**Clause Ease AI Based Contract Language Simplifier**

**Project Documentation**

**Submitted by:**

**Group 5**

Y.Bhavya Sri

B.Yasokeerthi

M.subrajit

**Submitted to**

**Mentor:** Dr . A . Kalaivani

Infosys SpringBoard

Infosys SpringBoard Virtual Internship 6.0

**Table of content**

|  |  |
| --- | --- |
| **SNO** | **Contents** |
| 1. | Title |
| 2. | Project Statement |
| 3. | Outcomes |
| 4. | Modules to be Implemented |
| 5. | Evaluation Criteria |
| 6. | Workflow Diagram |
| 7. | Architecture Diagram |
| 8. | Database Schema |
| 9. | Future Enhancement |
| 10. | Conclusion |

**Clause Ease AI Based Contract Language Simplifier**

1. **Title:** Contract Language Simplifier



1. **Project Statement:**

Legal and policy documents, including contracts, terms and conditions, and privacy policies, are essential for defining rights, responsibilities, and obligations in personal and business contexts. However, these documents are often written in complex, formal language, laden with technical jargon and lengthy, convoluted sentences. This complexity makes it challenging for the average reader to fully comprehend their content, creating confusion, misinterpretation, and sometimes legal or financial risks. Small businesses, individuals, and non-native English speakers are particularly affected, as they may struggle to grasp critical clauses and implications. The "Contract Language Simplifier" project seeks to address this problem by leveraging an AI-powered text simplification model that translates complex legal English into clear, plain language. By making legal information accessible and understandable, the tool aims to empower users to make informed decisions, reduce misunderstandings, and promote transparency in contractual and policy communications.

1. **Outcomes:**

* **Clear Understanding:** Users can effortlessly grasp complex legal and policy documents without needing specialized legal knowledge.
* **Time-Saving Automation:** The tool reduces the manual effort involved in simplifying lengthy and dense contracts.
* **Broader Accessibility:** Makes critical legal information understandable and accessible to everyone, including non-experts and small business owners.
* **Customizable Simplification:** Offers multiple levels of simplification so users can choose how much detail or simplicity they want.
* **Readability Insights:** Provides metrics to compare the readability of the original versus the simplified text, helping users see improvements.
* **Secure Management:** Includes user authentication and an admin panel for managing content and monitoring system activity, ensuring safety and controlled access.

1. **Modules to be Implemented:**

**1. User Authentication & Profile Management:**

* Registration and login using standard email/password.
* Personal user profiles to manage uploaded documents and save simplified versions.

**2. Document Ingestion & Preprocessing:**

* Interface for users to paste text or upload legal/policy documents.
* Automated text cleaning, sentence segmentation, and tokenization for smooth processing.

**3. NLP Analysis & Simplification (Core Module):**

* **Readability Analysis:** Compute readability scores like Flesch-Kincaid and Gunning Fog for the input text.
* **Text Simplification Model:** Utilize a text-to-text AI model (e.g., fine-tuned FLAN-T5 or BART) to convert complex sentences into simpler language.
* **Summarization (Optional):** Condense longer passages while simplifying content.
* **Key Term Extraction:** Identify complex legal terms and provide explanations where necessary.

**4. Output & Visualization:**

* Display original and simplified texts side-by-side for easy comparison.
* Show readability scores for both versions.
* Allow users to select different levels of simplification based on their needs.

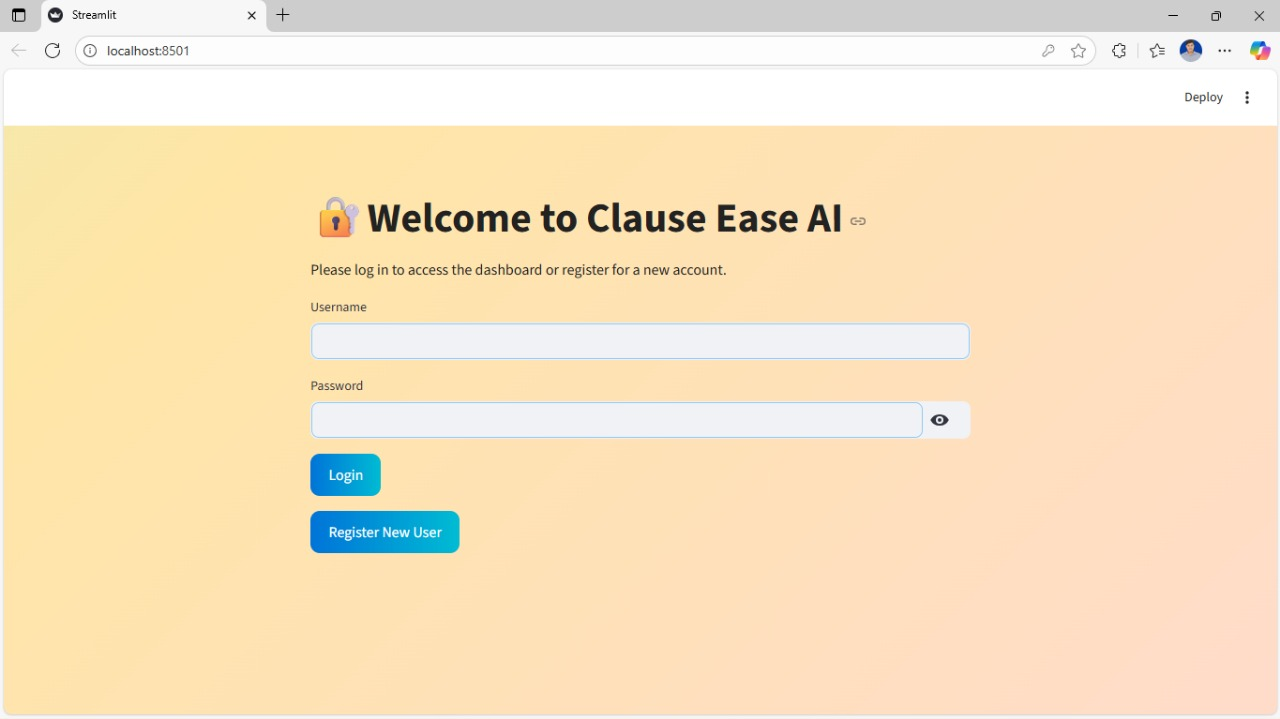
**5. Admin Dashboard:**

* Manage simplification model parameters.
* Monitor system usage and performance.
* Review and ensure the quality of simplifications.

**Module 1: User Authentication & Document Ingestion**

**High-Level Requirements:**

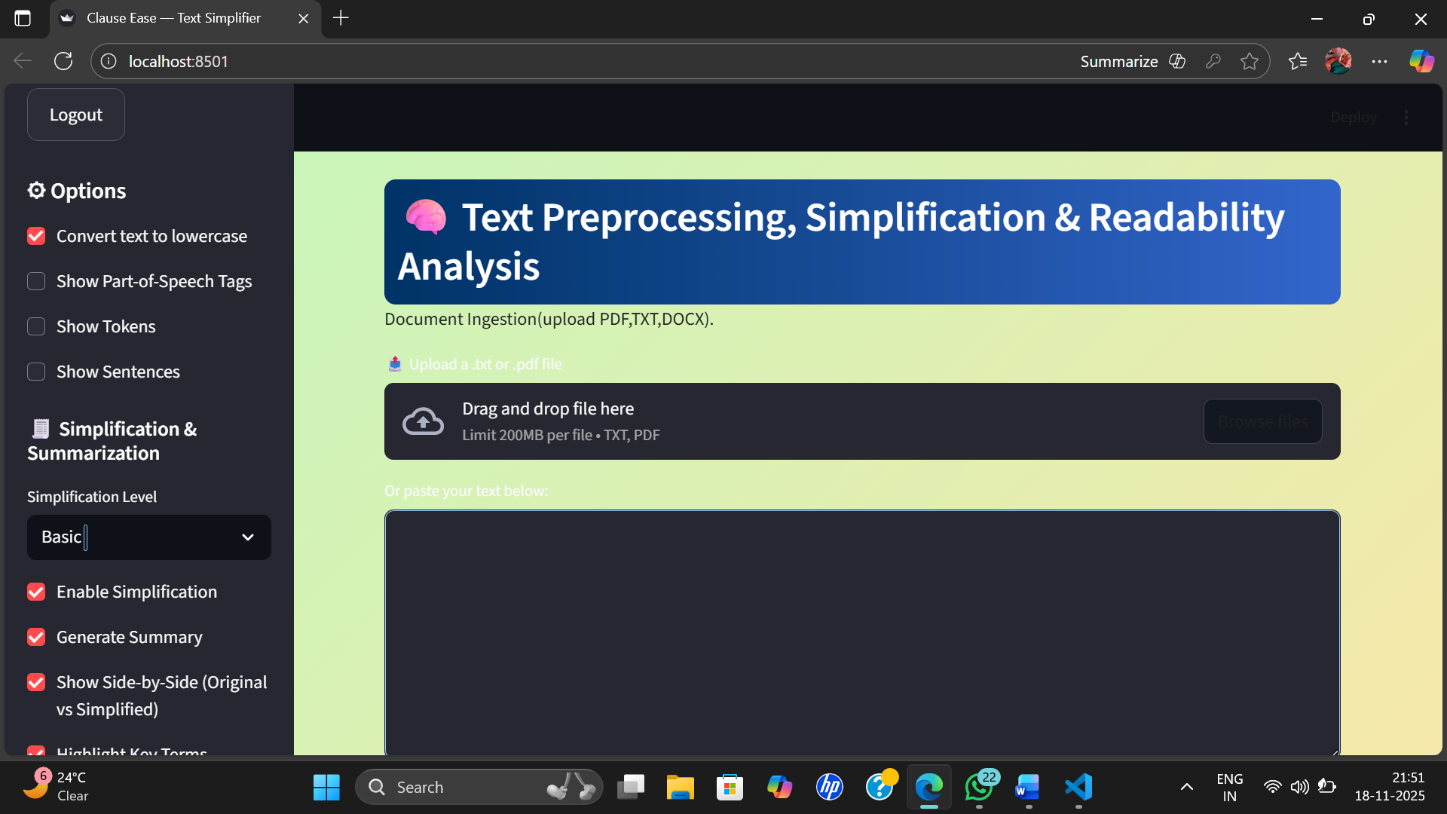
* **User Registration:** Implement a secure registration system where users can sign up using email and password, with JWT (JSON Web Token) based authentication to ensure secure sessions.
* **Login System:** Develop a reliable login mechanism that authenticates users and protects their data.
* **Profile Management:** Provide personal user profiles where users can view, manage, and organize their uploaded documents and previously saved simplified versions.
* **Document Input Interface:** Create a user-friendly web interface (using Flask or Streamlit) that allows users to either paste text or upload legal and policy documents in plain text format for processing.



**Module 2: Text Preprocessing & Readability Analysis**

**High-Level Requirements:**

* **Text Preprocessing Pipeline:** Build a robust preprocessing system that performs text cleaning, sentence segmentation, and tokenization using NLP libraries like NLTK and spaCy.
* **Readability Score Calculation:** Integrate readability metrics (e.g., Flesch-Kincaid Grade Level, Gunning Fog Index) using NLTK or custom implementations to evaluate the complexity of the input text.
* **Initial UI for Input & Readability:** Enhance the web interface to allow users to submit text and immediately view readability scores, giving instant feedback on the complexity of their document.



**Module 3: Core Simplification & Summarization**

**Objective:**

The main goal of this milestone is to make the system capable of **simplifying** and **summarizing** complex text using a pre-trained transformer model.

Users can upload or paste text, and the app will automatically:

1. Rewrite the text into a **simpler, human-friendly form**, and
2. Generate a **short, clear summary** of the same text.

Both outputs are displayed **side-by-side** in the user interface for easy comparison.

**High-Level Requirements:**

* **Text Simplification Model Integration:**
  + A **Hugging Face Transformer model** (such as **FLAN-T5-base** or **BART-large-cnn**) is integrated to automatically simplify complex sentences.
  + The model processes the uploaded or typed text and produces a simpler version with clear, natural language while keeping the original meaning intact.
* **Basic Summarization:**
* A built-in summarization feature generates a short, meaningful summary of the document.
* This helps users quickly understand the main points without reading the entire text.
* Side-by-Side Display:
  + The user interface displays the **original text** and its **simplified version** next to each other.
  + This allows easy comparison of sentence structure, vocabulary, and clarity.

**User Interface Description:**

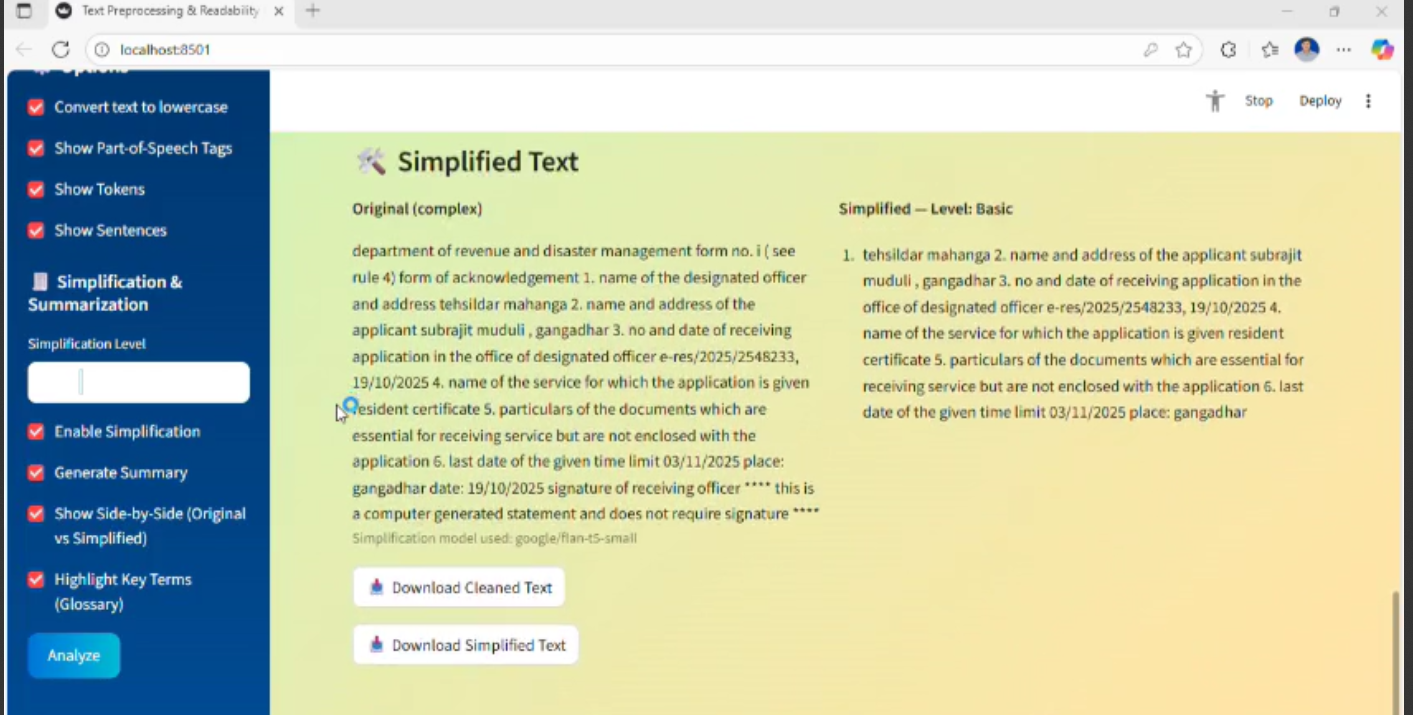
Users can **upload .txt or .pdf files** or directly **paste text** into the input area.

Options on the left panel include:

* Convert text to lowercase
* Show Part-of-Speech tags
* Show Tokens and Sentences
* Enable Simplification
* Generate Summary

A **Simplification Level** slider allows users to control the degree of simplification.

Once text is processed, the **simplified and summarized outputs** appear side by side below the input section.



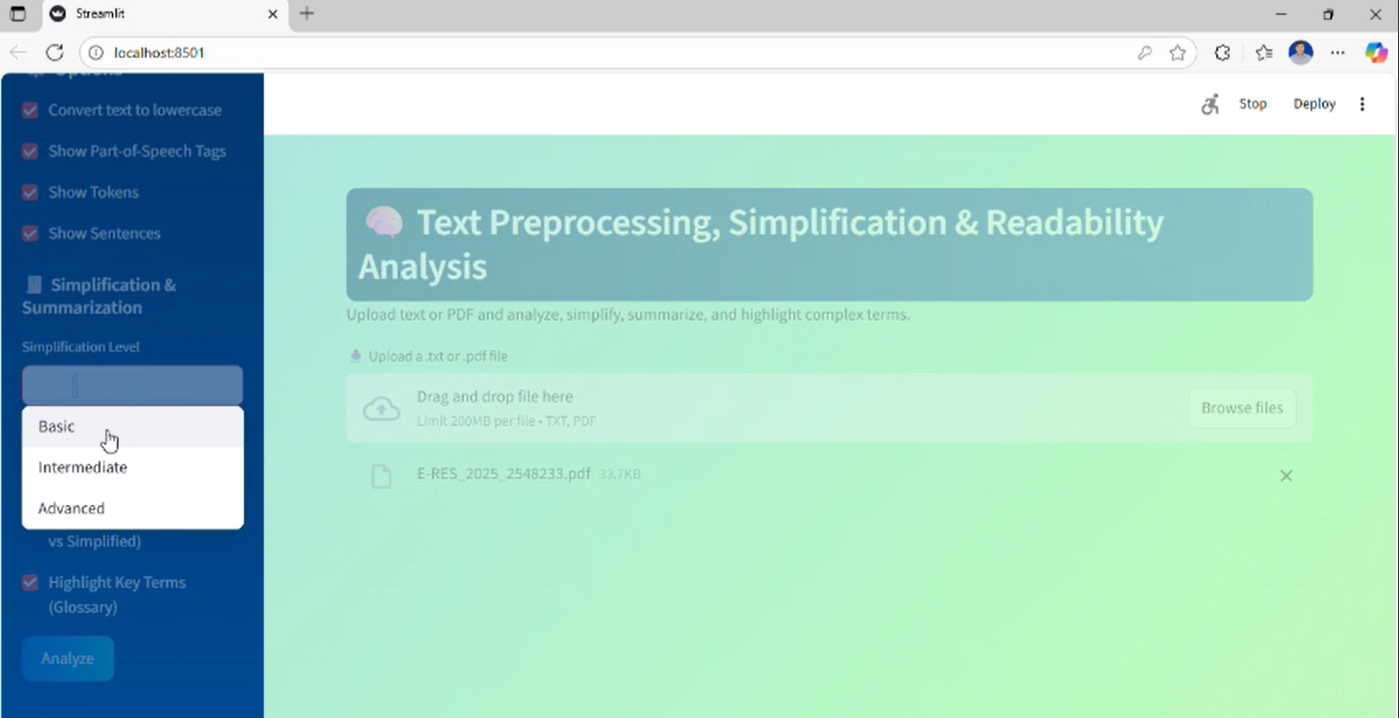
**Module 4: Multi-Level Simplification, Admin & Deployment**

**Objective:**

Enhance the text simplification system with multiple simplification levels, admin management, and deployment readiness for real-world use.

**Key Features:**

* **Multi-Level Simplification:**
  + **Basic:** Minimal rephrasing, sentence shortening.
  + **Intermediate:** Moderate simplification with easier vocabulary.
  + **Advanced:** Deep rewording and phrase restructuring.
  + Levels use either fine-tuned models (FLAN-T5, BART) or post-processing rules.



* **Key Term Highlighting:**
  + Complex terms are highlighted with tooltips or glossary definitions for easy understanding.
* **Admin Dashboard:**
  + Monitor simplification requests and response times.
  + Review and correct outputs.
  + Manage glossary entries.
  + Track metrics like total simplifications and top terms.



* **Deployment:**
  + Containerized using Docker for portability.
  + Ready for hosting on Streamlit Cloud, Hugging Face Spaces, or platforms like Render/Railway.
* **Documentation & Presentation:**
  + Complete project docs, code comments, and presentation slides including workflow, UI screenshots, and results.

**Functional Flow:**

1. User selects simplification level.
2. Text is processed and complex terms highlighted.
3. Model generates simplified text.
4. Admin reviews outputs and updates glossary.
5. Deployed via Docker on cloud platforms.

**Outcome:**  
A fully deployable, user-friendly system with scalable multi-level simplification, term highlighting, and admin control — suitable for demonstration or real-world applications

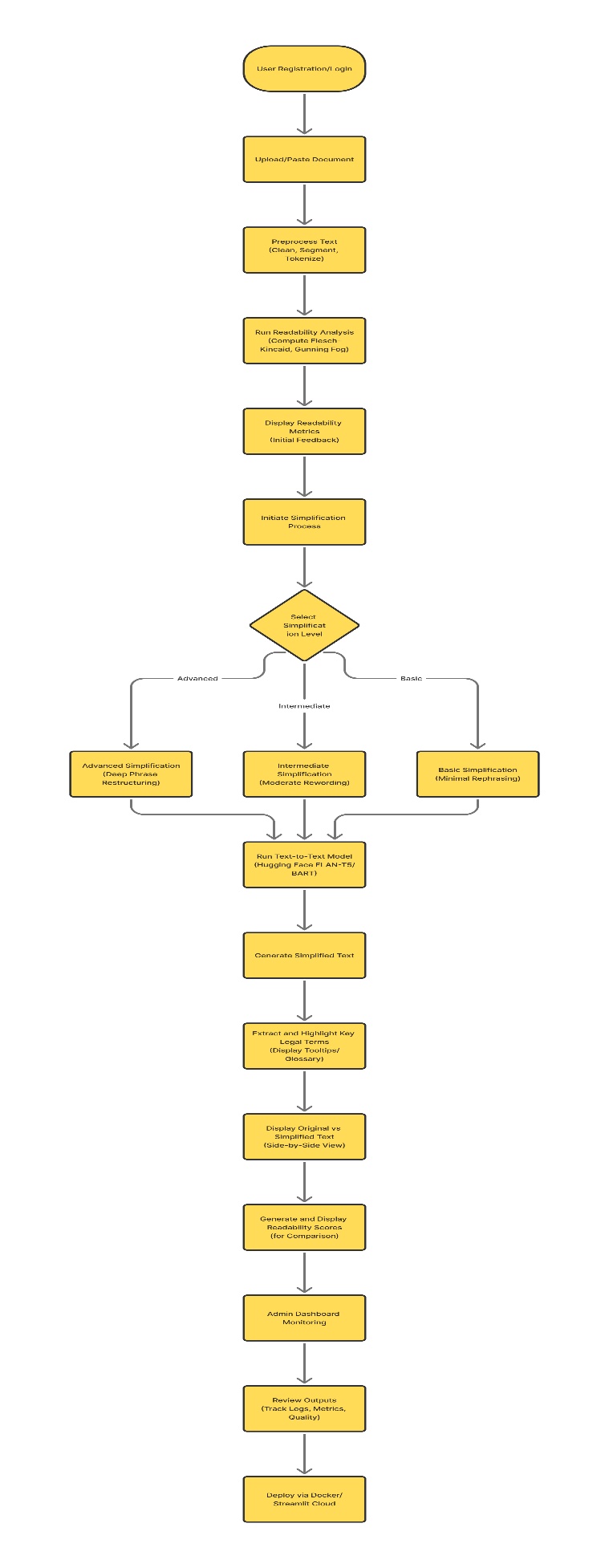
**5.Evaluation Criteria:**

* **Module 1:**
  + User registration/login works.
  + Text input interface supports pasting/uploading.
* **Module 2 :**
  + Text preprocessing handles multiple formats.
  + Readability scores calculated accurately.
  + UI displays input text with readability metrics.
* **Module 3 :**
  + Core simplification model integrated.
  + Basic summarization works.
  + Original and simplified text displayed side-by-side.
* **Module 4 :**
  + Multi-level simplification implemented.
  + Key term highlighting/explanation functional.
  + Admin dashboard operational.
  + App containerized for deployment.
  + Complete documentation and final presentation delivered.

**6. Workflow Diagram**

This diagram illustrates the user's journey through the Contract Language Simplifier

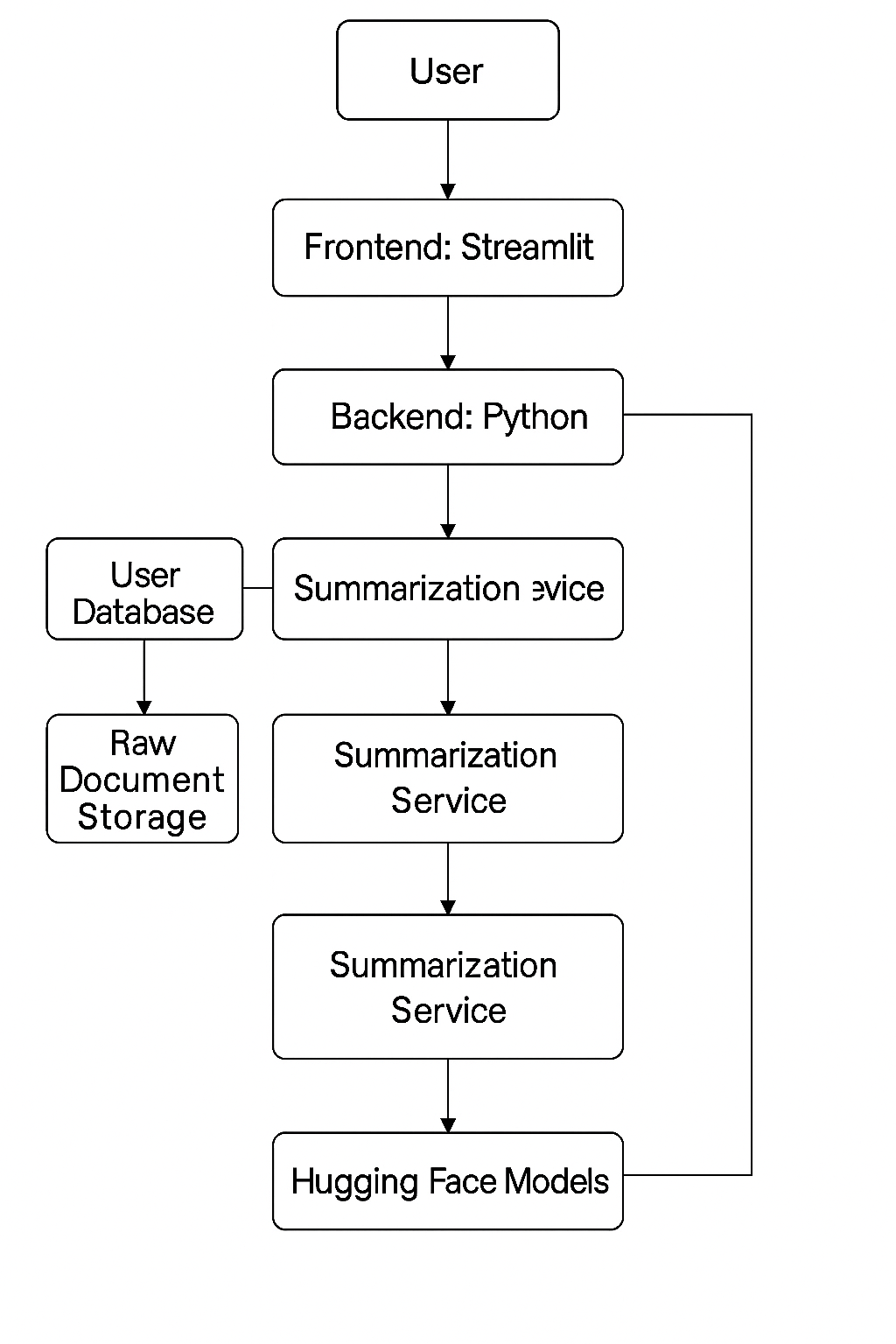
application and the system's underlying process flow.



**7. Architecture Diagram**

This diagram outlines the major components of the Contract Language Simplifier system and

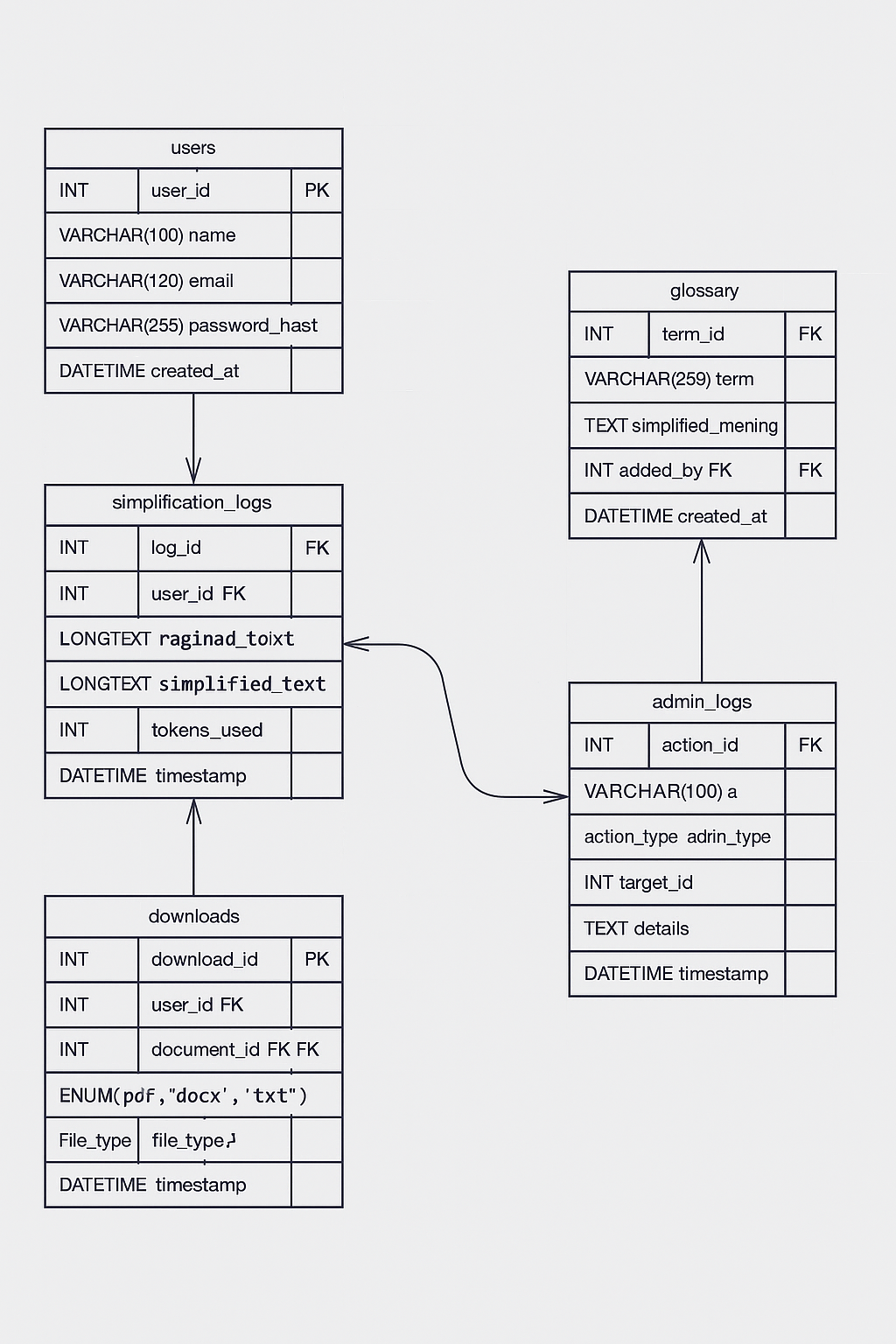
how they interact.



**8. Database Schema**

This schema details the tables and their relationships within the Contract Language

Simplifier's database.



**9. Future Enhancement**

**Multilingual Support**

* Extend the model to simplify contracts written in multiple languages (e.g., French, Spanish, Hindi).
* Include translation + simplification features for global use.

**User Feedback Learning System**

* Allow users (lawyers, students, or clients) to rate simplified outputs.
* The model can learn from feedback to improve future simplifications.

**Voice-Based Contract Explanation**

* Integrate text-to-speech to **verbally explain** simplified versions of complex legal terms.

**Offline Access or Mobile App**

* Develop a mobile or desktop version that can work without internet connectivity for secure environments.

**Data Privacy and Security Enhancements**

* Use encryption and secure storage for uploaded contracts.
* Add compliance with GDPR and other privacy regulations.

**10. Conclusion**

The project **ClauseEase** shows how Artificial Intelligence can make legal language easier for everyone to understand. Many people find contracts confusing because of the complex and technical terms used in them. ClauseEase helps solve this problem by using **Natural Language Processing (NLP)** to read and simplify the clauses while keeping their original meaning.

This tool helps users save time, reduce confusion, and understand legal documents without always needing expert help. It can be especially useful for students, businesses, and individuals who deal with agreements and want to know what they are signing.

In the future, ClauseEase can be improved by adding more features like support for multiple languages, automatic detection of important clauses, and checking if the content follows legal rules. These updates would make it even more useful and reliable. Overall, this project proves that AI can play a big role in making legal information simpler, clearer, and more accessible to everyone.