Lawvia AI: ML-Powered Legal Assistance with Authority Routing and Document Risk Analysis Using NLP

Jayakrishnan U V[1], Aparna Prakash[2], R Govinda Sivam[3], Anas N S[4], Alfie G Anil[5],

Ancey Varghese[6]

Research Scholar, Computer Science & Engineering Department, St. Thomas Institute for Science and Technology,

Trivandrum, India[1-5]

Assistant Professor, Computer Science & Engineering Department, St. Thomas Institute for Science and Technology,

Trivandrum, India[6]

Abstract: Lawvia AI is an intelligent legal assistance and document analysis platform designed to improve legal awareness and accessibility. Users such as students, employees, and tenants often face legal issues including harassment, cybercrime, and rental disputes without adequate guidance. The system accepts text or voice input and applies Natural Language Processing (NLP) to identify the legal context. A BERT-based transformer model embedded in the chatbot performs legal intent and domain classification, enabling simplified explanations of relevant laws and recommended actions. Lawvia AI includes a Complaint Draft & Action (DocuAction) module for generating structured complaint letters with options to download or forward them to external authorities, and a Legal Support (LegalAid+) module that provides access to nearby legal aid resources. The platform also supports document risk analysis, where uploaded legal documents are processed using Optical Character Recognition (OCR) and NLP-based keyword mapping to detect risky clauses and generate simplified summaries. Additionally, the ESpace authority-routing module assigns complaints to appropriate internal authorities based on severity. The system is implemented using ASP.NET with C#, integrates Python-based NLP components, BERT models, LLM APIs, and SQL Server, offering a modular and socially impactful legal assistance solution.

Keywords: Legal Assistance, Document Risk Analysis, Natural Language Processing (NLP), Machine Learning (ML), Complaint Letter Generation, Authority Routing.

I. INTRODUCTION

Legal awareness is a fundamental requirement for ensuring justice, equality, and social stability in modern societies. However, a significant portion of the population—particularly students, employees, tenants, and first-time legal users—lacks adequate understanding of their legal rights, responsibilities, and available remedies. This lack of legal literacy oftenprevents individuals from taking timely and informed action when faced with issues such as workplace harassment, cybercrime, consumer fraud, or rental disputes.

Legal documents including employment contracts, offer letters, and rental agreements are typically drafted using complex legal terminology and formal structures that are difficult for non-experts to comprehend. As a result, individuals may unknowingly agree to unfavorable clauses, hidden penalties, or restrictive conditions. In many cases, fear of legal complexity, financial constraints, and uncertainty about appropriate procedures further discourage individuals from seeking professional legal assistance, leading to unresolved grievances and exploitation.

Recent advancements in Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML) have introduced new opportunities for improving access to legal information. AI-powered legal chatbots and document analysis systems can assist users by interpreting legal text, identifying relevant legal domains, and presenting simplified explanations of laws and procedures. Transformer-based language models such as BERT have demonstrated strong performance in understanding contextual meaning within legal text, making them suitable for intent classification and domain identification tasks. Despite these advances, most existing solutions address legal question answering or document analysis as isolated functionalities, without offering an integrated workflow that supports users from initial legal awareness to actionable resolution.

To address these limitations, this paper proposes Lawvia AI, a unified AI-driven legal assistance platform that combines conversational legal guidance, document risk analysis, and structured complaint handling within a single system. Lawvia AI enables users to describe legal issues using text or voice input, receive simplified explanations of applicable laws,

analyze uploaded legal documents to identify potential risks, and generate structured complaint letters when necessary. In addition, the platform introduces an institutional complaint routing mechanism that allows grievances to be forwarded to appropriate authorities based on severity, supporting organized escalation within institutions.

By leveraging BERT-based transformer models, NLP-driven document processing, and a modular system architecture, Lawvia AI aims to enhance legal literacy, reduce dependency on costly legal consultations, and promote informed decision- making. The proposed system is designed to be user-friendly, scalable, and socially relevant, making it suitable for deployment in educational institutions, workplaces, and public legal awareness initiatives.

II. LITERATURE REVIEW

The growing complexity of legal systems and limited legal literacy among the general public have encouraged the development of AI-based legal assistance solutions. Legal documents are often written in complex language, while access to professional legal services is constrained by cost and availability. As a result, AI-driven legal information systems have emerged as effective tools for improving legal accessibility.

Sharma et al. proposed LAW-U, an AI-based legal guidance chatbot aimed at supporting sexual violence victims through conversational legal assistance [1]. Although effective within its domain, the system is limited in scope and does not support document analysis, complaint drafting, or escalation mechanisms. Rao et al. introduced AI Jurist, a conversational platform designed to improve legal literacy in India [2]. While the system enhances public legal awareness, it primarily functions as a question–answer tool and lacks integrated document analysis or grievance-handling features.

Several studies have explored chatbot-based legal assistance using NLP and machine learning techniques. Kumar et al. developed a rule-based and intent-classification chatbot for legal queries [3]. However, such approaches rely on predefined responses and exhibit limited adaptability. In parallel, research on legal document processing has focused on clause extraction and classification. Nair et al. proposed a machine learning framework using named entity recognition to analyze legal documents [4], but the system does not provide simplified explanations or integrate legal advisory workflows.

Transformer-based models have recently gained prominence in legal text analysis. Singh et al. demonstrated that BERT-based models outperform traditional machine learning methods for legal clause classification [5]. Similarly, Chalkidis et al. evaluated Legal-BERT for legal document classification, showing improved contextual understanding [6]. However, these models are commonly deployed as standalone classifiers without user-oriented legal guidance or complaint-handling capabilities.

Other works have addressed document risk analysis using NLP-based keyword extraction and clause-level classification. Kumar et al. proposed a system to detect risky contractual clauses [7], while Patil et al. applied supervised learning for complaint severity classification to assist administrative prioritization [8]. Although effective in isolation, these systems do not integrate legal explanations, document analysis, and escalation workflows.

Gupta et al. proposed a hybrid legal assistance framework combining chatbots with document processing modules [9]. While this approach moves toward integrated legal support, it lacks structured authority routing and escalation mechanisms. Overall, existing literature highlights the absence of a unified platform that combines legal awareness, document risk analysis, complaint drafting, and institutional routing. Lawvia AI addresses this gap by integrating these components into a single, cohesive legal assistance system.

III. METHODOLOGY

The methodology of Lawvia AI follows an integrated pipeline that transforms unstructured legal inputs into simplified guidance, actionable documents, and structured grievance handling. The system combines Natural Language Processing (NLP), Machine Learning (ML), and deep learning techniques to support legal awareness, document risk analysis, and institutional complaint routing.

User interaction begins through text or voice input, with voice queries converted into text using speech-to-text APIs. The input text undergoes NLP preprocessing, including normalization and tokenization, before analysis. A BERT-based transformer model within the chatbot generates contextual embeddings to identify the legal domain, such as cybercrime, harassment, or tenancy disputes. Based on the detected domain, simplified legal explanations are generated using controlled language generation and LLM APIs.

For document analysis, uploaded legal documents are processed using OCR and PDF parsing tools to extract text. Risky clauses are identified through keyword-based mapping using a predefined legal risk lexicon, and key sentences are summarized using a frequency-based approach to produce concise risk summaries.

Institutional grievance handling is supported through a severity-based routing mechanism. Complaints are analyzed using BERT embeddings combined with a Random Forest classifier trained on labeled complaint data. The predicted severity is mapped to predefined authority hierarchies, enabling automated routing and escalation.

All components are integrated using a modular architecture with ASP.NET and C# for the application layer, Python-based services for NLP and ML processing, and SQL Server for data management.

A. User Interaction and Input Processing

Lawvia AI supports both text and voice-based user interaction. Voice inputs are converted into text, followed by NLP

preprocessing to standardize input and reduce noise. This ensures reliable downstream processing for legal understanding and complaint handling.

B. Legal Awareness Chatbot

The chatbot employs a BERT-based transformer model to generate contextual embeddings and classify user queries into legal domains. Simplified legal explanations are generated
using NLP techniques and LLM APIs, enabling accurate and user-friendly guidance.

C. Document Risk Detection and Summarization

Legal documents are analyzed using OCR and NLP preprocessing. Risky clauses are detected through keyword mapping against a legal risk dictionary. Important clauses are summarized using frequency-based scoring to highlight key risks.

D. ESpace: Institutional Complaint RoutinG

Complaints are analyzed using BERT embeddings and a Random Forest classifier to predict severity levels. Based on severity, complaints are routed to appropriate authorities. Unresolved complaints can be escalated automatically to higher authorities.

E. Complaint Draft and Action (DocuAction)

Complaint letters are generated using predefined templates corresponding to the identified legal domain. User details are dynamically populated, and the final document is produced as a downloadable PDF.

F. Legal Support Resource Mapping (LegalAid+)

LegalAid+ provides access to nearby legal resources such as police stations and legal aid centers using stored regional data and optional location services. This module bridges the gap between legal awareness and real-world assistance by guiding users toward appropriate support channels.

G. Dataset Preparation

A locally curated dataset was created to support complaint classification and severity analysis. The dataset consists of approximately 1000 complaint records, each labeled with a legal category (e.g., harassment, cybercrime, tenant dispute). Text preprocessing steps such as tokenization, stop-word removal, and normalization were applied. This dataset is used for fine-tuning transformer-based models.

H. System Integration and Data Management

Lawvia AI follows a modular architecture integrating multiple components into a unified system. The frontend interface is developed using ASP.NET with C#, while NLP and ML operations are handled by Python-based services. Structured data such as user profiles, complaints, and document metadata are stored in SQL Server. Inter-module communication is managed through internal APIs, ensuring smooth data flow between the chatbot, document analysis, complaint management, and authority routing modules.SYSTEM ARCHITECTURE

A. Architectural Overview

The overall architecture of Lawvia AI follows a layered and modular design to ensure scalability, maintainability, and clear separation of concerns. The system consists of three primary layers: the User-Facing Layer, AI/Processing Layer, and Data Management Layer, as illustrated in Fig. 1.
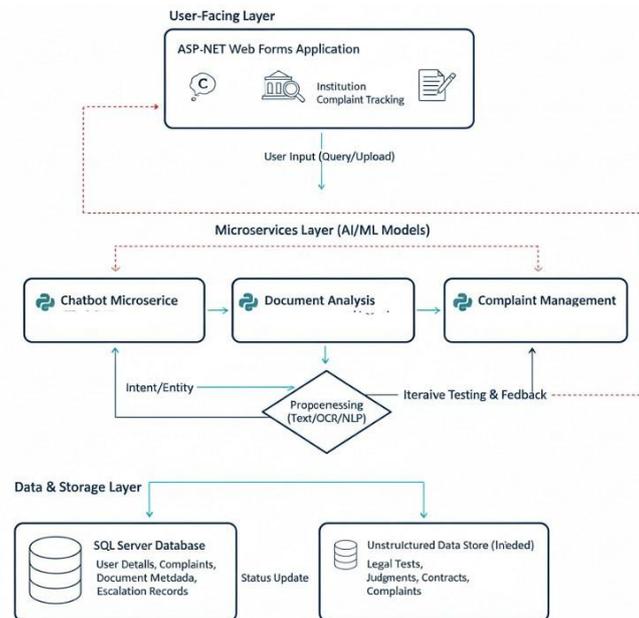
Fig. 1. Architecture diagram which demonstrates the Overall working of Lawvia AI.

B.   User-Facing Layer

The User-Facing Layer is implemented using ASP.NET Web Forms, providing interfaces for:
- Legal awareness chatbot interaction.
- Complaint drafting and institutional complaint tracking (Espace).
- Legal support resource access (LegalAid+).
- Legal document upload for risk analysis.

All user inputs—including text queries, voice inputs, and document uploads—are forwarded securely to the processing layer for analysis.

C.   AI and Processing Layer

The Processing Layer contains modular AI components responsible for legal intelligence tasks:
1.   Chatbot Microservice
- Uses a BERT-based NLP model for intent detection and legal domain understanding.

- Generates simplified legal explanations and guidance.
2.   Document Analysis Module
- Extracts text using OCR (Tesseract / PyMuPDF).
- Applies NLP-based keyword mapping for risk detection.
- Performs frequency-based summarization.
3.   Complaint Management Module (ESpace)
- Classifies complaint severity using ML models.
- Assigns complaints to appropriate authorities.
- Supports escalation   to   higher   authorities   if unresolved .

Preprocessing operations such as text cleaning, tokenization, and OCR normalization are shared across modules.

D.   Data and Storage Layer

The Data Layer manages both structured and unstructured data:
- SQL Server Database

Stores user profiles, complaints, escalation records, and document metadata.
- Unstructured Data Store

Maintains legal texts, uploaded contracts, and complaint documents for NLP processing.

IV. SYSTEM FLOWCHART AND OPERATIONAL FLOW

The operational flow of the Lawvia AI system follows a structured sequence to ensure accurate legal assistance, document analysis, and complaint handling.

1. The process begins with user login using valid credentials.
2. The user selects one of the system modules:
- Legal Support (LegalAid+).
- Complaint Draft & Action (DocuAction).
- Institutional Complaint Routing (ESpace).
- Document Risk Detection.
3. Input data undergoes preprocessing, including:
- Text cleaning.
- Speech-to-text conversion.
- OCR for document uploads.
4. The processed input is analyzed based on the selected module:
- Chatbot response generation.
- PDF complaint letter creation.
- Severity-based authority assignment.
- Risk detection and summarization.
5. The system produces final outputs such as:
- Legal explanations.
- Complaint documents.
- Authority routing decisions.Document risk summaries.

This modular operational flow enables seamless navigation from legal awareness to actionable outcomes, ensuring usability, scalability, and effective grievance handling.
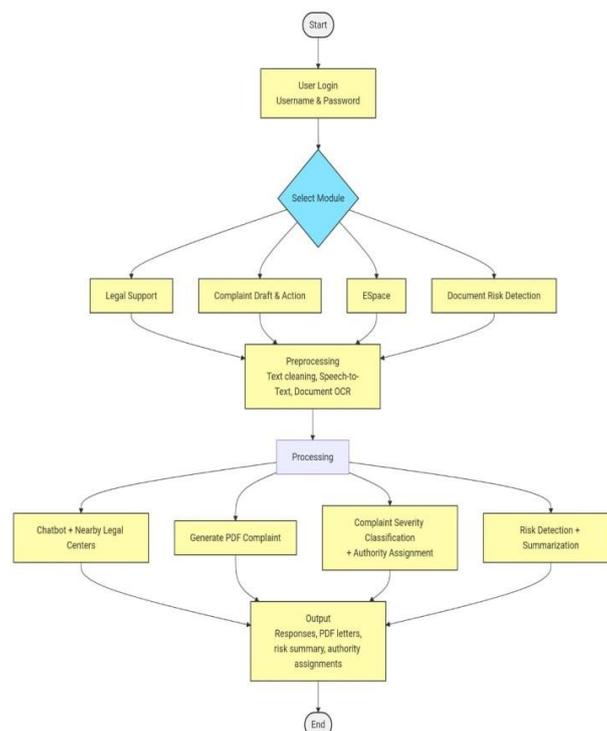

Fig. 2. Lawvia AI system flowchart diagram

V.   MODEL EVALUATION

The performance of the proposed Lawvia AI system was evaluated to assess the effectiveness of its machine learning models used for legal domain classification and complaint severity assessment. Evaluation was conducted using a supervised dataset of approximately 1000 labeled complaint records, prepared specifically for this study.

A.   Evaluation Setup

The dataset consisted of user complaint texts categorized into multiple legal domains (e.g., workplace harassment, cybercrime, rental disputes) and assigned severity levels (low, medium, high). A BERT-based transformer model was fine- tuned using this dataset to perform intent understanding and severity prediction. Model training and evaluation were managed through the Admin Dashboard, which allows controlled dataset upload, training execution, and result visualization. The dataset was divided into training and testing subsets using an 80:20 split, ensuring unbiased evaluation. Standard preprocessing steps such as text normalization and tokenization were applied prior to training.

B.   Evaluation Metrics errorsoccurred between semantically similar categories, which is expected in real-world legal text classification scenarios.
Severity classification performance showed strong alignment with ground truth labels, enabling reliable authority routing within the ESpace module. This supports automated escalation decisions while maintaining transparency through admin-level monitoring.

Model performance was measured using commonly accepted classification metrics:
- Accuracy: Overall correctness of predictions
- Precision: Correctness of positive predictions
- Recall: Ability to identify relevant cases
- F1-Score: Balance between precision and recall Additionally, a confusion matrix was used to visually analyze prediction outcomes across different legal categories and severity levels.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

C.   Evaluation Process

- The dataset contains approximately 1000 labeled complaint records across multiple legal categories.
- The dataset is split into training and testing sets.
- Predictions from the fine-tuned BERT model are compared with ground-truth labels.
- A confusion matrix is generated, and metrics are computed.

*D.   Confusion Matrix Analysis*

The confusion matrix highlights the relationship between predicted and actual labels. Results indicate that the BERT-based model achieves high accuracy in distinguishing between legal   domains   with   minimal   misclassification. Most RESULTS AND DISCUSSION

This section presents the observed outcomes and functional results obtained from the implementation of the Lawvia AI system. The evaluation focuses on usability, module effectiveness, and system responsiveness across different components.

A.   User Authentication and Profile Management

The system successfully implements a secure login mechanism using username and password authentication. Upon login, users are directed to a personalized dashboard where profile details and previous interactions are maintained.
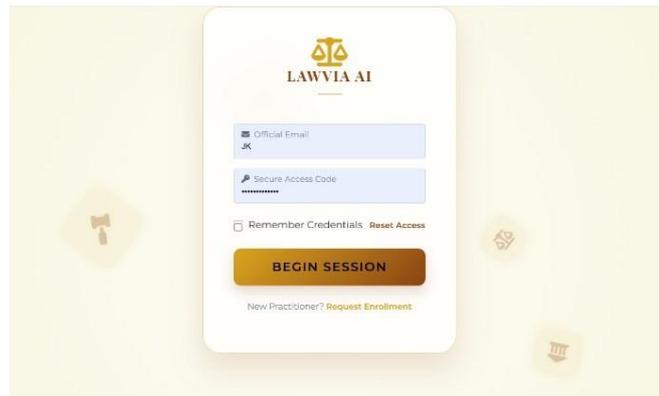
This ensures controlled access and user-specific data persistence, which is essential for complaint tracking and document history management.

Fig. 3. Lawvia AI User authentication

B.   Chatbot Module Output

The chatbot demonstrates effective legal domain understanding by accurately interpreting user queries related to common legal issues such as cybercrime, workplace harassment, and rental disputes. Using a BERT-based NLP model, the system classifies the intent of the query and provides simplified legal explanations and guidance. The chatbot output is concise, context-aware, and user-friendly, enabling non-technical users to understand legal concepts easily.
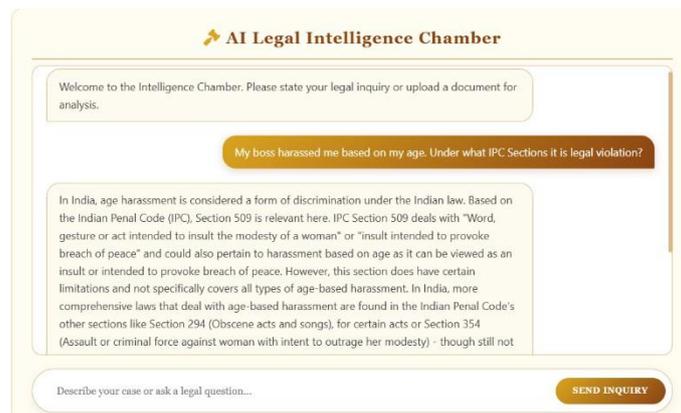

Fig. 4. Legal chatbot response

C.   Document Risk Analysis and Summarization

The document risk analysis module successfully processes uploaded legal documents such as contracts and offer letters. Text extraction using OCR techniques enables the system to analyze document content and identify risky clauses, including hidden charges and restrictive conditions. Detected risks are highlighted and summarized, and the final output can be downloadable as a PDF.
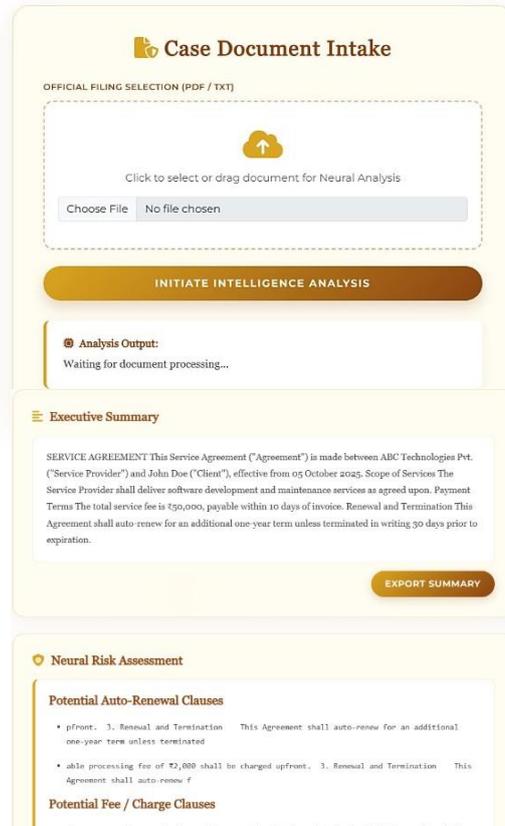
Fig. 5. Document risk analyzer module response

Admin Dashboard and System Monitoring

The admin dashboard provides comprehensive control over system operations. Administrators can monitor user growth, manage registered users, and oversee complaint statistics. The dashboard also supports model training and evaluation workflows, including dataset upload and performance assessment. A notification system is integrated to allow administrators to send alerts or updates to all users or selected users, enhancing communication and system transparency.

Fig. 6. Admin dashboard which shows overall user status.

Fig. 7. Users manage option for admin
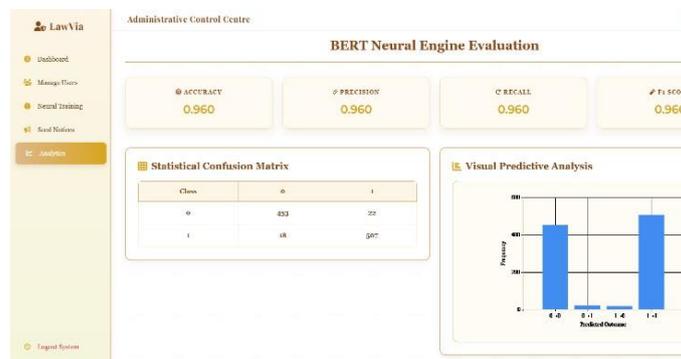
Fig. 8. Model training and evaluation by admin

D.   Authority Routing

The Authority Routing (ESpace) module automatically analyzes user complaints using machine learning to determine their severity level. Based on this severity, complaints are routed to the appropriate authority within the institution. If a complaint remains unresolved, the system supports automatic escalation to higher authorities, ensuring transparent and efficient grievance handling.

a.   Send Compliant

It allows users to easily create and submit formal complaint letters by entering basic details about the issue. The system automatically generates a legally structured complaint document and enables users to download or forward it to the concerned authority.
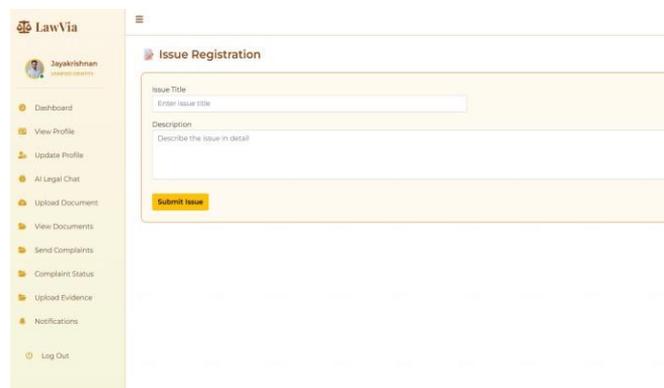

Fig. 9. Complaint Registration

b.   Upload Evidence

The Upload Evidence feature allows users to attach supporting files such as documents, images, or screenshots along with their complaints.
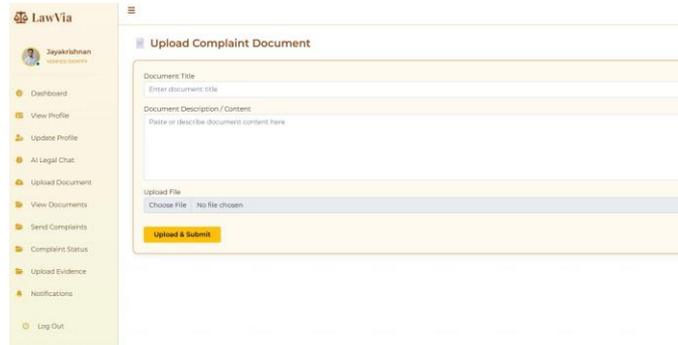

Fig. 10. Upload Complaint Document

c.   Authority Complaint Management

Authority Complaint Management enables authorized officials to view, track, and respond to complaints assigned to them based on severity.
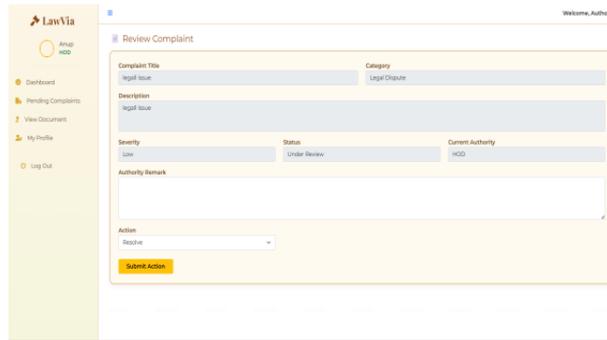

Fig. 11. Complaint Review

d.   Complaint Status

Complaint status allows users to track the current state.

Fig. 12. Status of registered complaints

VI.   CONCLUSION

This Lawvia AI, an intelligent legal assistance platform designed to enhance legal awareness and simplify access to legal guidance for the general public. The proposed system integrates NLP and ML techniques to provide users with legal domain understanding through a conversational chatbot, automated complaint drafting, document risk analysis, and structured authority routing via the ESpace module. By combining a BERT-based chatbot for legal intent

understanding, keyword-driven document risk detection, and ML-based severity assessment for complaint routing, Lawvia AI offers a unified solution that bridges the gap between legal awareness and actionable outcomes.

Experimental observations demonstrate that the system effectively classifies legal issues, identifies risky clauses in documents, and generates simplified summaries that improve user comprehension. The modular architecture enables scalability, institutional adaptability, and secure data management, while the admin dashboard supports model evaluation, user monitoring, and notification management. Compared to existing legal chatbot or document analysis systems, Lawvia AI provides a more comprehensive and socially relevant framework by integrating awareness, documentation, and escalation mechanisms within a single platform .Overall, Lawvia AI highlights the potential of AI-driven legal assistance systems in promoting legal literacy, supporting informed decision-making, and enabling structured grievance redressal. The proposed approach can serve as a foundation for future intelligent legal systems aimed at improving access to justice through technology.

VII.   FUTURE SCOPE

Lawvia AI can be further enhanced in several directions to increase its effectiveness, scalability, and real-world adoption. Future work may focus on integrating official government legal databases and portals to provide more up-to-date and jurisdiction-specific legal information. Expanding multilingual support to include additional regional languages can further improve accessibility for diverse user groups.

The document risk analysis module can be strengthened by incorporating advanced transformer-based legal language models to improve clause classification accuracy and contextual understanding. Similarly, the authority routing mechanism in the ESpace module can be extended using adaptive learning techniques to refine severity prediction based on historical complaint resolution patterns.

Future versions of Lawvia AI may also include a mobile application to enhance usability and outreach. Integration with institutional grievance redressal systems and government complaint portals could enable seamless complaint forwarding and status synchronization. Additionally, incorporating explainable AI (XAI) techniques would improve transparency by allowing users and administrators to understand how legal classifications and severity decisions are made.

REFERENCES

[1]. S. Sharma, R. Verma, and P. Gupta, "LAW-U: Legal Guidance Through Artificial Intelligence Chatbot for Sexual Violence Victims and Survivors," IEEE Access, vol. 11, pp. 45621–45632, 2023.

[2]. A.Joseph, R. Thomas, and K. Nair, "AI Jurist: Empowering Legal Knowledge and Practices in India," in Proc. IET Conf. on Artificial Intelligence and Law, 2022, pp. 1–6.

[3]. P. K. Singh and M. R. Tripathi, "Leveraging AI-Driven Chatbots for Legal Literacy," in Proc. IEEE Int. Conf. Advanced Computing and Communication Technologies, 2024, pp. 215–220.

[4]. R. Nair, S. Menon, and A. Krishnan, "Design and Implementation of a Chatbot for Automated Legal Assistance Using NLP and Machine Learning," in Proc. IEEE Int. Conf. Computing, Communication and Automation, 2021, pp. 412–417.[5]. J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre- training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019, pp. 4171– 4186.

[6]. A.Chalkidis, I. Androutsopoulos, and N. Aletras, "Legal- BERT: The Muppets Straight Out of Law School," in Proc. ACL Workshop on NLP for Legal Texts, 2020, pp. 1–10.

[7]. A.Kumar and S. Agarwal, "Legal Document Risk Analysis Using Natural Language Processing," IEEE Access, vol. 10,pp. 98734–98745, 2022.

[8]. K. Menon and R. S. Pillai, "Automated Clause Detection and Risk Classification in Legal Contracts," IEEE Trans. Artificial Intelligence, vol. 3, no. 4, pp. 512–521, 2022.

[9]. S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. Sebastopol, CA, USA: O'Reilly Media, 2009.

[10]. Y. Liu et al., "TextRank: Bringing Order into Texts," IEEE Intelligent Systems, vol. 19, no. 2, pp. 65–70, 2004.