Smart Legal Form Builder

- AI-Enhanced Document Creation

Seoyeon Kim
College of Social Science Hanyang University
Dept.of Tourism
Seoul, Korea
soen0814@hanyang.ac.kr

Taeyeon Kim
College of Engineering Hanyang University
Dept.of Information Systems
Seoul, Korea
tangboll5075@gmail.com

Suhyeon Yu
College of Engineering Hanyang University
Dept.of Information Systems
Seoul, Korea
sh265826@naver.com

Soohan Jeong
College of Engineering Hanyang University
Dept.of Information Systems
Seoul, Korea
soohan9909@gmail.com

Jiwon Han
College of Business Hanyang University
Dept.of Business
Seoul, Korea
jiwon8623@hanyang.ac.kr

***Abstract*— Smart Legal Form Builder is an app that automates the creation of important legal documents, making them more accessible to the general public. Legal systems around the world are often inaccessible to the general public due to their complex terminology, structured document formats, and the need for specialized knowledge. This inaccessibility, coupled with the significant financial and time costs associated with seeking legal assistance, prevents many individuals from effectively pursuing their rights and struggles with drafting complaints. There is a growing need for an easy and efficient app for drafting legal documents.**

The core of Smart Legal Form Builder is that it divides complaints into several categories— Secondhand transaction fraud, online abuse, sexual harassment, assault and injury—and uses AI to automatically draft each complaint through questions. The app adapts the questions based on the user's input, providing a personalized document drafting experience tailored to the specific legal context and scenario. This process eliminates the need for the user to interpret complex legal terminology or worry about legal requirements. Once all the necessary inputs and answers are collected, the application provides a complaint in legal form that is ready to be used in legal proceedings. To enhance its performance, the platform utilizes a fine-tuned GPT model trained on categorized legal complaint data. This tailored AI ensures the generated documents adhere to legal standards and accurately represent the user's case, improving the relevance and reliability of outputs.

The functionality of the application extends beyond document creation. Features such as real-time input validation, seamless integration of user interaction and AI-based automation ensure a smooth and accessible experience. Developed using Flutter, the platform offers cross-platform compatibility, allowing Android and iOS users to access the same responsive and user-friendly interface. The generated documents are processed via Syncfusion Flutter PDF, allowing users to preview, adjust, download, and submit professional-quality PDFs directly from the app. The app also features a user-friendly main menu and completion confirmation system, designed to enhance the overall user experience and encourage ongoing engagement.

On the backend, the platform is powered by AWS EC2, providing a scalable and secure environment to manage user data and process AI-based document generation. This infrastructure ensures high availability and responsiveness even with a large number of users.

This paper explores the design and development of the platform and addresses the need for an app that can help people without prior legal knowledge and who struggle to draft legal documents. By reducing the reliance on expensive legal professionals and streamlining the document drafting process, Smart Legal Form Builder can be of use to ordinary people in an increasingly complex legal environment.

Keywords—Legal technology, AI legal assistance, automated complaint drafting, legal accessibility, contextual AI, Step-by-step legal guidance, Personalized document creation

I. INTRODUCTION

1. Motivation

For many, the law is difficult and inaccessible. While everyone has had to draft a complaint at some point in their lives, this basic task is often difficult for individuals who lack the legal expertise or resources to do so. This problem disproportionately affects marginalized communities, amplifying social inequality and widening the gap between those who can and cannot access professional legal assistance.

However, advances in AI and automation are helping to solve this problem through the development of apps. Smart Legal Form Builder leverages these technologies to streamline the legal process and help individuals draft legal documents professionally and independently. The app specializes in creating complaints for four key categories: secondhand transaction fraud, online abuse, sexual harassment, assault and injury. This not only addresses immediate barriers, but also makes filing a complaint more inclusive and accessible.

2. Problem Statement

A. Lack of Public Understanding of Legal Terms and Documents

According to the 2019 National Legal Awareness Survey, 76.3% of respondents reported having difficulty understanding legal terms, and 78.4% found legal documents difficult to understand. This lack of understanding is a major cause of individuals' difficulties in participating in the legal system. Misinterpretation of legal terms often leads to errors in document preparation, which creates further obstacles to achieving legal resolution.

B. Limited Interest and Familiarity with Legal Information

According to the 2021 National Legal Awareness Survey, 79.9% of respondents reported never having searched for legal information online. Additionally, only 23.4% discussed legal topics with acquaintances to better understand the law, indicating a lack of interest in legal information. This suggests that even though legal tools exist, most individuals do not know how to use them effectively.

C. High Costs and Barriers to Accessing Legal Professionals

Legal services are often prohibitively expensive, especially for tasks such as filing a complaint. Many individuals face time and financial burdens when seeking professional help. Low-income people cannot even afford legal consultation fees. Despite the availability of legal knowledge online, it is still common to rely on expensive legal professionals due to the complexity of the complaint.

These statistics and challenges highlight the need to streamline the legal document drafting process and provide tools that enable individuals with limited legal knowledge to draft complaints on their own. They also highlight the urgent need to bridge the gap between professional services and everyday users. Smart Legal Form Builder aims to address these systemic issues by automating the drafting of complaints, ensuring accuracy, accessibility, and compliance with legal standards while significantly reducing costs and efforts.

3. Vision and Goals

The core of Smart Legal Form Builder is to provide an accessible, intuitive and reliable solution. The app allows users to reduce the burden of filing a complaint and instantly create the necessary legal documents. The goal is to make filing a complaint easy and convenient for anyone, rather than just a professional.

The platform aims to eliminate the dependency on legal professionals and allow users to independently draft complaints. It automates complex processes and provides step-by-step questions and guidance. The platform aims to simplify the process of drafting complaints for four major categories: secondhand transaction fraud, online abuse, sexual harassment, assault and injury. By providing templates and guidance for these cases, Smart Legal Form Builder helps users complete complaints without the hassle of complex legal procedures. In doing so, the app reduces the burden of legal jargon, simplifies procedural requirements, and allows users to fully articulate their claims and demands.

Identify applicable funding agency here. If none, delete this text box.

At the end of the process, Smart Legal Form Builder envisions a more equitable legal environment:

- Users save significant amounts of time and money on legal documents.
- The legal system becomes more accessible to marginalized groups.
- Users experience the power and confidence to independently address their legal needs.
- It develops into a scalable framework that can be extended to other areas of complaint.

The success of this feature will be measured not only by the efficiency and accuracy of the complaints it generates, but also by its transformative impact in closing the gap in legal access and creating a more just society.

II. DATASETS

The dataset used to train the Complaint Generation AI Model consists of three CSV files, each representing different sections of a legal complaint. These datasets are provided in text format and include the essential information required to generate a complaint, such as details of the complainant, defendant, and incident description.

1. Dataset Structure:

The dataset is divided into the following categories:

- Complainant Information: Name, Address, Phone Number, National ID, Occupation
- Defendant Information: Name, Address, Phone Number, National ID, Occupation
- Incident Description: Incident Date, Location, Incident Details, Outcome
- Complaint Intent: Desired Legal Outcome
- Attached Documents: Medical Certificate, Witness Statements

Each of these fields is provided in text format and structured to allow the AI model to generate a complaint that follows a standardized legal format.

2. Dataset Generation Method:

The dataset was generated using GPT-2, which was employed to produce various text formats and variations needed for complaint generation. The model directly generated data for complaint templates, including complainant information, defendant details, incident descriptions, and complaint intent. These generated texts were then used as training data for the AI model.

3. Why We Didn't Use Existing Datasets:

We opted not to use pre-existing datasets for complaint generation and instead chose to generate the dataset using GPT-2 for the following reasons:

- Lack of Comprehensive Legal Document Datasets: Existing complaint datasets are often limited in scope or only cover specific types of incidents. This limitation makes it difficult to train a model that can generalize across various legal scenarios. Furthermore, the complexity of legal terminology and the variability of incidents necessitate the creation of diverse templates to ensure that the model can handle multiple types of complaints effectively.
- Leveraging GPT's Data Augmentation Capabilities: By utilizing GPT, we were able to automatically generate a wide variety of complaint documents tailored to different legal situations. This approach allowed us to expand the training dataset without requiring manual effort, thereby facilitating a more generalized learning process. The diverse set of generated complaints provided the model with the flexibility to handle a wide range of real-world legal scenarios, making the system more robust and adaptable.

III. METHODOLOGY

1. Model and Environment Set:

In this research, we selected the GPT-2 model for complaint generation due to its effectiveness in generating coherent and contextually relevant text. GPT-2 is a natural language generation model that specializes in text generation and is highly capable of automating the creation of legal documents, such as complaints, ensuring the generated text adheres to legal standards.

The environment setup for fine-tuning the GPT-2 model involves installing several essential libraries, including:

- Python 3.8+
- Hugging Face Transformers library
- PyTorch or TensorFlow (based on preference)
- Additional utilities: datasets, accelerate, and wandb (optional)

Once the environment is set up, the GPT-2 model and tokenizer are loaded to preprocess the text data.

2. Dataset Preparation

For this project, we use a text dataset formatted in plain text for training the GPT-2 model. The dataset consists of text entries that represent different sections of legal documents, such as complainant details, defendant information, incident description, and complaint intent. This dataset is used to train the model to generate legal documents based on predefined templates.

Dataset Structure:

The dataset is divided into the following categories:

- Complainant Information: Name, Address, Phone Number, National ID, Occupation
- Defendant Information: Name, Address, Phone Number, National ID, Occupation
- Incident Description: Incident Date, Location, Incident Details, Outcome
- Complaint Intent: Desired Legal Outcome
- Attached Documents: Medical Certificate, Witness Statements

The dataset is divided into four categories based on specific legal issues commonly encountered:

- Secondhand Transaction Fraud: Legal cases involving fraudulent transactions in secondhand goods.
- Online Abuse: Cases related to harassment, defamation, or abuse occurring through digital platforms such as social media.
- Sexual Harassment: Cases involving unwanted sexual advances or behaviors, often within workplaces or public spaces.
- Assault and Injury: Legal matters involving physical harm or threats of harm, such as assault or injury due to negligence.

Each category is used to generate a specific legal complaint, allowing the model to learn to handle various types of legal cases.

3. Data Preprocessing and Tokenization

Before training the model, the text data is tokenized. Tokenization involves breaking down sentences into smaller units (tokens) that the model can process. The text is preprocessed and tokenized using the GPT-2 tokenizer, ensuring it is in a suitable format for the model to understand.

The preprocessing step involves adding a prefix (e.g., "Generate complaint:") to each sentence and then tokenizing it into a format the GPT-2 model can process. Padding and truncation are also applied to ensure the text fits the model's input length requirements.

4. Model Fine-Tuning

Fine-tuning involves adapting the pre-trained GPT-2 model to the specific task of complaint generation. The model is trained on a dataset containing complaint templates and various legal document elements. This fine-tuning process adjusts the model's weights, enabling it to generate text that adheres to legal structures and templates.

A blank complaint template is created to guide the training process. The template includes four primary sections: complainant details, defendant information, incident

description, and complaint intent. The model learns to accurately fill in these sections based on the dataset.

The training process involves configuring training arguments, including the number of epochs, batch size, and learning rate. These parameters help control the learning process and ensure the model adapts well to the task of legal document generation.

5. Model Evaluation and Validation

Once the model is trained, it is evaluated using a validation dataset. The evaluation metrics used include accuracy, precision, recall, and F1-score to assess the model's performance in generating legally accurate complaints.

Additionally, the generated complaints are reviewed by legal experts to ensure compliance with legal standards. This validation step ensures that the model generates coherent text that also meets the necessary legal criteria.

6. Hyperparameter Optimization

During the training process, hyperparameters such as learning rate, batch size, and the number of epochs are optimized to improve the model's performance. These hyperparameters are tuned based on experimental results to achieve better performance in generating legally compliant text.

By fine-tuning the GPT-2 model on complaint templates, this methodology enables the automatic generation of legally compliant complaints. The process includes training the model on high-quality datasets, fine-tuning it to adhere to legal document standards, and evaluating its performance using various metrics. This system significantly advances the automation of legal document generation, making legal services more accessible to individuals and small businesses.

IV. Evaluation & Analysis

The Smart Legal Form Builder is currently undergoing training, and performance evaluation and analysis will be conducted once the model is fully trained and validated. However, based on the current development progress, the following evaluation and analysis directions have been established.

1. Evaluation Metrics

The key metrics for evaluating the model's performance include accuracy, precision, recall, and F1-score. These metrics are essential in determining how well the model generates text that aligns with legal standards for complaint generation.

- Accuracy: Measures how closely the generated complaints match the predefined templates and meet legal requirements.
- Precision: Evaluates the accuracy with which the model generates legally relevant information in the complaint text.
- Recall: Ensures that all necessary elements, such as complainant details and incident descriptions, are included in the generated complaint text.
- F1-score: Provides a balance between precision and recall, offering an overall performance metric that combines both aspects.

These evaluation metrics will guide the assessment of how effectively the model generates legally compliant and relevant complaints.

2. Initial Results and Analysis

Based on initial results from the model, basic details such as complainant and defendant information are being generated accurately. However, more complex sections like the incident description and legal outcomes require additional refinement. There is currently a potential for inaccuracies in describing the incident details or the legal outcomes, which will be addressed in future iterations of model training.

Furthermore, there is an observed data imbalance, particularly in categories like sexual harassment and assault, where there is limited training data. This imbalance may lead to lower model accuracy for these specific legal cases, and it will be a key area of focus for improvement.

3. Improvement Strategy

To enhance the model's performance, the following improvement strategies will be implemented

- Data Augmentation: Increase the training data for various incident categories, particularly for underrepresented legal cases, to improve the model's ability to generalize across diverse scenarios.
- Hyperparameter Tuning: Optimize key hyperparameters such as learning rate, batch size, and the number of training epochs to enhance the model's performance.
- Fine-tuning: Further fine-tune the model to specialize in different incident types, ensuring that legal requirements specific to each category are met effectively.

By employing these strategies, the model's performance is expected to improve over time, leading to more accurate and legally compliant complaint generation.

V. Related Work (e.g., existing studies)

1. Existing Tools and Platforms

A. Rocket Lawyer

Rocket Lawyer is a versatile platform that provides users with tools to create, manage, and store legal documents. It offers a variety of pre-built templates for common legal needs, including contracts, leases, and wills. The platform streamlines document creation with guided questionnaires, allowing users to customize templates with case-specific details. Rocket Lawyer also connects users with licensed attorneys to provide on-demand legal advice and document

reviews, giving users access to expertise when they need it. This combination of automation and human assistance makes Rocket Lawyer especially valuable to individuals and small businesses seeking comprehensive legal assistance.

B. LegalZoom

LegalZoom is one of the most well-known platforms for creating personalized legal documents online. It offers an extensive library of templates and provides users with a guided process to input relevant details and create custom documents for personal and business use. LegalZoom's services extend beyond document creation, including trademark registration, business formation assistance, and access to expert legal advice. Although not specialized in complaints handling, its user-friendly interface and wide range of offerings make it a great solution for individuals and businesses looking for efficient and accessible legal tools.

C. LawDepot

LawDepot focuses on enabling users to create legal documents quickly and efficiently through a user-friendly platform. It offers an extensive library of customizable templates that cover a wide range of legal needs, including contracts, agreements, and estate planning documents. By answering a series of guided questions, users can create documents tailored to their specific situation. LawDepot is especially useful for individuals looking for a cost-effective solution to handle everyday legal issues without professional assistance. The focus on customization and simplicity is consistent with the growing demand for accessible and simple legal tools.

2. Libraries and Frameworks

A. Flutter

Flutter is used to develop cross-platform applications, ensuring compatibility across Android and iOS devices. The framework leverages a single codebase, allowing developers to create responsive and interactive user interfaces. Flutter's widget-based architecture simplifies the implementation of key features such as step-by-step questionnaire modules, real-time document previews, and PDF generation capabilities. The platform's cross-platform compatibility provides a consistent and seamless experience for users. This allows users to reduce development time and maintain high performance.

B. Axios

Axios is used to manage real-time communication between the front end and the back end via HTTP requests and API calls. It plays a critical role in processing user input submissions from the questionnaire to the back end for processing by the AI model. Once the questionnaire is generated, Axios returns it to the front-end for the user to preview and download. This streamlined integration ensures fast, reliable, and lag-free interactions.

C. PDF Generation with Flutter Libraries

PDF generation and export is managed using Flutter-compatible libraries such as [syncfusion_flutter_pdf], which converts AI-generated text into professional-quality PDF documents. These documents are formatted to include appropriate headers, footers, and jurisdictional compliance elements, ready for submission or sharing. The ability to create custom layouts ensures that documents meet both legal and professional expectations.

D. Cassandra

Cassandra, a distributed NoSQL database, serves as the backend storage solution for user data, questionnaire responses, and generated complaints. The distributed architecture ensures high availability and scalability, allowing the platform to efficiently handle large amounts of data. Cassandra's high-traffic durability ensures that user data remains safe and accessible. This allows for seamless user interaction and reliable document storage, with future searchability.

3. AI and NLP Resources

A. OpenAI GPT-4 API

The OpenAI GPT-4 API is the underlying technology behind Smart Legal Form Builder, which can automatically generate well-formed and context-sensitive complaint forms. To optimize performance, the model was fine-tuned using a curated dataset categorized into four legal case types: second-hand fraud, online abuse, sexual harassment, and assault. The fine-tuning process involved collecting and preprocessing legal documents to highlight the specific terminology, logical structure, and formatting required for each category. By training the model on this categorized dataset, it was able to adapt to the nuances of legal language and format standards, ensuring that the output matches the user-provided input and legal requirements. Validation and testing demonstrated the model's ability to efficiently generate professional-quality documents, reduce reliance on manual drafting, and improve accessibility to legal services.

B. Hugging Face Transformers

Hugging Face Transformers played a pivotal role during the development phase of Smart Legal Form Builder, offering a flexible framework for testing various natural language processing tasks. Early prototypes leveraged its capabilities to explore text summarization, dynamic question-answering, and contextual understanding, all of which informed the design of the platform's step-by-step questionnaire interface. The framework also enabled the evaluation of different models' adaptability to legal datasets, providing valuable insights that shaped the fine-tuning strategy later implemented with the GPT-4 API. Although primarily used for research and experimentation, Hugging Face Transformers proved essential in refining the AI-based workflows and establishing a robust foundation for the platform's user-centric solution.

C. spaCy

spaCy is integrated into the platform to verify the structure, grammar, and correctness of the AI-generated text. Advanced language processing capabilities, including dependency parsing and named entity recognition, ensure that all generated documents are professional and compliant with legal standards. By providing this additional layer of verification, spaCy improves the reliability and trustworthiness of the feature output

4. Relevant Studies and Documentation

A. "AI in Legal Tech: Opportunities and Challenges"

This study highlighted the transformative potential of AI in legal technology, with a particular focus on automating document creation and streamlining complex legal processes. This enabled the platform to meet the nuanced needs of users by using dynamic question flow and context-aware text generation.

B. National Legal Awareness Surveys (2019, 2021)

This survey revealed widespread public concerns about legal terminology and accessibility, which in turn suggested the challenges of legal document creation and the need for a complaint writing app. The key findings of difficulty understanding legal documents and limited online engagement with legal resources led us to consider simplified processes and question-based complaint writing capabilities.

C. Blogs and Developer Resources

Practical resources from platforms such as OpenAI, React Native, and Cassandra helped us implement scalable and user-friendly systems. These guides helped streamline the integration of AI with front-end and back-end systems, ensuring optimal performance and seamless user experience.

VI. Conclusion: Discussion

Smart Legal Form Builder offers an innovative approach to overcome existing barriers in the legal system. Emphasizing user convenience, it utilizes AI technology, intuitive design, etc. to enable users to independently create accurate and legally compliant legal documents. The app simplifies the process of filing a complaint, reducing reliance on expensive legal services, and reducing the distance from the law that was perceived as complicated and difficult.

Key Achievements:

- Improved accessibility: The platform simplifies complex legal tasks, making legal documents accessible to non-experts.
- Cost and time efficiency: Users can create professional-quality documents in minutes, significantly reducing reliance on expensive legal services.
- Technological innovation: By leveraging cutting-edge AI and integrating scalable backend solutions, the platform sets a new benchmark in legal technology.

Challenges and Future Directions

While Smart Legal Form Builder has achieved its initial goals, several challenges remain:

A. Expansion of Case Types

- Expand beyond the services provided (e.g. Secondhand transaction fraud, online abuse, sexual harassment, assault and injury) by developing templates for additional case types such as real estate disputes, labor disputes, consumer complaints, and public service-related issues.
- Expand the scope of the platform to address a wider range of legal challenges by ensuring comprehensive support for diverse user needs.

B. Enhanced Integration with Legal Procedures

- Develop features that provide users with guidance to relevant legal authorities, such as nearby police stations or judicial offices, based on location and case type.
- Integrate with electronic legal systems to enable seamless complaint filing, allowing users to file complaints online directly from the platform.

C. Personalized User Support

- Introduce multilingual features for international users and improve accessibility for international users.
- Implement user profiling to analyze frequently handled case types and interests to provide tailored legal support and personalized recommendations.
- Utilize feedback data to continuously optimize the UI/UX and functionality of the platform to ensure continuous improvement based on real user experience.

D. Collaboration with Legal Professionals

- Integrate optional review features that allow users to connect with legal professionals to verify and improve draft documents.
- Work with attorneys to ensure that templates reflect the latest legal precedents and regulatory changes, ensuring that the platform's output is accurate and up-to-date.

E. Synthetic Data Utilization

- Develop advanced methods for generating synthetic data to supplement limited real-world examples in

specific case categories. This ensures that the AI model remains accurate and reliable even when user-provided data is sparse.

- Collaborate with legal professionals to review and validate synthetically generated datasets, ensuring they align with legal standards and practical scenarios.
- Leverage synthetic data to enhance the model's ability to generalize across diverse legal scenarios, enabling it to handle new case types or variations more effectively.

By addressing these challenges, Smart Legal Form Builder will solidify its position as a versatile and user-centric legal solution while expanding its capabilities to better serve its global audience.

In conclusion, Smart Legal Form Builder is not only a tool for automating legal documents, but also an innovative solution that creates a more inclusive legal ecosystem. Through continuous development and improvement, it will further democratize access to justice and help the legal system serve everyone equally.

VII. REFERENCES

[1] National Legal Awareness Survey, "Survey on public awareness of legal terminology and access to legal resources," 2019 & 2021. [Data set].

[2] OpenAI, "GPT-4 API documentation," 2024. [Online]. Available: https://platform.openai.com/docs

[3] Flutter Team, "Flutter documentation," 2024. [Online]. Available: https://flutter.dev

[4] Axios Project, "Axios documentation," 2024. [Online]. Available: https://axios-http.com

[5] Apache Software Foundation, "Cassandra documentation," 2024. [Online]. Available: https://cassandra.apache.org

[6] Syncfusion, "Syncfusion Flutter PDF documentation," 2024. [Online]. Available: https://www.syncfusion.com/flutter-widgets/flutter-pdf-viewer

[7] Hugging Face, "Transformers documentation," 2024. [Online]. Available: https://huggingface.co/docs/transformers

[8] Explosion AI, "spaCy documentation," 2024. [Online]. Available: https://spacy.io

[9] Brookings Institution, "Embracing AI in legal practice: Opportunities and challenges," May 6, 2024. [Online]. Available: https://www.brookings.edu/articles/embracing-ai-in-legal-practice-opportunities-and-challenges-the-techtank-podcast/