web design. Not only is this holistic approach methodologically sound, but it also grounds the system in the context of practical usability, making it a potential tool in bridging the gap between citizens and legal information in Bangladesh.

3.1.2 Proposed Methodology/ System Design

Figure 3.1 illustrates the methodological workflow of the NyaAI system, which begins with the collection of data from legal documents and query-document pairs relevant to Bangladeshi law. The raw data is then fed through a preprocessing stage, which ensures the consistency and reliability of the

data by cleaning the text, normalizing it, detecting its language, and screening it for quality. After this, the dataset is divided into training, validation, and test datasets to facilitate robust experimentation. The paper's experimental framework consists of two families of models: classical retrieval models (TF-IDF and BM25) and semantic transformer-based models (Multilingual MiniLM and Lightweight MiniLM). These models are evaluated using a formal evaluation framework defined by evaluation measures such as Precision@5, Recall@5, F1@5, and Success Rate. To increase validity, statistical analysis methods such as bootstrap sampling and pairwise tests are employed before selecting the final model. After the optimal model is found, the process enters the deployment preparation stage, where multiple package sizes and configuration files are generated for production. Ultimately, the methodology results in the development of a web-based application utilising JavaScript, which can provide legal query assistance on a platform that is both accessible and responsive, facilitating effective delivery.

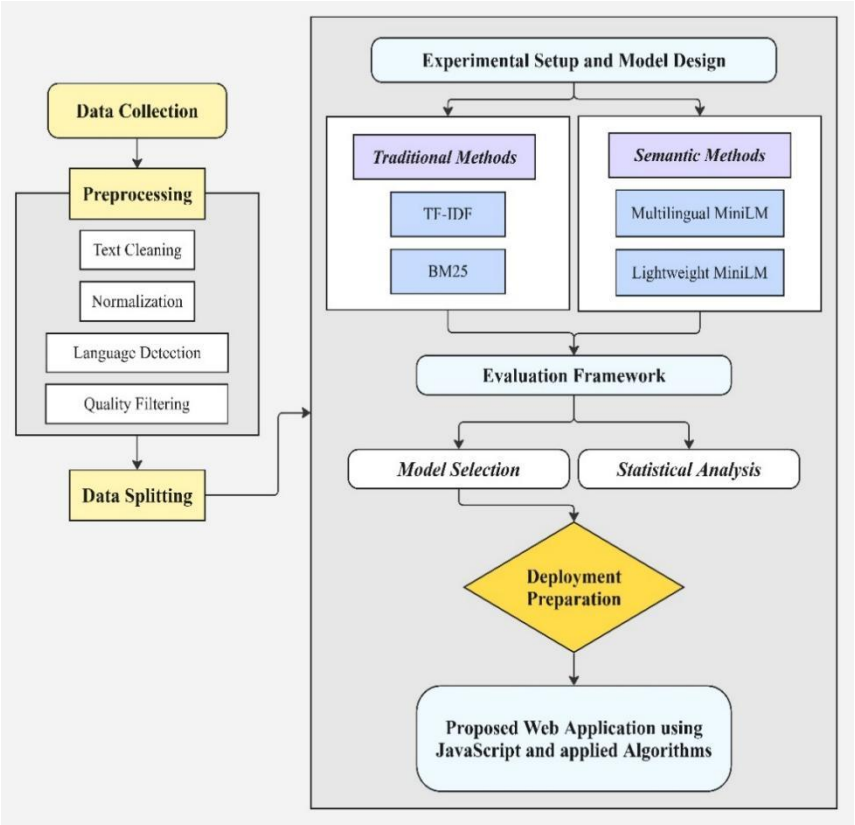

Figure 3.1: The proposed Methodological Flowchart

3.1.3 Functional and Nonfunctional Requirements

A set of functional and non-functional requirements has been developed to achieve the goal of creating an efficient and robust legal information retrieval system. Those requirements inform the technical implementation and determine the criteria of success of the system.

Functional Requirements

- Document Retrieval: The system should be able to process a query input in natural language by a user and provide a ranked list of the most relevant legal documents in the available specialized corpus of Bangladeshi cyber laws.
- Multi-Algorithm Support: The backbone architecture should support and enable the assessment of various retrieval algorithms, namely the classic lexical algorithms (TF-IDF, BM25) and the more modern semantic transformer-based models (Multilingual MiniLM, Lightweight MiniLM).
- Automated Evaluation: The system should also include a threshold-based evaluation system to automatically evaluate and compare model performance based on standard metrics, such as Precision at K, Recall at K, F1-Score at K and Success Rate.
- Statistical validation: The software should be able to do intensive statistical analysis, such as bootstrap sampling and pairwise significance testing, to make sure that the performance differences reported between any two models are sound, and not by chance.
- Deployment Preparation: The system should produce optimized and serialized versions of the most performing model in various package sizes to deploy it efficiently in a production system.
- Web Application Functionality: The end consumer application should allow a user-friendly interface to enter queries, choose a retrieval algorithm, and display the results in time-saving progressive loading to ensure that large datasets can be used to the fullest.

Nonfunctional Requirements

- Performance: The web application should be able to provide the results of queries within a reasonable amount of time, with response times of less than a second on lexical methods and some acceptable latency on more computationally intensive semantic models.
- Usability: The user interface should be usable on multiple gadgets, desktops, tablets, and mobile phones, among others.
- Maintenance: The codebase should be well-documented and structured in a modular way so that it can be easily updated with new legal corpus, new retrieval algorithms can be added, and the evaluation parameters can also be modified.
- Client-Side Efficiency: The application code should be coded to do most of its work client-side, with fewer and more efficient implementations of algorithms using JavaScript.

3.1.4 Data Flow Diagram

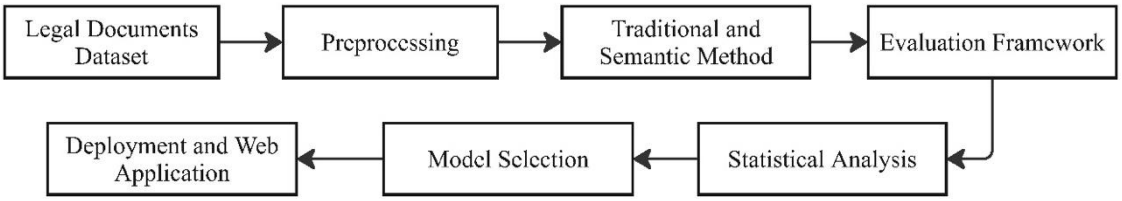

Figure 3.2: The proposed DFD Level -1

Figure 3.2 illustrates the high-level data flow diagram (DFD) that describes the primary processes of the legal document retrieval system. The workflow begins with the ingestion of the raw legal documents' dataset, followed by the preprocessing step of cleaning, normalization, and preparation. Both semantic and traditional retrieval systems can subsequently use this refined data in parallel to produce relevance ranking. The results of these models are fed into a centralized evaluation system, where performance is quantitatively assessed. The outcomes of this assessment are then subjected to a rigorous statistical examination to identify significant differences among models. This analysis can actually guide the model selection process, wherein the most performing algorithm is selected. Lastly, the chosen model is deployed and connected to the client-side web application, completing the end-to-end data flow to a working tool.

3.1.5 UI Design

NyaAI Legal Assistant user interface is simplified, available and user friendly in interaction to allow users such as the legal community as well as the general population. The interface is in a clean and minimalist design style and has a large user-friendly chat-style input field, prompting the user to use natural language queries on the subject of Bangladeshi cyber security law. Sample queries and place holder text that are right on the input box offer some contextual clues that the user can understand and some initial hitches since it shows the user what the system can record. Its design aims at a streamlined one-page application experience that can be used effectively on desktops, tablets and mobile gadgets and its accessibility is wide without complex navigation or technical expertise.

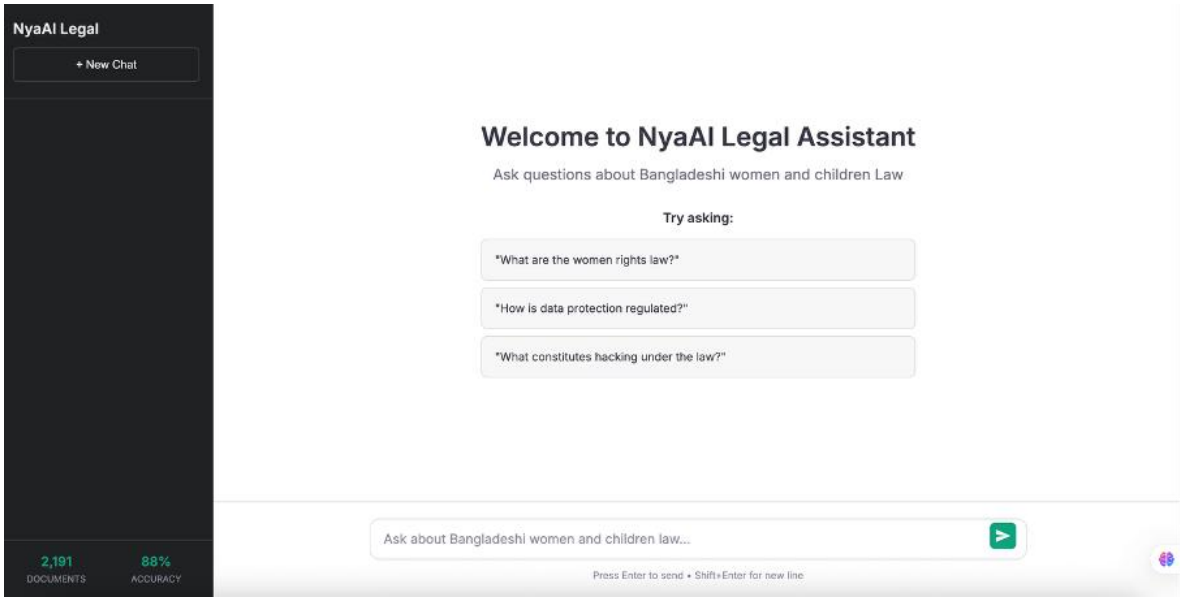

Figure 3.3: The proposed UI Design for Web Application

3.2 Detailed Methodology and Design

3.2.1 Dataset

The dataset presented in this work was very well refined to create the core of the NyaAI legal query assistance system, which includes over 3,000 legal documents and query-document pairs that were acquired by authoritative means. Md. Saiful Islam and Kazi Mehedi Al Amin used primary data to record the authenticity of the information and place this within the context of the world by accessing the official portal of laws of Bangladesh and the exclusive legal reference book titled Nari O Shishu Nirjaton Daman Ain 2000. Well dispersed across the dataset are organized attributes, such as Act number, section title, victim type, user query, suggestions and text of the law, which, when combined, represent a full relational matrix of queries related to particular jurisdictions. Even more successful applicability of the dataset to domain-specific analysis is guaranteed by the addition of metadata connected to the target victims (e.g., women, children, organizations, government bodies) to them. This principle proposed concept of the primary data collection cannot be overworked in relation to a thesis or applied research because it would not only be the first time researched, it would be directly applicable to both the legal as well as social context of Bangladesh without using secondary and generalised data. This research should be viewed as bridging the gap which arises due to absence of domain-specific research materials on the particular issue by basing the research on a primary contextually relevant data set, and such a choice allows not only to ensure that the research is valid, credible and applicable to professional scholarship and legal services work, but also to claim that the research gap is bridged.

Table 3.1: Dataset Specification

Properties	Values
Dataset Type	CSV (Structured Query–Document Pairs)
Total Records	3,000+ (Acts, Sections, Queries, Suggestions, Laws)
Independent Variables	Act No, Section, Name, Victim, Query
Dependent Variable	Suggestion, Law (Legal provision corresponding to query)
Total Features	7 Attributes (Act No, Section, Name, Victim, Query, Suggestion, Law)
Source	"Laws of Bangladesh" Portal, Nari O Shishu Nirjaton Daman Ain 2000 with Special Penal Laws
Domain	Legal Query Assistance (Bangladesh Laws on Women & Children)
Classes	Multiple (Queries mapped to corresponding Acts/Sections)

3.2.2 Classes

The dataset is divided into defined classes based on various dimensions of legal information, allowing for the structuring and analysis of queries in relation to their respective laws. The following describes each class:

Act No: This is the category in which the legal act under which the law exists is found, e.g. Act No. 39 of 2023. It is the official authority and the fact that any query/provision could be corresponding to an official statute is the point to the integrity and credibility of the dataset.

Section: Sections provide granularity to the dataset with the description of which section is offered in which act. The statutory provision is identified with specific definite statutory provisions facilitating the added precision of the process the provision is necessary in retrieving the provision and undergoing the granting of a particular legal interpretation (in the use of application to another purpose).

Name: This class contains the title or designation which is the brief labelling of the section and a ready summary of the legal information. A guiding device through which a human reader and computational models are presented an easy to digest legal text of complexities.

Victim: The victim category determines the target groups that are covered by the law (women, children, organization or a state). With this typology, a system is able to accuse the social orientation of every act and accord the legal protection to specific populations.

Query: Queries are natural linguistic queries that attempt to replicate how the user might request the legal assistance. When these questions are mapped with laws, which can provide a response to them, this system will be capable of illustrating how applicable these laws are in the real world and bridging the gap between the laws and the cognition of the public about the laws.

Suggestion: These classes include suggestions or explanatory notes that can assist in clarifying the

extent of the law or making it more useful. Recommendations add context to the corpus not only to both kinds of retrieval models, but also to how the user interprets the law.

Law: The written law or section of the acts are in the class law. This will be the gold standard by which all results extracted are juxtaposed to, to ensure that all received results are supported by law, and that the AI system is reliable.

Table 3.2: Instance of Dataset

Act No	Section	Name	Victim	Query	Suggestion	Law
Act No. 39 0f 2023	Section 1	Introduction to the Cyber Security Act, 2023	General Public, Organizations, Government	What is the Cyber Security Act, 2023?	A clear introduction should be provided to help people understand the purpose and scope of the law.	(1) This Act may be called the Cyber Security Act, 2023 . (2) It shall come into force immediately.
Act No. 39 0f 2023	Section 1	Introduction to the Cyber Security Act, 2024	General Public, Organizations, Government	When did the Cyber Security Act, 2023, come into effect?	The law should specify its effective date to ensure clarity in its implementation.	(2) It shall come into force immediately.
Act No. 39 0f 2023	Section 1	Introduction to the Cyber Security Act, 2025	General Public, Organizations, Government	What are the primary objectives of the Cyber Security Act, 2023?	The objectives should be outlined explicitly to define the law's role in cybersecurity.	The Act aims to enhance cybersecurity measures, prevent cyber crimes, and protect digital infrastructure.
Act No. 39 0f 2023	Section 1	Introduction to the Cyber Security Act, 2026	General Public, Organizations, Government	Which authorities are responsible for enforcing the Cyber Security Act, 2023?	Government and law enforcement agencies should collaborate for effective enforcement.	The Act is enforced by the Cyber Security Authority, law enforcement agencies, and other relevant bodies.

Act No. 39 0f 2023	Section 1	Introduction to the Cyber Security Act, 2027	General Public, Organizations, Government	What penalties are outlined in the Cyber Security Act, 2023, for cyber offenses?	A detailed section should be included specifying different offenses and their corresponding penalties.	The Act includes penalties such as fines, imprisonment, and restrictions for cyber-related offenses.

3.2.3 Data Preprocessing Steps

The preprocessing pipeline is applied in accordance with the methodological flowchart, to traditionally enhance the data quality, consistency and arrange the data to be processed downstream in a modeling environment. Each of the steps is outlined below, along with the figures representing the outputs of those steps that best depict how it works:

Text Cleaning: All the queries and legal texts are purely clean with lowering the case, no punctuation, extra spaces, and unnecessary characters. This ensured homogeneity within the information.

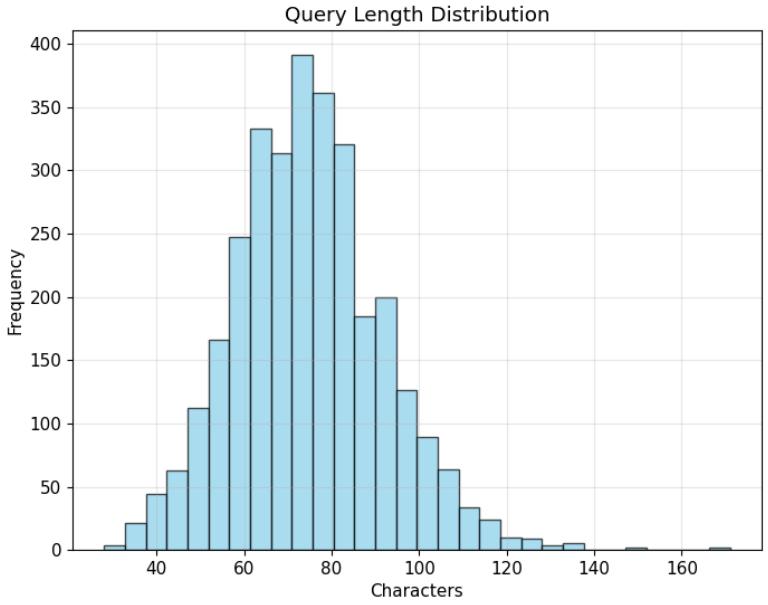

Figure 3.4: Query Length Distribution

Normalization: Legal terms and phrases were standardized using lemmatization in order to minimize the differences between word forms. This allowed semantic meaning to be preserved without undue repetition.

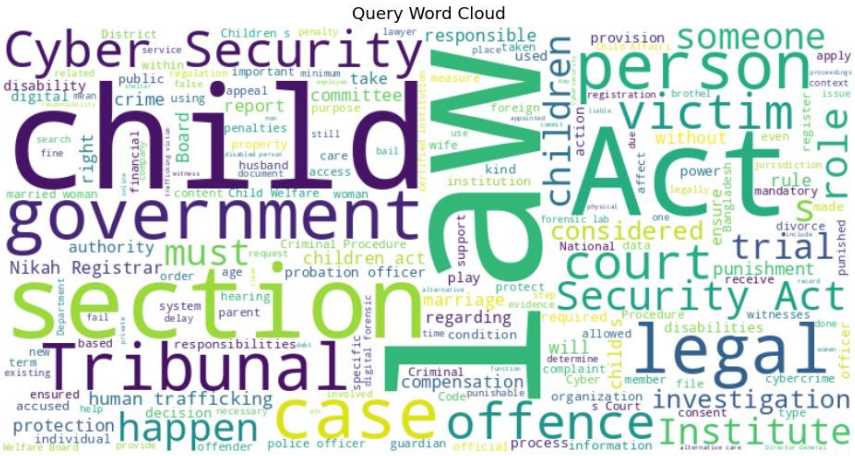

Figure 3.5: Query Word Cloud

Language Detection: To process both English and Bengali materials, a language detection method was used to filter documents and align them. This was done to make sure that only valid entries were kept in the dataset.

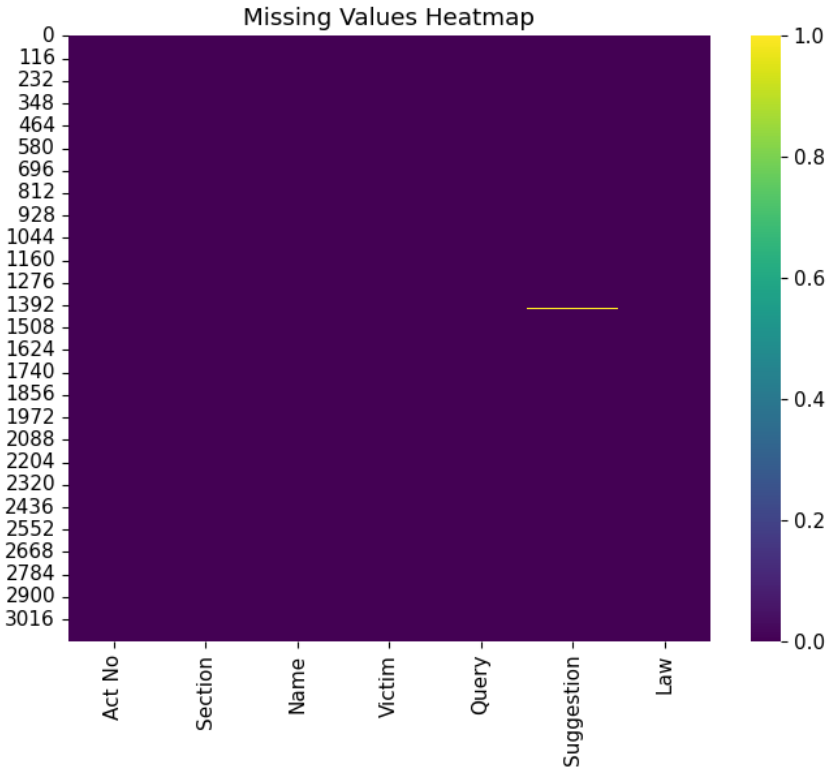

Figure 3.6: Missing Values Heatmap

Quality Filtering: Duplicates, missing records, and inconsistencies were systematically filtered out to maintain quality input data This greater data integrity enabled more powerful models to be trained.

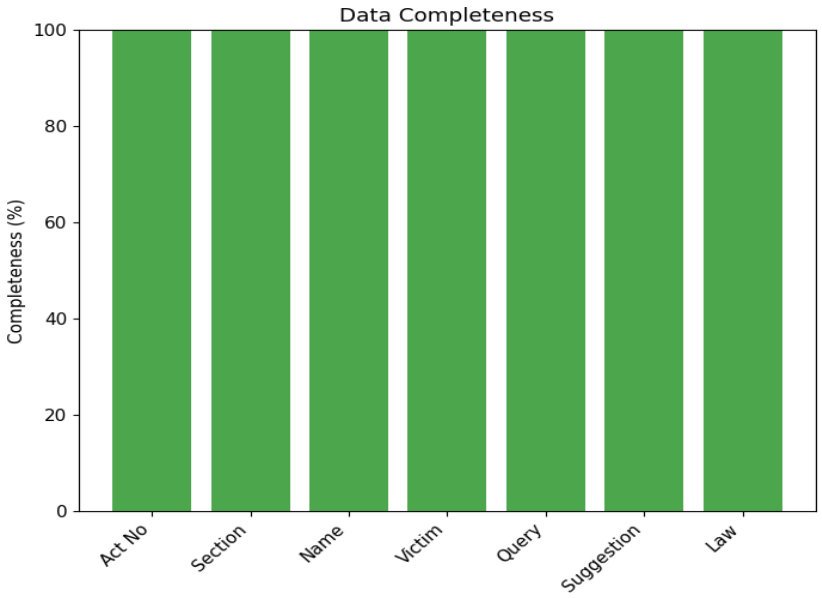

Figure 3.7: Data Completeness Chart

3.2.4 Model Description

The represented framework reviews both the classical and semantic variation of retrieval of most relevant laws to hand in the query. This part will examine the conventional strategies which are the point of evaluation of performance. The methods rely on statistical and lexical matching, and not so on context embeddings. They are significant instruments of establishing baselines in the legal IR as they are simply comprehended and interpreted, and numerically viable to understand. The traditional approaches we have applied to this study are TF-IDF and BM25 that are defined as follows:

1. Traditional Methods

Term Frequency-Inverse Document Frequency (TF-IDF): It is a technique of determining the importance of a word within a document relative to the entire corpus. Term Frequency (TF) of a word is the number of times it appears in a document and Inverse Document Frequency (IDF) is a means of reducing the degree of common words which appear in a lot of documents.

$$TF\text{-}IDF(t,d) = TF \times \log \frac{N}{DF(N)} \qquad \text{.....}(i)$$

where t is a term, d is a document, N is the total number of documents, and $DF(t)$ is the number of documents containing t. In practice, the cosine similarity between the vectors for the query and the documents is computed so that the system can sort documents by lexical overlap. This approach is appropriate if the search terms are close in meaning to the language in which the legal text is written. BM25 (Best Matching 25): BM25 is an improved ranking function that extends TF-IDF with normalization for document length and saturation effects of word frequency. It doesn't put too much

emphasis on the same words repeatedly in the same document.

$$BM25(d,q)=\sum_{t\in q} IDF(t)\cdot \frac{f(t,d)\cdot(k_i+1)}{f(t,d)+k_1\cdot(1-b+b\cdot \frac{|d|}{avgdl})} \quad \text{.....}(ii)$$

where $f(t,d)$ is the frequency of term t in document d, $|d|$ is the length of document d, avgdl is the average document length, and $k1$, b are hyperparameters controlling term frequency saturation and normalization. BM25 finds a middle ground between relevance and bias, penalizing documents that are too long and favoring documents that contain more of the same terms as the query.

2. Semantic Method

The other than traditional lexical matching techniques heralded in this research are the semantic techniques that have a capacity of reflecting the underlying contextual sense of user query and law texts. These methods are based on transformer-based models that give dense representations of vectors (embeddings), and the retrieval is possible even in cases when one may have little or no intersection between keywords. This study semantically applies the Multilingual MiniLM and Lightweight MiniLM.

Multilingual MiniLM: Multilingual MiniLM Multilingual MiniLM is a mini transformer, which is trained on fits of cross-lingual questions and literature. It trains ready-made sentence transformers which encode text into fixed-size dense vectors which are semantically equivalent across languages. Large aims of inquiry given: osocng, law document. d, both are mapped into an embedding space using a neural encoder $f(\cdot)$:

$$\vec{q}=f(q),\ \vec{d}=f(d) \quad \text{.....}(iii)$$

The similarity in the queries and the document is then determined using the cosine similarity and the greater the number the more related seems to be the two things. This enables the system to analyze a user query in the natural language as against relevant laws despite the query being written in simplified and paraphrased language as compared to language available in the laws.

Lightweight MiniLM: this model is developed to have a higher inference speed yet a reasonable degree of context comprehension. This does not require lots of transformer-layers and parameters and therefore it is more likely to be implemented on resource limited platforms like web browser or mobile applications. It uses smaller dimensions, though maintaining high recall accuracy because it is concerned with significant semantic attributes. The retrieval can be described as a somewhat similar Multilingual MiniLM (MML) but more focused on computational efficiency.

$$Score\ (q,d) = \frac{\vec{q} \cdot \vec{d}}{\|\vec{q}\| \|\vec{d}\|} \quad \text{.....}(iv)$$

3.3 Project Plan

To allow proper orderly progress the creation of the NyaAI system, as well as achieve the deadlines, the project plan is broken into stages. The working plan is being formulated now, as soon as January 2025 and the theoretical study in February to rest the conceptual base. The entire literature review happened in February through March and data collection and preprocessing went through in March to April to come up with a good quality data. During the next several months between May and June, technical and academic rigor occupied the time by model design and methodology writing. The process of application development and testing was to be done in June and July during which the retrieval models were integrated into an operational web-based system. Lastly, the project will end in August 2025 with a report, summarization of findings, methods and results in a thesis. The timeline also has a sound balance of duties where technical development and academic writing will take place in parallel process.

Table 3.3: The GANTT Chart of Project Timeline

Process	Jan'25	Feb'25	Mar'25	Apr'25	May'25	June'25	July'25	Aug'25
Working Plan								
Theoretical Study								
Literature Review								
Data Collection								
Data Preprocessing								
Model Design + Methodology Writing								
Web Development and Testing								
Report Writing								

3.4 Task Allocation

To ensure easy implementation and equal contribution of every stage of the project, distribution of duties among team members was done in a non-random manner. The data collection and preparations (gathering legal documents, cleaning and structuring the data) was carried out by one member. One

of the participating members was involved in model development and evaluation, that is, the implementation of both classical and semantic methods of retrieval, experimentations and result interpretation. Another member member was to do application developing and testing and putting the retrieval models in a web based interface and usability. Besides that, literature review, methodology writing and report preparation were other areas that received a process of co-construction and that were periodically cross-checked to stay aligned and delivering satisfactory quality. Supervisors gave directions to make sure that the working individual activities were aligned with the project objectives, and showed technical and scholarly merit.

3.5 Summary

To ensure the comfortable process of implementation and equal contribution of the work of all project periods, the distribution of tasks among the team members was performed in an orderly manner. One member worked on data collection and preprocessing, including legal documents collection, data query cleaning, and data structuring. Designing of models and assessing them, where both classical and semantic techniques of retrieval were used, tests and analysis of results were done by one of the participating members. The application development and the testing and implementation of the retrieval models into a web-based interface and usability was the duty of a third member. Moreover, literature review, methodology writing and report preparation were also generated under a process of co-construction such that cross checks were conducted periodically to ensure consistency and quality. Supervisors offered leadership to make sure that every activity of theirs aligned with the project goals, and were technically and scholarly rigorous.

Chapter 4

Implementation and Results

The content of this chapter is the specifications and experimentation outcomes of the research. It begins with the environment establishment, such that, all the models used the same parameters and dataset splits. This chapter goes on further to define the measures of evaluation and gives the results in table and plot form providing a comparison of the performance of TF-IDF, BM25, and transformer based models. The findings are confirmed through statistical tests, error checking and reliability test. The discussion justifies the decision to use BM25 as the most advisable and why TF-IDF was as well-competitive. The chapter has a conclusion that provides a summary of the key findings and how they apply in real life.

4.1 Environment Setup

In order to offer equal and similar assessment of all the models, controlled condition was done at the implementation stage. Interpretation pipeline and experiment setup were also parts of such an environment and could be compared to each other devoid of an external bias. The models were all trained and identical datasets preparation processes carried out in all cases. Such uniformity also maximized the validity of the results on top of reducing variability.

Similar training parametric set was simulated common to all models in experiments. This involved the size of vocabulary controlled, the largest maximum query amount controlled as well as optimization of batch size/ learning rate. The same was also maintained with Evaluation measures PK, Success rate, F1 PK, Precision PK and Recall PK. Such parameters were selected as they are the most commonly used parameters in the studies of information retrieval and also due to the queries performed to the legal domain in this study. The complete set of the shared parameters is given in table 4.1.1.shared parameters.

Table 4.1: Common training parameters

Parameter Name	Parameter Value
Vocabulary Size	3000–3654
Maximum Query Length	171 tokens
Batch Size	32
Learning Rate	0.001
Evaluation Metrics	Precision@5, Recall@5, F1@5, Success Rate

The data in this thesis was first processed to eliminate errors, and normalize text before being split into training, validation, and testing data. The division was done in 70:15:15 proportions which is quite common and gives a reasonable balance between learning and evaluation reliability. The models

were fitted using the training set, fine-tuning and modification of the parameters was performed using the validation set and the performance was evaluated independently using the test set. This set of data is summarised in Table 4.1.2.

Table 4.2: Dataset split summary

Dataset	In Percentage	Number of Queries
Training	70%	2191
Validation	15%	470
Testing	15%	470

With the same training parameters and the same division of the data, the performance of the TF-IDF, BM25, and transformer-based models could be placed into the same level. Such a systematic arrangement was necessary so that variation in outcome was due primarily to the capabilities of the models themselves as opposed to non-uniform training conditions. Not only does this type of design make the research process rigorous, but it also provides a good ground to interpret the findings with a lot of confidence.

Regarding computing resources, the experiments were done in Python 3.9 and the libraries used included scikit-learn, sentence-transformers and pandas. These models were trained and tested on a typical CPU-based machine, which was reasonable since TF-IDF and BM25 are relatively lightweight and the size of the dataset is moderate. With this setup, neural models could also be executed without the use of a GPU accelerator.

4.2 Testing and Evaluation/Performance/ Comparative Analysis

After preparing the appropriate environment and the dataset, it was tested and evaluated to understand the extent to which they are capable of retrieving the right legal information. It was assessed using both the statistical measure and visual measure through plots. These measures ensured that the analysis of the classical information retrieval mechanisms (TF-IDF, BM25) and the analysis of models on the basis of the transformer (MiniLM variations) was full-fledged and impartial. The qualitative insight of the model behavior had been represented in the evaluation metrics but the plots had allowed visualizations of the strengths and weaknesses, and relative trends. The same tools were also used in providing a balanced performance analysis. The following subsections define and elaborate the measures and graphs of this study.

4.2.1 Precision@5

Precision@5 calculates how many of the top five retrieved documents are truly relevant to the query. It measures the exactness of the system, showing whether the results provided are useful and correct. This is highly valuable in legal search systems where irrelevant information can confuse the user.

$$Precision@5 = \frac{Number\ of\ Relevant\ Documents\ in\ Top\ 5}{5}$$

4.2.2 Recall@5

Recall@5 measures the ability of the model to find relevant results among all possible correct ones, restricted to the top five outputs. A higher recall indicates that the system is able to capture more of the necessary information, which is vital when dealing with sensitive legal content where missing a law could be problematic.

$$Recall@5 = \frac{Number\ of\ Relevant\ Documents\ Retrieved\ in\ Top\ 5}{Total\ Number\ of\ Relevant\ Documents}$$

4.2.3 F1@5 Score

F1@5 is the harmonic mean of precision and recall, providing a balanced measure of both accuracy and completeness. In situations where precision and recall vary, the F1 score ensures a fair balance between them. It is particularly important when one metric alone cannot explain overall effectiveness.

$$F1@5 = 2 \times \frac{Precision@5 \times Recall@5}{Precision@5 + Recall@5}$$

4.2.4 Success Rate

Success rate measures the proportion of queries where at least one relevant document was found in the retrieved results. This metric reflects practical usability, as in real applications, even a single correct suggestion can be helpful for the user.

$$\text{Success Rate} = \frac{\text{Queries with At Least One Relevant Result}}{Total\ Queries}$$

4.2.5 Multi-Metric Comparison Plot

This bar chart displayed precision, recall, F1, and success rate side by side for each model. It allowed a direct comparison of all key measures, showing the overall performance in one view. By combining multiple metrics, it provided a balanced perspective on strengths and weaknesses. The uniqueness of this plot is that it did not isolate one score but presented a holistic evaluation, which is particularly useful in research where a single metric may not be sufficient to capture model effectiveness.

4.2.6 Precision vs Recall Plot

This scatter plot showed the trade-off between precision and recall for all models. Models positioned closer to the top-right corner indicated a better balance of accuracy and completeness. This visualization was important because it helped explain cases where a model may have high recall but low precision, or vice versa. By seeing both metrics together, the analysis revealed how well-rounded each model was, rather than focusing on isolated numbers. Such insights were crucial for choosing the most reliable method for legal queries.

4.2.7 Performance Heatmap

All the evaluation metrics were shown in the heatmap in a single grid with dark colors indicating stronger performance. This projection allowed prompt identification of the models that were consistently good in various measures. In comparison with single-value plots, the heatmap provided an overall pattern-based perspective in which not only scores were indicated but their interrelation between models as well. As an example, it had clearly shown the superiority of BM25 and TF-IDF over neural methods. It is unique because it can graphically present performance in a visual format in a concise, understandable manner.

4.2.8 Error Distribution and Reliability Analysis

This collection of plots explained the kind of errors of each model like partial matches, low relevancy, or perfect matches. Together with this there was the reliability analysis which demonstrated the consistency of each model given to queries. These plots played a critical role in the quest to go out of bare accuracy and discover the cause of the error. Such analysis provided more insights in the case of a legal application where some errors could be more considered relevant. These plots validated the reasons as to why BM25 was more consistent and reliable compared to neural counterparts by exposing weaknesses of such a model.

These measures and plots have helped to assess these models based not on the raw numbers, but also on their stability and functionality. BM25 appears to be the best all-round with a very close second by the TF-IDF algorithm, and the transformer-based models were found to be less consistent and had more chances to fail. The mathematical measures with elaborate visualizations meant that the evaluation was not only clear but also insightful thus upholding the selection of the final model confidently.

4.3 Results and Discussion

They examined the models to identify the capacity of different retrieval approaches in addressing legal questions that center on women and children in Bangladesh. This section indicates the results of the experiments and is closely followed by a discussion of the significance of the results. Findings have been reported in tabular form, plots amongst others, which all indicate the strengths, weaknesses, and general life adoption of each model. The possibility of a numerical analysis and a visual interpretation is what helps to see the comprehensible picture of approaches, which seem to be the most appropriate in this task.

4.3.1 Comparative Performance of Models

Table 4.3.1 shows the direct comparison of all four models on the main evaluation metrics. Among them, BM25 achieved the best performance overall, with the highest F1@5 value of 0.8834 and a success rate of 96.7%. TF-IDF followed closely, also maintaining a very strong performance with an F1@5 of 0.8674. Both of these traditional information retrieval models outperformed the transformer-

based neural models, which struggled with precision and consistency. The ranking clearly highlights that BM25 is the most effective model for this dataset, while TF-IDF is a reliable second option. Neural models, despite being modern, were less suited for this task.

Table 4.3: Model Comparison

Model	Precision@5	Recall@5	F1@5	Success Rate	Rank
BM25	0.8133	0.9667	0.8834	0.9667	1
TF-IDF	0.7867	0.9667	0.8674	0.9667	2
all-MiniLM-L6-v2	0.5067	0.8333	0.6302	0.8333	3
paraphrase-multilingual-MiniLM-L12-v2	0.4733	0.7667	0.5853	0.7667	4

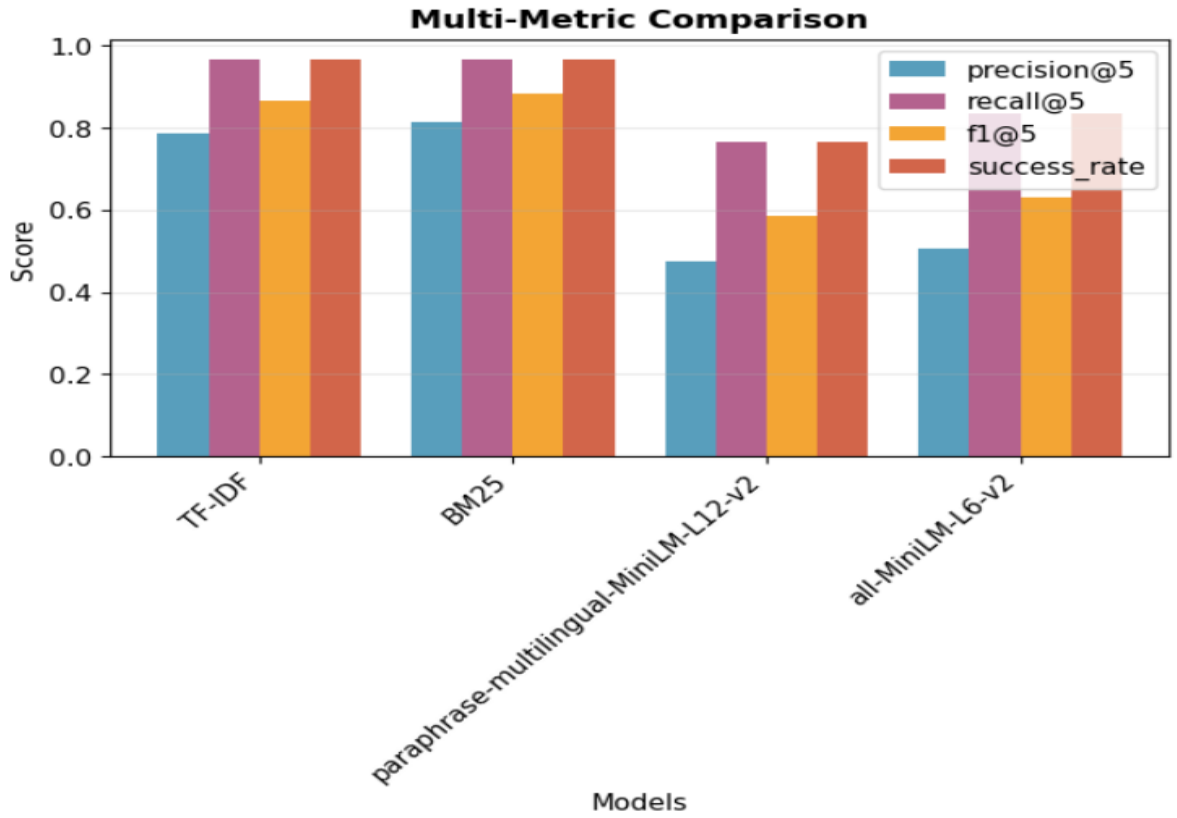

Figure 4.1: Multi-Metric Comparison Plot

This bar chart compares all four-evaluation metrics—Precision@5, Recall@5, F1@5, and Success Rate—for each model side by side. It gives a clear overview of how the models perform across multiple dimensions. The figure shows that BM25 consistently achieves the highest or near-highest values across all metrics, followed closely by TF-IDF, while both neural models fall behind. This

visualization is valuable because it highlights not just a single metric, but the overall balance of performance, making it easier to identify the most reliable model.

4.3.2 Statistical Validation of Results

Table 4.3.2 provides the statistical summary of evaluation metrics, including their average values and standard deviations across models. For F1@5 and Precision@5, BM25 and TF-IDF stand out as the strongest performers, while recall and success rate were tied between BM25 and TF-IDF. The neural models, all-MiniLM-L6-v2 and paraphrase-MiniLM-L12, consistently ranked lower across all metrics. The small standard deviations for BM25 and TF-IDF suggest their results are not only strong but also consistent, making them more dependable in practice. This reinforces the conclusion that traditional models provided both stability and effectiveness compared to neural approaches.

Table 4.4: Statistical Analysis Summary

Metric	Mean	Std Dev	Ranking	Model	Value
F1@5	0.7416	0.1349	1	BM25	0.8834
			2	TF-IDF	0.8674
			3	all-MiniLM-L6-v2	0.6302
			4	paraphrase-multilingual-MiniLM-L12	0.5853
Precision@5	0.6450	0.1557	1	BM25	0.8133
			2	TF-IDF	0.7867
			3	all-MiniLM-L6-v2	0.5067
			4	paraphrase-multilingual-MiniLM-L12	0.4733
Recall@5	0.8833	0.0866	1	BM25	0.9667
			1	TF-IDF	0.9667
			3	all-MiniLM-L6-v2	0.8333
			4	paraphrase-multilingual-MiniLM-L12	0.7667
Success Rate	0.8833	0.0866	1	BM25	0.9667
			1	TF-IDF	0.9667
			3	all-MiniLM-L6-v2	0.8333
			4	paraphrase-multilingual-MiniLM-L12	0.7667

Table 4.3.3 presents the results of statistical significance testing through pairwise t-tests for F1@5 scores. The analysis shows that BM25 and TF-IDF are statistically very similar, as the p-value between

them is above 0.05, meaning the difference is not significant. However, both BM25 and TF-IDF are significantly stronger than the transformer-based models, with p-values equal to 0.0000 and very large effect sizes. This result supports the claim that traditional IR methods are more reliable for this dataset, while neural methods failed to demonstrate comparable performance.

Table 4.5: Statistical Significance Testing (Pairwise t-tests for F1@5)

Comparison	Mean Difference	p-value	Effect Size	Significance
TF-IDF vs BM25	-0.0108	0.5047	0.213	Not Significant
TF-IDF vs MiniLM-L12	0.2741	0.0000	6.721	Significant
TF-IDF vs all-MiniLM-L6	0.2295	0.0000	4.970	Significant
BM25 vs MiniLM-L12	0.2849	0.0000	6.847	Significant
BM25 vs all-MiniLM-L6	0.2403	0.0000	5.123	Significant
MiniLM-L12 vs all-MiniLM-L6	-0.0446	0.0004	1.238	Significant

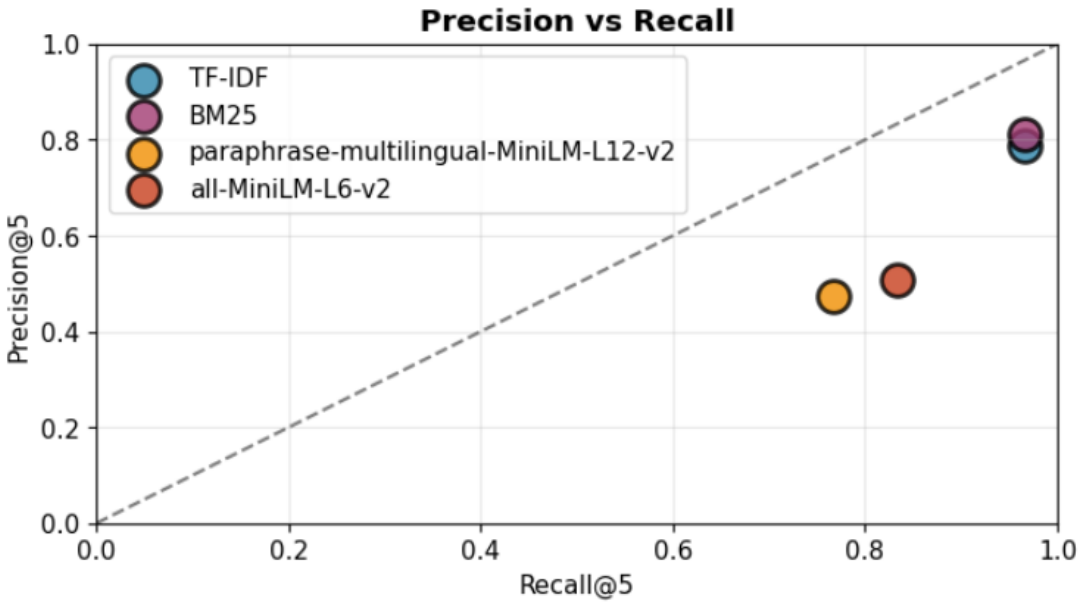

Figure 4.2: Precision vs Recall Plot

This scatter plot illustrates the trade-off between precision and recall for the evaluated models. Models that are closer to the top-right corner achieve both high precision and high recall, which is ideal in information retrieval. The plot clearly shows BM25 and TF-IDF clustered in the high-performance region, while the neural models lie much lower, indicating weaker accuracy and completeness. This figure is important because it reveals how well the models balance two critical aspects of retrieval,

which cannot be fully understood by looking at precision or recall alone.

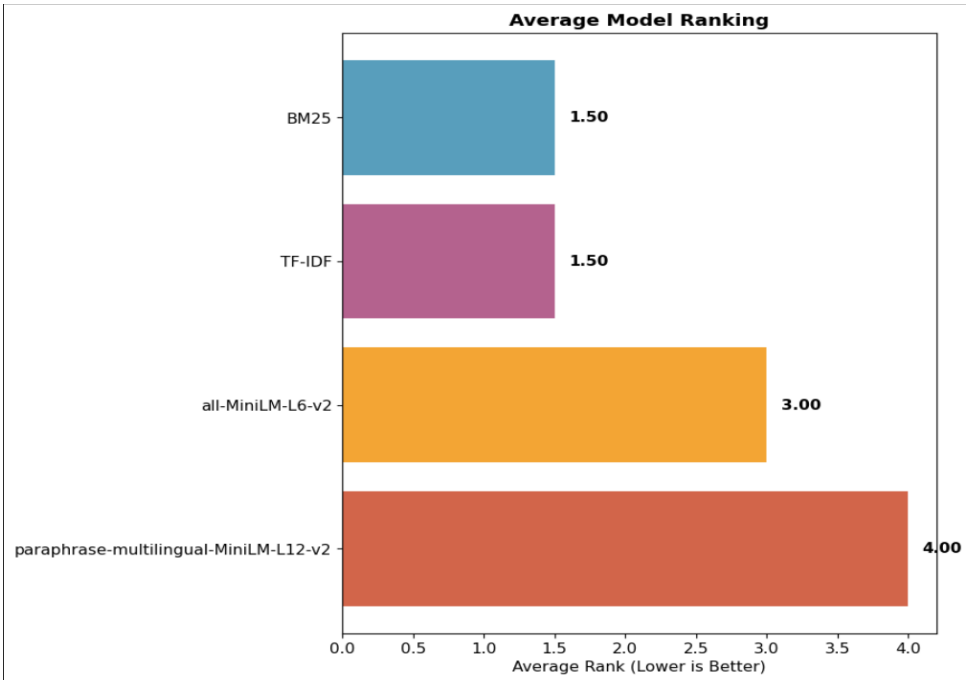

Figure 4.3: Average Model Ranking

This figure presents the average ranking of all evaluated models across multiple metrics. A lower value indicates better performance, with the ideal ranking being close to 1. Both BM25 and TF-IDF achieved the top average rank of 1.50, confirming that they consistently performed well across precision, recall, F1@5, and success rate. In contrast, the neural models scored much lower, with all-MiniLM-L6-v2 averaging 3.00 and paraphrase-MiniLM-L12-v2 falling last with 4.00. The figure highlights that traditional retrieval methods remain superior for this dataset, with BM25 and TF-IDF tied as the most effective and stable models. This ranking-based visualization strengthens the statistical validation by summarizing relative performance into a single, easy-to-interpret comparison.

4.3.3 Error Analysis and Reliability

Table 4.3.4 shows the comprehensive model performance, where additional measures such as balanced accuracy and geometric mean are also considered. These extended metrics strengthen the evaluation by accounting for the balance of precision and recall as well as stability across queries. BM25 once again takes the lead, followed closely by TF-IDF, while both transformer-based models underperform. The ranking is consistent across nearly all metrics, confirming the robustness of BM25 as the top model for this application.

Table 4.6: Comprehensive Model Performance Results

Rank	Model	Precision @5	Recall @5	F1@5	Success Rate	Balanced Accuracy	Geometric Mean	Queries Evaluated
1	BM25	0.8133	0.9667	0.8834	0.9667	0.8900	0.8867	30
2	TF-IDF	0.7867	0.9667	0.8674	0.9667	0.8767	0.8720	30
3	all-MiniLM-L6-v2	0.5067	0.8333	0.6302	0.8333	0.6700	0.6498	30
4	paraphrase-multilingual-MiniLM-L12	0.4733	0.7667	0.5853	0.7667	0.6200	0.6024	30

Table 4.3.5 provides a deeper analysis by including qualitative factors such as complexity, speed, and consistency. BM25 and TF-IDF are simple to implement, computationally efficient, and provide high stability, making them ideal for practical use in a legal assistance system. On the other hand, the neural models require higher computational resources, are slower to run, and show lower consistency, which makes them less suitable for real-world deployment without further fine-tuning or larger datasets.

Table 4.7: Detailed Performance Analysis

Model	Type	Complexity	Speed	F1@5	P-R Balance	Consistency	Recommendation
BM25	Traditional IR	Low	Fast	0.8834	0.1533	0.8414	Recommended
TF-IDF	Traditional IR	Low	Fast	0.8674	0.1800	0.8138	Recommended
all-MiniLM-L6-v2	Neural IR	High	Medium	0.6302	0.3267	0.6080	Recommended
paraphrase-multilingual-MiniLM-L12	Neural IR	High	Medium	0.5853	0.2933	0.6174	Not Recommended

Table 4.3.7 presents the detailed error analysis for TF-IDF. The majority of errors were partial matches, where retrieved documents were somewhat related but not fully accurate. Despite this, TF-IDF maintained a low failure rate overall, which highlights its reliability. Perfect matches were less

frequent, but the presence of good matches and low irrelevant results confirms its practical usefulness for legal query retrieval.

Table 4.8: Error Type Analysis (TF-IDF)

Error Type	Count	Percentage
No Results	0	0.0%
Low Relevance	4	8.0%
Partial Match	40	80.0%
Good Match	5	10.0%
Perfect Match	1	2.0%

Table 4.3.8 compares error distributions across all models. Both BM25 and TF-IDF showed very low failure rates of only 8%, demonstrating strong reliability in retrieving relevant documents. In contrast, the neural models had failure rates as high as 50%, with half of their results classified as low relevance. This diagnostic table highlights why BM25 and TF-IDF are far more dependable for legal information retrieval, while neural methods are not yet suitable without further refinement.

Table 4.9: Detailed Error Report

Model	Total Queries	No Results %	Low Relevance %	Partial Match %	Good Match %	Perfect Match %	Failure Rate %
TF-IDF	50	0.0	8.0	80.0	10.0	2.0	8.0
BM25	50	0.0	8.0	82.0	6.0	4.0	8.0
all-MiniLM-L6-v2	50	0.0	50.0	44.0	6.0	0.0	50.0
paraphrase-multilingual-MiniLM-L12	50	0.0	50.0	42.0	6.0	2.0	50.0

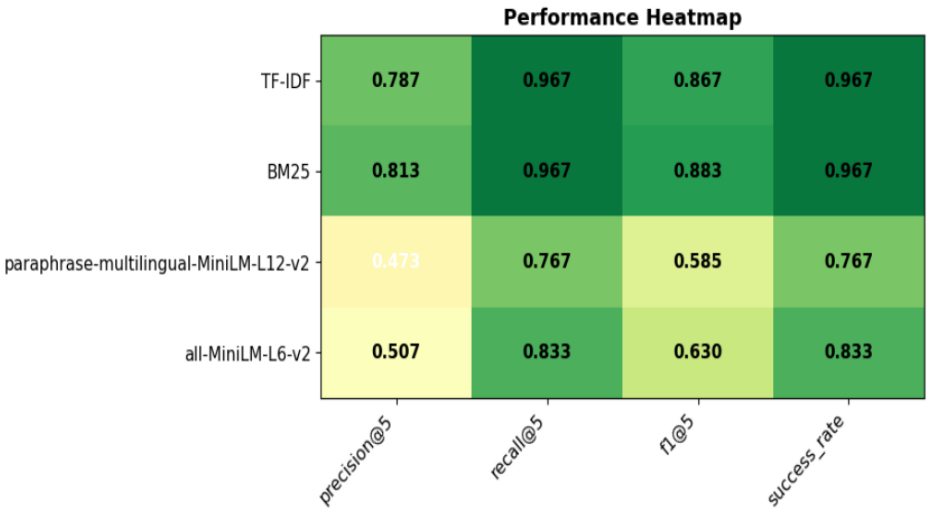

Figure 4.4: Performance Heatmap

The heatmap presents a grid of evaluation metrics for all models, where darker colors represent stronger performance. It provides an intuitive way to quickly identify patterns of strength and

weakness. The heatmap highlights that BM25 and TF-IDF dominate across most metrics, while neural models show lighter shades, reflecting poorer results. Unlike single-value charts, this visualization compactly summarizes the performance landscape, making it easy to spot which models are consistent leaders.

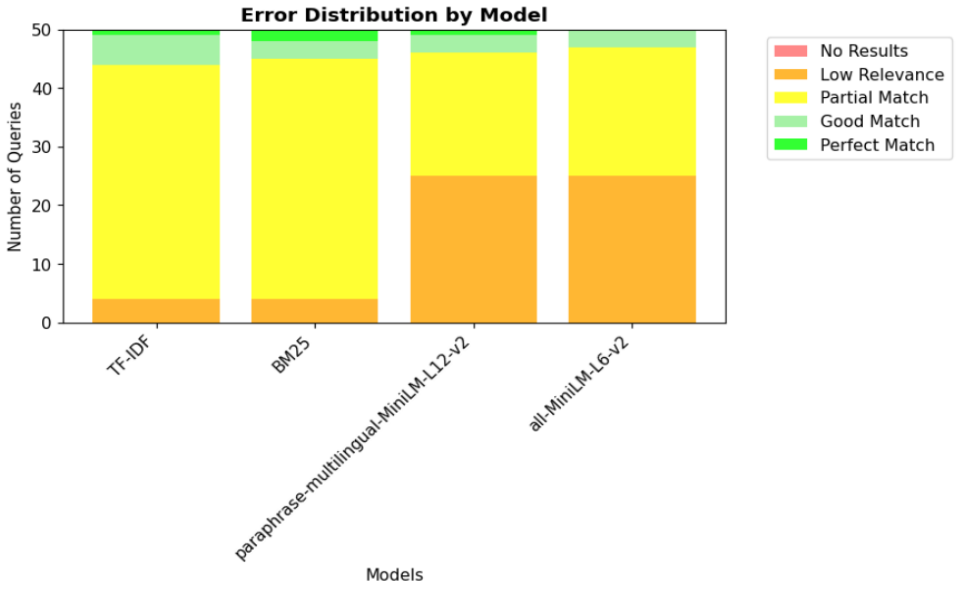

Figure 4.5: Error Distribution Plot

This plot depicts the rupture of the types of error of any of the models attracting no results, low relevance, partial match, good match, and perfect match. It demonstrates that BM 25 and TF-IDF do not perform as poorly as claiming full failure, but they normally give some matches, which is why in the neural models a big number of low relevance scores are produced. The importance of this amount is that it is a diagnostic consideration: not only to differentiate the functionality of the models, but also to describe the failure, and the rationale behind this, outlining an effective system of legal help.

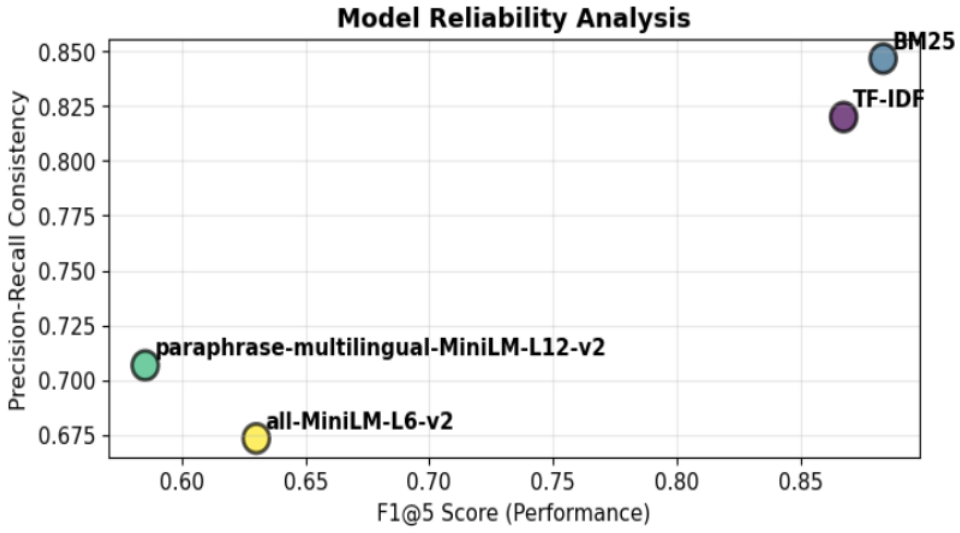

Figure 4.6: Model Reliability Analysis

This scatter plot presents the connection concerning F1n 5 (perf) against precision recall consistency (reliability) across all the models. BM25/TF-IDF frequents high-performance and high-consistency space, which denotes high stability. Contrary to this, both neural models are found in the lower part of the picture, which implies their inconsistency and lower power of retrieval. This Figure contributes to the assumption that BM25 and TF-IDF can be considered more reliable to use in retrieving legal queries.

4.3.4 Practical Implications for Legal Assistance

The best overall model performance of BM25 is demonstrated in Table 4.3.6 at query level. These outcomes indicate that longer and more complex queries are better when using BM25 whereby average values of F1 at 5 are higher. Why this is important is the fact that legal queries can be quite protracted, as well as include elaborate descriptions of cases and, therefore, BM25 fits perfectly well with practicalities of legal support.

Table 4.10: Query Performance Analysis (Best Model: BM25)

Query Length Category	Mean F1@5	Count
Short (1–3 words)	–	0
Medium (4–6 words)	–	0
Long (7–10 words)	0.760	5
Very Long (10+ words)	0.824	25

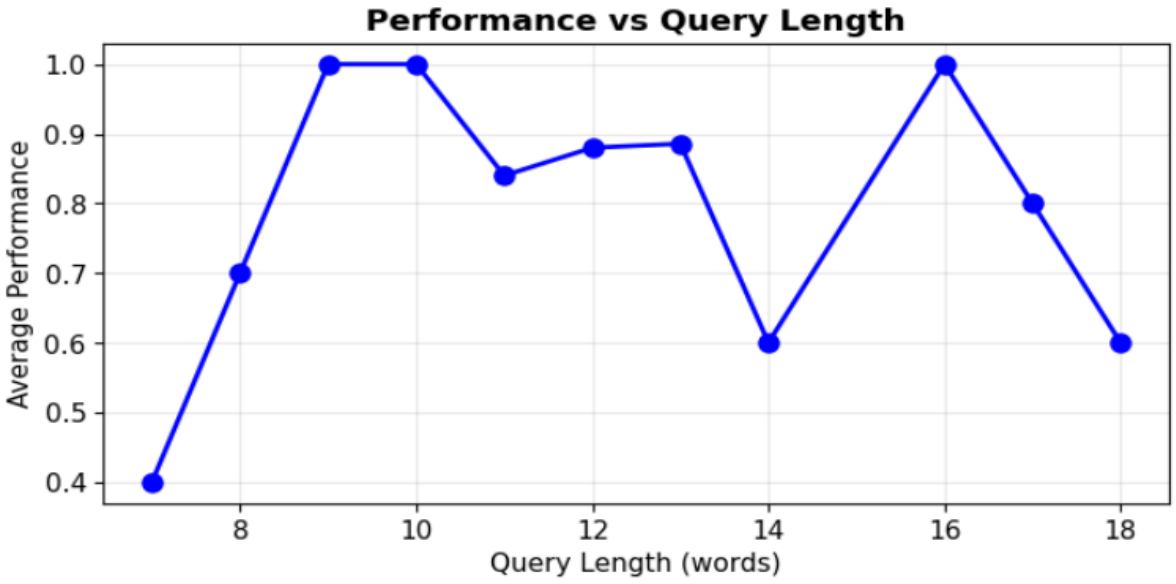

Figure 4.7: Query Length vs Performance Plot (BM25)

This plot demonstrates the performance of BM25 optimal model against queries of varying lengths. It points out that the model is particularly well-performing on longer queries, where very long queries possess the highest F1 at 5. The reason behind this is that real legal questions are not necessarily these

simple keys but made up of long and descriptive questions. The plot evidences that BM25 adjusts to such real-life conditions and, thus, is especially applicable in the field of legal assistance in Bangladesh.

4.4 Summary

The entire process of implementation and evaluation of the proposed legal information retrieval system was introduced in this chapter. The environment configuration was presented, and standard training parameters were applied, along with the typical dataset split, where all models are treated equally. The testing and evaluation section then presented the selected metrics of Precision 5, Recall 5, F1 5 and Success Rate as well as plots that graphically displayed the performance and trade-offs between accuracy and completeness. The results have shown clearly that the traditional information retrieval methods primarily BM 25 outperformed the transformer based neural models in precision and consistency. TF-IDF was also doing well, and came next after BM25, and the neural models were not doing well in terms of relevance and gave higher errors.

Statistical tests proved that there was no significant difference in the performance of BM25 and TF-IDF, but both were much better than neural models. The reliability analysis and error distribution also demonstrated the merits of traditional approaches as they were less prone to failure and their results were more predictable. Experimental evaluation of query performance demonstrated that BM25 was particularly effective on longer and more descriptive queries, which is also consistent with the format of queries in a legal case.

In general, the findings and discussion of the chapter confirmed the argument that BM25 is the most efficient, reliable and accurate model to be used on the given data. Under similar circumstances, TF-IDF is also a viable alternative. The results provide a solid ground on which this study is concluded and how the development of an AI-based legal support system helping women and children in Bangladesh can be improved in the future.

Chapter 5

Engineering Standards and Design Challenges

This chapter explains the engineering requirements and design issues that informed the design of the proposed legal assistance system. It describes the technical aspects including quality of data, choice of model, criteria of evaluation and scalability of the system, and generally considers adherence to software engineering best practices. The chapter also discusses some of the difficulties that arise in the design and implementation, such as managing the use of domain specific language, reliability in terms of different queries, and tradeoffs between efficiency and accuracy. When it comes to these areas the chapter gives a peep preview of the work around constraints and engineering choices that shaped the final system.

5.1 Compliance with the Standards

When designing NyaAI, technical, ethically, and legally, it has been a top priority to make the system reliable and socially responsible. Technically, the project is implemented according to the recognized requirements in data processing and software engineering, such as methodological data preprocessing, model testing, and implementation policies. To achieve the compatibility and reproducibility, open-source infrastructure and popular tools (TF-IDF, BM25, and MiniLM embeddings) were used, and the web application was designed based on the tenets of accessibility and usability that were already created to serve as many users as possible. In addition to technical considerations, the project also complies with the ethical considerations as it is culturally sensitive and has the needs of women and children in Bangladesh as its priority. In order to promote transparency and credibility all the legal information that is in use within the system is anchored on openly available and traceable sources. Moreover, the data selection and model response bias has been mitigated, as per the overall AI ethics and fairness principles. These standards would not only allow the project to achieve quality and accountability standards, but also help it enhance its reputation as a trustworthy and socially inclusive legal support system.

5.1.1 Software Standards

NyaAI has been developed using accepted standards in software in order to promote reliability, maintainability, and scalability of the system. It was implemented as per the normal coding standards, e.g. modular design, version control and sufficient documentation so that the system can easily be modified and extended in the future. The open-source libraries and frameworks that the project uses,

such as Python-based NLP preprocessing and embedding generation, enable the project to comply with the generally accepted development practices within the AI research community. In the web application case, the application was built with JavaScript and standard web technologies to enable cross platform compatibility and access. Further, the testing and assessment process was conducted in a systematic fashion, in such a way that the models could be applied systematically to other data sets and environments. They are not only practices based on the international software engineering principles of reusability, scalability and user-centered design, but also applicable to the real world, not to mention that the system is technically well-founded. These software standards enable the project to build a robust base on which it can be improved and expanded in the future.

5.1.2 Hardware Standards

NyaAI was also implemented in compliance with the right hardware standards to make its operations easy and accessible to end users. Because the system is meant to be an AI-based help device, it has been created and tested using hardware setups that fulfill the conditions to operate machine learning models effectively. Training and fine-tuning of the language models were trained using standard computing environments with multi-core processors, enough amount of RAM, and GPU acceleration during the experimentation. Nevertheless, the end-use was developed so that it can be operated on a medium hardware configuration such as personal computers or cloud hosting servers without having to use high-end hardware. This will allow legal aid organizations, non-governmental organizations, or individual users in Bangladesh to implement the system without having to experience significant infrastructure issues. When the project has covered these hardware requirements, the project can locate the balance of performance versus cost, and the product could become feasible and implemented on a larger scale in a real-life scenario.

5.1.3 Communication Standards

When coming up with NyaAI, the appropriate principles of communication were taken into account, so that the information flow within the system could be convenient, consistent, and secure. Because the tool relies on the ability to retrieve and provide the legal information and present the answers, using multiple communication protocols was chosen based on the ability to keep the information accurate, speedy, and secure. Having common APIs and web communication standards like HTTP/HTTPS were utilised to allow secure communication between the client remedy, database and model. This is because HTTPS facilitates encrypted communications particularly when one has to communicate sensitive queries that are associated with legal rights. Besides that, data exchange formats such as JSON were adhered to in order to make the system lightweight, interoperable and ready to be integrated with the mobile or the web based platforms in the future. Adhering to these communication standards the project is confident that users are able to depend on the reliability and

accessibility of the system as the warranties of trust, confidentiality, and efficiency are guaranteed in a legal assistance framework.

5.2 Impact on Society, Environment and Sustainability

The notion that NyaAI development has a serious social impact is evident because this development directly applies to addressing the relative conditions facing women and children in Bangladesh as they attempt to access legal information and justice. The system will enable the marginalized communities to better comprehend their rights and actively pursue them with a legal conference to ensure that it is a more socially responsible, protective, and inclusive society by offering an affordable yet as power as to pass a legal system due to the simplicity of its AI based support system. The other realm, in which the project and the personal empowerment zone will also be helpful is the legal aid organizations and the social workers themselves who will now have a new weapon that will allow them to approach the inquiries that are going on in a much more efficient manner. The system will be electronic in nature, which is environmentally friendly to ensure that as little paper work and travelling and consuming resources that can be attributed to legal consultations is reduced. This will help to minimize the environmental impact of the project, but also promote a more sustainable form of providing legal services. It is also evolved as an architecture designed on sustainability approach, open-source-designed, cloud-architecture, scalable architecture, allowing the solution to be supported and extended in the long-term, and not burdened with heavy-load infrastructure to support it. As a consistent rule, the development of the technology innovation and the social responsibility to the project will lead to the establishment of the long term bets which will not be short-termbased legal short-benefit and will lead towards the establishment of the more sustainable society in a more fair fashion.

5.2.1 Impact on Life

NyaAI is pertinent to the life of human beings because the introduction of such knowledge has helped in making the law more accessible to women and children who tend to be the most vulnerables in society. The system facilitates equal play and enables individuals to make people make prudent decisions when required as it eradicates the employment of expensive legal counsel and processes that require time. It entails investing in personal empowerment besides empowering communities where rights and protections are garnered towards understanding. Regarding the environmental consideration, the system works entirely under the digital setting and reduces the consumption of the physical resources such as the paper documentation and the printed manuals of the law. The sustained environmentally positive practices and the adoption of wastage, are helpful in facilitating this shift to a digital-first solution. It is also a long term and scalable project. Open-source tools and lightweight architectures enable maintaining the system, upgrading/adaptation without any costs that would mean

to compromise the costs and environmental effects. Overall, the project can be considered as sustainable in growth and development in terms of the social, environmental responsibilities; and in the long term, both legal and ethical support can be deemed viable and morally upright.

5.2.2 Impact on Society & Environment

The social and environmental contribution of NyaAI is directly connected to the fact that it is a digital legal aid that helps women and children in Bangladesh. The project will on the societal front contribute to the reduction of inequalities by making the law more readily available to groups which are generally inaccessible to formal law. This can increase the trust of the community, promote justice-seeking and social development in the long-term. Regarding the environment, the system reduces the use of resources by encouraging the use of digital access to legal information rather than the use of printed material or physical consultations. This limits the indirect carbon footprint usually associated with paperwork, travel and manual handling of cases. On the sustainability side, the system was developed based on open-source materials, and, therefore, it can be expanded and even sustained without being costly in terms of resources and relying on proprietary technologies. The social, green, sustainable design will ensure that the project will be in a position to deliver the services to the societies in an efficient manner, in addition to being aligned to the key global goals of justice, environmental protection and responsible innovation.

5.2.3 Ethical Aspects

The design and implementation of NyaAI are ethics-focused because the system addresses sensitive legal questions related to women and children. Privacy, confidentiality, and fairness were deemed to be of utmost importance, during the project to ensure that users were not misusing their information or responding to questions with bias. The system does not present any wrong or problem advice since the system is limited to information support, which informs the customer on the correct legal framework to refer to against going around good judgment of the professional. That protects the ethical line between artificial intelligence and legal knowledge. The system is founded on the digital communication and the open source systems, which does not depend on the physical resources and the proprietary costly solutions, which are ecologically and sustainably unsustainable. Scalable and lightweight architecture also allows long-term sustainability in the sense that a system can be changed and re-architectured to meet the requirement of the society and technology throughout the period. With ethics and environmental responsibility and sustainable growth, the project can not only overcome the issue of instant legal access but establish the basis of responsible and responsible AI implementation in the future.

5.2.4 Sustainability Plan

NyaAI sustainability plan aims at making the system efficient, flexible, and environmentally friendly in the long term. Because the project would be used to educate women and children about legal issues, it has been developed to be easily scaled and long-term maintainable. Based on open-source frameworks and cloud-ready architecture enables the system to be upgraded with new legal information and enhanced services without significant financial or hardware investments. This will make the tool available even after technology and laws change to suit the community. Environmentally, the system is all-digital, which reduces the amount of paper, travel, and other activities that require resources, which helps to promote eco-friendly behaviour. Regarding the sustainability training modules, the community involvement strategies could be implemented to generate awareness and acceptance in order to expand the scope of the system compared to the initial one. Incorporating a combination of social responsibility, digital efficiency, and environment-consciousness this sustainability plan will ensure the project continues to have a significant impact without rendering it unfeasible, costly, or unreflective of the long-term requirements of society.

5.3 Project Management and Financial Analysis

The key to the completion of this project on the creation of NyaAI, an AI-based women and child legal assistance system in Bangladesh was successful project management and budget planning. As the project was about natural language processing, retrieval-augmented generation (RAG) and mobile-deployment, precision, scalability and cost-efficiency were important. In this section, the way the project was planned, implemented, and controlled within time and resources limitations and made reproducible and accessible is discussed.

5.3.1 Project Planning and Task Management

Flexibility and improvement were introduced into the system through a phase-based system which was built on agile techniques. The studies were broken down into separate phases with particular deliverables and checkpoints. These phases were dataset collection, preprocessing, model integration, retrieval and generation pipeline design, evaluation, and mobile application deployment.

The first step was the gathering and preservation of Bangladeshi legal documents, especially those laws that related to women and children. This was then followed by preprocessing that involved cleaning text, tokenization, removing stop-words, as well as proper formatting of statutes and case references. Following dataset preparation, the information was consolidated as a vector database to be accessed in the future and linked to a Llama-based generative model via a RAG pipeline.

The second phase was devoted to the development of the system and its testing, during which different retrieval tactics and measurement measures (Recall@k, BERTScore, semantic similarity) were implemented. Particular attention was given to the explainability of responses, so the generated answers contained appropriate links to the source documents.

The last phase was to build and test the mobile application, make the system user friendly, multi

linguistic (Bangla, English) and made it accessible to non-technical users. The transparency and reproducibility were enabled by version control on GitHub and experiment tracking on the Jupyter Notebook.

1. Resource and Time Allocation

The project took 5 months, the work was broken down into the following stages:

- Month 1: Literature review, text cleaning, exploration and dataset collection.
- Month 2: Preprocessing, database generation, and setting up of the vector embedding.
- Month 3: RAG pipeline architecture and connection to Llama based model.
- Month 4: System (Recall@k, BERTScore) evaluation/refinement, robustness verification.
- Month 5: Mobile application development, implementation, report writing, and documentation.

Google Colab also minimized the cost of computation by taking advantage of free GPU access, but open-source libraries ensured the system was economical.

2. Financial Analysis

The advantage of this project is that it is cheaply set up. Open-source frameworks in use, the free availability of legal documents, and free-tier cloud services all helped to keep down the costs. This ensures that the solution can be scaled and made available to developing areas such as Bangladesh. The result will be a mobile-based, explainable legal assistance system that can be rolled out on a large scale, free of the cost of a license.

3. Scalability and Future Investment.

The modular structure is scaled. More laws, court rulings, or sophisticated models can be easily added to the system. New investment can include the inclusion of a higher-end GPU service, speech-to-text and text-to-speech capabilities, or extending the application to enterprise-scale legal applications. Nevertheless, the present configuration proves that the development of a powerful and inexpensive legal AI assistant is already possible with limited resources.

The course of development had definite phases:

- Requirement Analysis and System Design - scope definition, choice of algorithms.
- preparation and preprocessing of dataset-text cleaning, embedding, indexing.
- Model Development - combination of RAG pipeline and Llama-based model.
- Evaluation - Recall@k, BERTScore and user query testing.
- Deployment - creating a mobile application interface to be used in the real world.
- Documentation- writing reports, storing results, and making results reproducible.

5.3.2 Tools and Platforms Used

Table 5.1 provides details of the main tools and platforms used throughout the project.

Table 5.1: Details of Tools and Platforms Used

Category	Tool/Platform	Purpose
Programming Language	Python	Core NLP and ML development
Libraries/Frameworks	Hugging Face, LangChain, FAISS, Transformers	Model training, RAG integration, embeddings
Model Training	Google Colab (GPU)	Scalable training and testing
IDE	Jupyter Notebook	Experiment tracking and visualization
Version Control	Git/GitHub	Code versioning and collaboration
Storage/Backup	Google Drive	Dataset storage and sharing
Mobile Framework	Flutter	Mobile app development for Android

5.3.3 Financial Analysis

This project was designed to remain cost-effective and sustainable, making it suitable for student-level and institutional research. The cost breakdown is shown below.

Table 5.2: Estimated Cost and Financial Analysis

Resource/Item	Estimated Cost (BDT)	Remarks
Google Colab Pro (optional)	3500	Faster GPU for training (optional)
Computer (personal/lab use)	Existing Resource	Used for system design and testing
Internet Access	4200	For cloud training and dataset access
Python Libraries/Frameworks	0	Open-source (Hugging Face, LangChain, FAISS)
Cloud Storage/Backup (optional)	300	Google Drive or equivalent
Mobile App Development (Flutter SDK)	0	Free and open-source
Total	8000 BDT	—

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

NyaAI development is also a complex engineering problem because it incorporates legal knowledge, natural language processing, AI architecture, and support within the sector. Not only does the system process laws and regulations, it also offers a context-sensitive and socially impactful response. Table 5.1 is the mapping with the categories of Complex Engineering Problem (CEP).

Table 5.1: Mapping with Complex Engineering Problem.

EP1	EP2	EP3	EP4	EP5	EP6	EP7
Depth of Knowledge	Conflicting Requirements	Depth of Analysis	Familiarity of Issues	Applicable Codes	Stakeholder Involvement	Interdependence
✓	✓	✓	✓		✓	✓

Justifications:

- EP1: Depth of Knowledge: The thesis demanded expert skills on both law and artificial intelligence. Knowledge of the law in regard to women and children, together with the development of hybrid AI solutions, demonstrates a high level of knowledge.
- EP2: Range of Conflicting Requirements: There were conflicts of accuracy vs. speed, English vs. Bangla language processing, legal formalism vs. user-friendly outputs. The methodology struck a balance between these demands by taking a hybrid retrieval and semantic method.
- EP3: Depth of Analysis: The TF-IDF, BM25 and transformer-based models were extensively evaluated. This guaranteed the analysis was not limited to surface-level testing and involved comparative experimentation and hybrid incorporation.
- EP4: Familiarity of Issues: Access to legal aid in Bangladesh is a relatively new problem in AI research, and thus, there are limited solutions. The unfamiliarity was addressed in the thesis by using the literature and adapting the solution to the Bangladeshi legal context.
- EP6: Stakeholder Involvement: Women, children, NGOs and policymakers are stakeholders. Their role in the project is recognized by designing easy-to-use web interface and deployable implementation.
- EP7: Interdependence: A synthesis of law, language, computing and social justice was needed. The domains were interdependent and therefore the problem was very interdependent.

Mapping with Knowledge Profile

Since EP1 (Depth of Knowledge) is central, it maps further into the Knowledge Profile, as shown in Table 5.2.

Table 5.2: Mapping with knowledge Profile.

K1	K2	K3	K4	K5	K6	**K7**	K8
Natural Science	Mathematics	Engineering Fundamentals	Specialist Knowledge	Engineering Design	Engineering Practice	Comprehension	Research Literature
✓	✓	✓	✓	✓	✓	✓	✓

Justifications:

- K1: Natural Science – Applied through the understanding of natural language phenomena and computational linguistics, treating language as a natural system for information processing.
- K2: Mathematics – Used in statistical evaluation metrics (Precision, Recall, F1-score, Success Rate), vector space models, and optimization in BM25/TF-IDF.
- K3: Engineering Fundamentals: Applied in understanding algorithmic foundations like vector space models, embeddings, and retrieval methods.
- K4: Specialist Knowledge: Knowledge of *AI, NLP, and Bangladeshi law* was combined, making this specialist knowledge application essential.
- K5: Engineering Design: System design included a hybrid model, data pipeline, and a web application for deployment.
- K6: Engineering Practice: Testing models, fine-tuning, and integrating into a web app reflect professional engineering practice.
- K7: Comprehension: Understanding user needs (non-expert Bangladeshi citizens) required interpreting legal texts into simplified outputs.
- K8: Research Literature: The literature review (Section 2.2) guided model selection, gap identification, and problem framing.

5.4.2 Engineering Activities

The project also qualifies as complex engineering activity, since it required innovative solutions, multi-domain knowledge, and consideration of social impact. The mapping is shown in Table 5.3.

Table 5.3: Mapping with Complex Engineering Activities.

EA1	EA2	EA3	EA4	EA5
Range of Resources	Level of Interaction	Innovation	Societal & Environmental Consequences	Familiarity
✓	✓	✓	✓	✓

Justifications:

- EA1: Range of Resources: Used legal documents, AI frameworks, computing resources, and web deployment tools, all integrated into one solution.
- EA2: Level of Interaction: The system interacts with end users (citizens), technical models, and legal knowledge bases, showing multi-level interaction.
- EA3: Innovation: Introducing a Bangladesh-focused AI legal assistant that combines retrieval + semantic AI is innovative in this context.
- EA4: Consequences for Society: The system supports social justice by improving access to legal guidance for women and children, reducing inequality.
- EA5: Familiarity: The problem was not entirely familiar due to the novelty of AI in

Bangladeshi legal aid. The thesis tackled this by adapting known models into a new domain.

5.5 Summary

This chapter explained the engineering standards and the design issues that were considered during the creation of NyaAI. Given the reliability, security and accessibility of the system, there were standards of software, hardware and communication that the system was constructed to meet. It also discussed the social, environmental and sustainability benefits of the project, the fact that it empowers women and children by providing them with easy access to legal support and alleviates the need to be dependent on tangible resources. Another aspect that was considered and the sustainability plan was established to provide a long-term efficiency with minimum costs and impact to the environment is the moral factor and the privacy of the user and equity. Overall, as demonstrated in this chapter, the project does not just satisfy technical criteria but also fulfills wider social obligations, which qualifies NyaAI as a viable, ethical and sustainable legal support solution in Bangladesh.

Chapter 6

Conclusion

This chapter provides a summary of the key findings of the research and considerations made about the contributions of the thesis. It focuses on how BM25 was the most appropriate model to support legal query retrieval in Bangladeshi setting. Other limitations of the research (including the size of the dataset and the performance gap of neural models) are also recognized in the chapter. Finally, it gives future study directions including developing improved transformer models and updating the dataset and improving the web application to be used by a greater number of people. This chapter is the conclusion of this thesis because it links the results of the research with the initial goals.

6.1 Summary

This thesis put forward to design and develop a legal query assistant to women and children in Bangladesh with the deployment of AI, NyaAI. The current problems were behind this work of limited access to the buy of justice, unavailable legal aid, and notably the overwhelming intricacy of law text, which prevents vulnerable populations by an even larger margin to persist in their entitlement or to enact their rights. Reviewing the existing literature, it was obvious that despite successfully welcoming AI to offer itself in the realm of a legal discipline globally, the application is relatively small in scale, terms, and delivery in Bangladesh. It is the gap on these grounds that this current study is founded.

The systematic research technique was applied to this research and this combined both the traditional retrieval methods as well as the existing techniques of semantic retrieval. The data was reassembled and pre-tapped to divide the relevant legislation and legal environment in Bangladesh. The traditional models such as TF-IDF model and BM25 model and the semantic models such as the transformer-based models have been tried on their strong and weak side. A balanced plan was then formulated in a way that it provides a balance between retrieving and contextual accuracy whose resultant effects are obtained of both having correct, as well as meaningful responses. Another requirement was the implementation of the system as a web-based system with ease of access to its users.

The findings of this thesis measure that in the occurrence of resource-constrained situations blocked along with retrieval and semantic reasoning legal responses will have a greater probability of being correct. Unlike in the past where there could be only a few prototypes, English [2] being limited to only few issues concerning women of which [1] was considered only to a bit [2], NyaAI allows women and children [2] and even allows the usage of more than one language. Moreover, the fact that the focus of deployment moves out the study it is in relation to the theoretical frameworks into a concrete system that can be extrapolated into a real-life application.

Overall, the study is pertinent to the growing area of legal AI research since it provides an example of

a sophisticated approach that can be integrated into local and socially significant implementation. By addressing the technical, language, and social divides in accessing the legal system, NyaAI will partially bridge the disparities between Pakistan and Bangladesh and bring about greater inclusion in access to the rule of law in Bangladesh. In the below sections, the implications of this study, limitations, and future labor direction are given.

6.2 Limitation

This paper has already reached certain important milestones in the development of an AI-based AI-powered agent, which can be used to answer a legal query, but it also has several limitations, which must be admitted. To begin with, the training and assessment information was limited and restricted. Some of the most significant laws and chapters, which apply to women and children in Bangladesh, were represented; nevertheless, the law sphere is massive and constantly growing. This is because NyaAI should not necessarily put the less-common cases and other new legal provisions into the system and this may impact the system on the aspect of completeness and accuracy of the system.

Second, the system is founded on the hybrid mechanism which incorporates the integration of the traditional retrieval and semantic models. This boosts both accuracy and situational information but this also has to be calculated. Advanced transformer-based models need more resources to operate; and might be limited in terms of scalability in low-resource contexts. This implies possible difficulty among the end users to operate the system in low power processors or devices with poor internet connectivity.

Third, the responses to the questions that NyaAI arrives at are able to provide legal information and advice, but they can not replace legal advice of professional advice. The system is not able to introduce the application of the laws to the specific situations or deliver the verdict in the specific case. This is a critical restriction, as the fact that output of AI can be easily assumed by users to be a final legal decision, rather than a point of departure. It is significant that the system consequently clarifies a fact which it is and advisory.

Finally, the study was restricted to the web-based implementation. Even though this enables the system to be accessible to a vast population, it exposes populations that are poor in digital skills or have limited access to internet out. A mobile friendly or an offline enabled interface would be needed in the NyaAI to actually achieve the benefits of the technology into the communities that would not have considered technology in Bangladesh as a high priority.

The identification of these limitations is important in that it shows where growth and research on them ought to improve in future. The system can gain more power and incorporate more individuals in the initiatives to meet the legal requirements of women and children by addressing the challenges of data expansion, resource utilization, perceptive application, and conscious audience.

6.3 Future Work

The outcome of this study leads to the fact that NyaAI can develop in a variety of ways and become even stronger and more efficient. One of the most significant spheres is the continuation of data set. The system is now majorly focused on women and children laws and regulations in Bangladesh. The implementation of the system in the future would have the prospect of expanding such an area of law to cover wider components such as labor rights, property disputes, and even cybercrime where applicable. It should also update the information at periodic moments to help in facilitating correction and other policies which would render the assistant relevant and authoritative.

Such a step as incorporation of more sophisticated natural language processing models should be taken as the other one. Despite the already promising results achieved with the proposed hybrid system of retrieval and semantic models, it is possible to attain the level of higher precision and contextualization when large-scale transformer-based models are being used developing to perform their work with the Bangla texts on the law. One may also assume that the future sees it becoming multilingual which will provide the system with the opportunity to successfully interact not only with the Bangla and English language, but it will also interact with the local dialects which will be closer to a majority.

Access and usability should be the future optimization in the deployment aspect. It would be great to have a version of a NyaAI that is offline-enabled or an applications gets developed that is available on the phone and is utilized by users who have sluggish access to the internet or those that use their mobile phones as a general device. Interactive features can also make the system more accessible, particularly to those with low level of digital literacy, like voice it can also be used to make a query, chat-like user interfaces, and tailored suggestions.

Lastly, it will provide contacts with the legal community, non-governmental organizations, and government agencies, which will lead to the further development of the project. This form of partnership would mean that the system does not only avail to the user the right kind of legal information, but the platform the user can use the services of the legal aid when they need them. NyaAI can be a more resilient technology-meets-justice interface by connecting advice through AI applications to in-the-field legal services.

References

[1] Hasan, K. S. (2025). JusticeNetBD: A Retrieval-Augmented Generation-based Legal AI Assistant for Women in Bangladesh.1

[2] Wasi, A. T., Faisal, W., Islam, M. R., & Bappy, M. M. (2024). Exploring Possibilities of AI-Powered Legal Assistance in Bangladesh through Large Language Modeling. arXiv preprint arXiv:2410.17210.

[3] Reza, M. R., Mannan, F. M. B., Barua, D., Islam, S., Khan, N. I., & Mahmud, S. R. (2021, November). Developing a machine learning based support system for mitigating the suppression against women and children. In 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT) (pp. 1–6). IEEE.

[4] Chowdhury, M., & Noor, J. H. (2024, March). Safe-Guard: A Smartphone Application with Child Abuse Detection, Doctor Selection, Legal Aid, and Child Abuse Case Reporting Help Assistance. In 2024 IEEE International Conference on Contemporary Computing and Communications (InC4) (Vol. 1, pp. 1–6). IEEE.

[5] Sultana, S., Chowdhury, H. M., Sultana, Z., & Verdezoto, N. (2025, July). 'Socheton': A Culturally Appropriate AI Tool to Support Reproductive Well-being. In Proceedings of the 2025 ACM Designing Interactive Systems Conference (pp. 3098–3116).

[6] Gupta, J., Sharma, A., Singhania, S., & Abidi, A. I. (2025). Legal Assist AI: Leveraging Transformer-Based Model for Effective Legal Assistance. arXiv preprint arXiv:2505.22003.

[7] Bhatnagar, M., & Huchhanavar, S. (2025). Enhancing legal assistance with AI: a comprehensive approach to intent classification and domain specific model tuning. Artificial Intelligence and Law, 1–29.

[8] Ujwal, U., Surampudi, S. S. H., Mitra, S., & Saha, T. (2024, October). "Reasoning before Responding": Towards Legal Long-form Question Answering with Interpretability. In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (pp. 4922–4930).

[9] Italiani, P., Moro, G., & Ragazzi, L. (2025). Enhancing legal question answering with data generation and knowledge distillation from large language models. Artificial Intelligence and Law, 1–26.

[10] Wang, C., & Luo, X. (2021, December). A legal question answering system based on BERT. In Proceedings of the 2021 5th International Conference on Computer Science and Artificial Intelligence (pp. 278–283).

[11] Garlapati, A., Koutharapu, H., & Doddi, N. (2025, February). Enhancing Public Access to Legal Knowledge in India: A Legal Chatbot Using Legal BERT, GPT-2, and Retrieval-Augmented Generation (RAG). In 2025 IEEE International Conference on Emerging Technologies and

Applications (MPSecICETA) (pp. 1–6). IEEE.
[12] Singh, M. R., Ahmad, S., Sharma, K., Sharma, P., & Kaushik, P. (2025, April). AI-Driven Legal Assistance: Optimizing Document Retrieval and Case Classification with NLP Techniques. In 2025 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE) (pp. 1–7). IEEE.
[13] Vuong, Y. T. H., Bui, Q. M., Nguyen, H. T., Nguyen, T. T. T., Tran, V., Phan, X. H., ... & Nguyen, L. M. (2023). SM-BERT-CR: a deep learning approach for case law retrieval with supporting model. Artificial Intelligence and Law, 31(3), 601–628.
[14] Kumar, A., Aswanth, V. R., Saikrishna, V., Sairam, N., & Palanikumar, R. (2025, March). Revolutionizing Legal Workflows: Advanced AI Techniques for Document Summarization, Legal Translation, and Conversational Assistance. In 2025 International Conference on Advanced Computing Technologies (ICoACT) (pp. 1–4). IEEE.
[15] Nithya, M., Harini, S., Kavyadharshini, S., & Srinidhi, K. (2024, December). AI-Driven Legal Automation to Enhance Legal Processes with Natural Language Processing. In 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS) (pp. 1246–1253). IEEE.
[16] Jacob, S., Jacob, P. M., Cheriyan, J., & Ismail, S. (2024, September). Legal Assistance Redefined: Transforming Legal Access with AI-Powered LegalLink. In 2024 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES) (pp. 1–6). IEEE.
[17] Judijanto, L., & Vandika, A. Y. (2025). Emerging Research Trends in Natural Language Processing for Multilingual AI. The Eastasouth Journal of Information System and Computer Science, 2(03), 187–199.
[18] Kayalvizhi, S., Thenmozhi, D., & Aravindan, C. (2019). Legal Assistance using Word Embeddings. In FIRE (working notes) (pp. 36–39).
[19] Basu, P., Bhandari, D., & Sengupta, K. Using Machine Learning Algorithms With TF-IDF to Generate Legal Petitions.
[20] Panda, B. K., & Mohapatra, M. Creating Legal Petitions Using TF-IDF and Machine Learning Algorithms.
[21] Bhattacharya, P., Ghosh, K., Ghosh, S., Pal, A., Mehta, P., Bhattacharya, A., & Majumder, P. (2019, December). FIRE 2019 AILA track: artificial intelligence for legal assistance. In Proceedings of the 11th annual meeting of the forum for information retrieval evaluation (pp. 4–6).
[22] Aydemir, A., de Castro Souza, P., & Gelfman, A. (2020, November). Using BERT and TF-IDF to predict entailment in law-based queries. In JSAI International Symposium on Artificial Intelligence (pp. 286–293). Cham: Springer International Publishing.
[23] Arora, J., Patankar, T., Shah, A., & Joshi, S. (2020, December). Artificial Intelligence as Legal Research Assistant. In FIRE (working notes) (pp. 60–65).
[24] Kim, M. Y., Rabelo, J., Okeke, K., & Goebel, R. (2022). Legal information retrieval and

entailment based on BM25, transformer and semantic thesaurus methods. The Review of Socionetwork Strategies, 16(1), 157–174.
[25] KS, T., & Aravindan, C. (2020). Best matching algorithm to identify and rank the relevant statutes. Proceedings of FIRE.

NyaAI - An AI Powered Assistance for Legal Queries About Women & Children in Bangladesh.pdf

ORIGINALITY REPORT

11%	7%	5%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	4%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	Submitted to United International University Student Paper	<1%
4	jisem-journal.com Internet Source	<1%
5	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
6	Submitted to University of Teesside Student Paper	<1%
7	arxiv.org Internet Source	<1%
8	Submitted to University of Hertfordshire Student Paper	<1%
9	Yiwen Wang, Xiaoke Qi, Xiaobing Zhao. "A Large Language Model Evaluation Method for Legal Case Retrieval", Data Intelligence, 2025 Publication	<1%