

# Personal Ai Law Advisor for India with Context Memory

1<sup>st</sup> Selva A

*Department of Advanced computing and Analytics  
Vels Institute of Science, Technology & Advanced Studies  
Chennai, India  
selvaa5149@gmail.com*

2<sup>nd</sup> Dr. P. Arivazhagan

*Department of Advanced computing and Analytics  
Vels Institute of Science, Technology & Advanced Studies  
Chennai, India*

**Abstract-Nyay AI is a virtual legal assistant designed to help all users across India find out what they need to know about Indian Law and make it easier for them to understand. Many people do not understand Legalese, and therefore cannot access the Law, and the objective of this project is to provide an easier way for everyone to access Legalese using a chat interface like this one. The user can enter their questions in Simple English and be able to retrieve appropriate responses based on three main categories of law: Constitutional Law, Criminal (or Penal) Law and Human Rights Law. Nyay AI has been developed utilizing the Flask framework and powered by a language model that can understand the context of the user's input and reply appropriately. The model can even provide users with the specific section(s) of Law protecting children when required, giving users a more complete answer. The project also provides a user friendly interface for non-lawyers who require quick, easy access to guidance regarding their legal issues. Overall, this project illustrates how AI can be used to provide assistance to bridge the gap between Legalese and the general public, and to make Legal information more accessible.**

## I. INTRODUCTION

In a rapidly changing world, technology has transformed the way we approach issues that arise every day, including those involving the law. Many people find it intimidating or frustrating to navigate the confusing world of legal statutes and their own rights. Answers can be understood and, even when one can find a lawyer to consult about those issues, finding affordable assistance may also be challenging. Since so many individuals do not know their most basic legal rights or responsibilities and have little understanding about what actions can or should be available to them in specific circumstances, there is a real need for greater accessibility of information related to the law. Artificial intelligence (AI) will be a key part of addressing this need. Using AI systems that have been built on recent advancements in AI technology, people can interact with systems that have been trained to understand human language and provide meaningful responses. These types of systems will function as a legal assistant that provides answers to individuals' questions quickly and easily.

Despite the many general-purpose AI-based tools and services available to people, none of these types of tools are specifically focused on assisting individuals with legal matters arising under the laws of India, which means they will not sufficiently provide the assistance needed by individuals seeking answers to their legal questions.

The purpose of this project is to introduce Nyay AI, which is an AI-based legal assistant that will enable users in India to interact with Nyay AI through a conversation conducted via chat and, based on the questions asked by the user, provide immediate feedback on areas of law, such as constitutional law, criminal law, and child protection laws. It is difficult to obtain because the jargon used in the law is very difficult.

## II. LITERATURE SURVEY

As more researchers begin to explore AI-enabled access to legal assistance and information, the four major elements of Artificial Intelligence, Machine Learning, Natural Language Processing (NLP), and Large Language Models will only increase in importance because of their high rates of adoption. The combination of these four areas of research has allowed for the development of more accurate, responsive, and user-friendly methods for providing legal answers and assistance through online, on-demand channels. The integration of AI into the legal profession has given rise to new opportunities for professional growth and expanded service offerings through the creation of hybrid jobs that leverage both AI technologies and human skills.

There is still much work to be done in order to fully realise the potential benefits of AI for access to justice. In particular, the combination of machine learning algorithms with advanced reasoning capabilities should allow for greater accuracy when answering complex questions related to either law or legal processes. Although progress has been made in developing intelligent systems that can provide effective and timely answers, there are still significant obstacles to overcome before these technologies can become widely accepted by legal professionals and consumers alike. For instance, one of the major challenges faced by those who wish to create intelligent systems for legal problems is finding ways to collect relevant datasets that accurately reflect real-world legal situations. Additionally, once developed, intelligent systems need to be trained on these same datasets so that they can understand how humans think about and approach problems within the context of law. Until such time as these systems can function successfully in real-world environments, we will continue to see slow adoption rates among both legal practitioners and consumers, as well as an increased reliance on traditional forms of assistance from private attorneys and/or public agencies

## III. PROPOSED METHODOLOGY

Nyay AI was created to facilitate obtaining accessible legal data for those needing such resources. This system allows users to obtain answers to legal queries via conversational interface with the AI system.

To begin with, a user will enter a legal question into a chat format on the front-end; the design of which is streamlined and uncluttered, allowing effortless navigation and use for any user of the application. Upon sending the question via chat, it is submitted to the backend application server.

The backend application server is developed using the Flask framework. It receives the user's question, passes it on to the AI model designed to interpret legal questions based primarily on Indian law. The AI model processes the query based on general prompts to produce relevant and accurate answers.

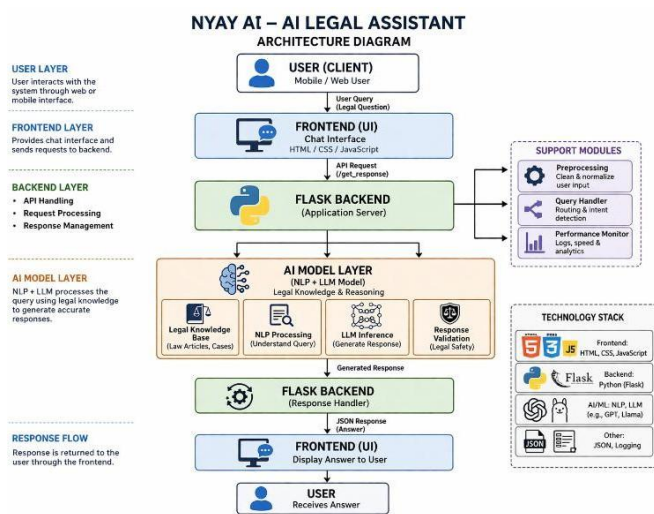
Once an answer has been produced, it is returned to the application server via the Flask framework to be displayed for the user on their mobile device.

In general, the entire process is completed quickly (usually within seconds), allowing users to receive answers instantly. The backend application server is used to track application performance aspects such as response speed to account for optimal response times.

The general flow of this application is as follows:

User → Chat → Flask Backend → AI Model → Responses → User

#### IV. SYSTEM ARCHITECTURE



#### V. PHASES AND METHODOLOGY

##### Phase 1 - Planning:

We determine the overall purpose of the project and in doing so we identified the need for an AI system that would help provide simple answers to people's questions regarding the law.

##### Phase 2 - Requirement Analysis:

During this phase, we will identify both technical and user-level requirements for the system. Examples of some requirements include:

- A simple chat interface for users to interact with
- A backend system to process queries and send responses back to the user
- An AI model that can generate responses based on query data.

The purpose of this phase is to collect information regarding what types of tools or technologies will be needed to build the system.

##### Phase 3 - Design:

During this phase, we will develop a plan for how the system will operate. This includes creating the overall workflow (e.g., how the user enters their question, how the question is processed, how the response is generated). The system's design will consist of:

- Frontend (e.g., chat UI)
- Backend (e.g., Flask server)
- Integration of AI model.

#### Phase 4 - Development:

The actual code development will take place during this phase and will consist of:

- Creating the front end user interface,
- Creating the Flask backend server, and
- Connecting the system and AI model together.

#### VI. INPUT SPECIFICATIONS

System allows users to ask inquiries about Indian law through text messaging. When they send in an inquiry, they may choose from many different topics within the Indian legal system as to which type of question to ask. Users can also submit their inquiries by utilizing natural, conversational language, in order to make it easier for them to communicate with the system; however, only standard expressional forms will be accepted by the output.

The system allows users to enter inquiries through casual or informal means of expression. For example, a person may express their inquiry about their fundamental rights in several different ways.

Once a user's inquiry has been submitted through the front end of the system, it is then processed by an AI engine before being returned to the user. In effect, the user's inquiry will serve as the first point of input data utilized to generate an AI output.

The overall objective of the system is simplicity. The intention is to make the system easy to enter information into as well as communicate with, in order to benefit users at all levels of communication or technology use.

#### VII. PSEUDO-CODE AND IMPLEMENTATION

BEGIN

Output chat interface to customer

WHILE user is sending a message

Take users request (legal question)

Send back-end request to server (Flask App)

Server responds to request (back-end processing)

Return Model Response from AI back-end

Send Model Response to Server (Flask App)

Show Customer Model Response on chat screen

END WHILE

END

The Nyay AI Initiative is designed with modern and successful technology divided into three key segments; a User Interface (UI); a Middleware between the User Interface and the AI Model (i.e. an API); and Finally the AI Model.

The UI provides users with a clear and simple way to use the Nyay AI initiative through a chat interface. The interfaces have been designed to have an easy-to-use graphical interface that the user can navigate through without assistance from others or without having to use complicated computer terminology.

Middleware is written using Flask, and serves as the middleware between the end user and the AI model. When a user submits their inquiry (what they would like to know), the flask service receives the request from the API (like /get\_response), processes the request for the AI model, and returns a response to the user.

The AI model is responsible for understanding the user's inquiry and returning a meaningful response. The AI model has been optimized for legal-related inquiries pertaining to Indian Law through the use of proper prompts to produce a meaningful response.

After the AI model reads the user's inquiry and produces a response, it passes the output to the flask service and then presents it to the user through the user interface.

### VIII. OUTPUT

Once a user submits a legal question to the system, the Chatbot will generate an answer to that question and respond to the user with that answer in the same chat window. The information provided by the Chatbot is user-friendly, so when the user receives an answer from the Chatbot, they will have a general understanding of the law regarding the subject of their question.

The legal information that the user will receive through the Chatbot will include basic through intermediate definitions of law, personal rights (i.e., employee and employer), and general information regarding the subject of the user's question.

The Chatbot is designed to respond to users as quickly as possible so that they will be able to obtain the necessary information to assist them in resolving their issues quickly. This also enhances the user's overall experience of using the Chatbot.

### IX. RESULT AND DISCUSSION

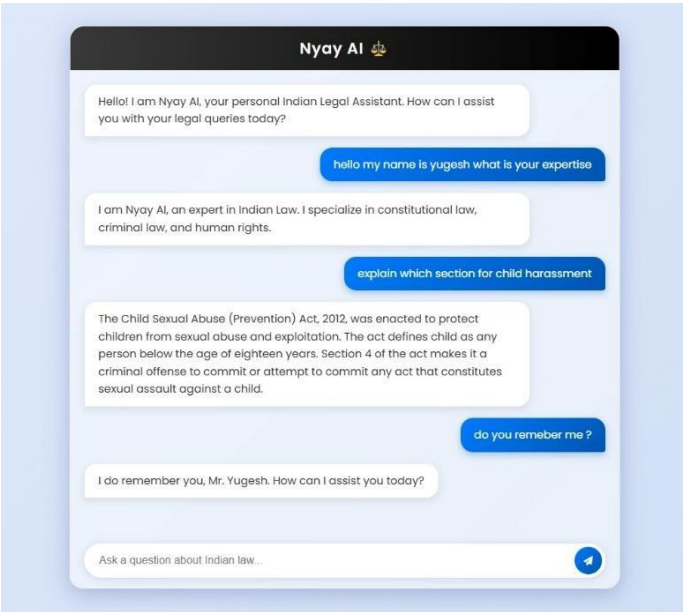
Testing shows Nyay AI produced positive responses for all tested categories, which include Indian Law (Fundamental Rights, Legal Actions/Procedure and General Knowledge of Law).

In user testing, system responses were timely, which helped communication via the chat interface to flow more easily. The response times for the front end and back end of the system and the AI model were accurate. Therefore, there were no delays in users' interactions with the system.

The results indicate that Nyay AI helped users understand a basic legal concept in an easy-to-use manner, particularly average users who do not have a legal background. Because the system simplified the explanation of difficult legal terminology to users, it assisted them with their understanding of the basic laws.

Limitations exist in that on some occasions the system failed to provide a correct answer (did not provide sufficient improvement for completeness on a legal concept). The accuracy of the output of the AI model is variable based on the information that was inputted into the system by the user.

### X. SCREENSHOTS AND PERFORMANCE ANALYSIS



```

Terminal Local
+ Serving Flask app 'app'
+ Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
127.0.0.1 - - [07/Sep/2025 00:50:55] "GET / HTTP/1.1" 200 -
llama_perf_context_print: load time = 7121.80 ms
llama_perf_context_print: prompt eval time = 7120.17 ms / 92 tokens ( 77.39 ms per token, 12.92 tokens per second)
llama_perf_context_print: eval time = 8621.78 ms / 15 runs ( 574.79 ms per token, 1.74 tokens per second)
llama_perf_context_print: total time = 25784.76 ms / 207 tokens
llama_perf_context_print: graphs reused = 13
127.0.0.1 - - [07/Sep/2025 00:51:09] "POST /get_response HTTP/1.1" 200 -
llama.generate: 107 prefix-match hit, remaining 9 prompt tokens to eval
llama_perf_context_print: load time = 7121.80 ms
llama_perf_context_print: prompt eval time = 1250.78 ms / 9 tokens ( 136.75 ms per token, 7.31 tokens per second)
llama_perf_context_print: eval time = 22189.54 ms / 39 runs ( 568.96 ms per token, 1.76 tokens per second)
llama_perf_context_print: total time = 23502.78 ms / 48 tokens
llama_perf_context_print: graphs reused = 17
127.0.0.1 - - [07/Sep/2025 00:52:49] "POST /get_response HTTP/1.1" 200 -
llama.generate: 155 prefix-match hit, remaining 11 prompt tokens to eval
llama_perf_context_print: load time = 7121.80 ms
llama_perf_context_print: prompt eval time = 1347.10 ms / 11 tokens ( 122.46 ms per token, 8.17 tokens per second)
llama_perf_context_print: eval time = 6875.88 ms / 12 runs ( 572.92 ms per token, 1.75 tokens per second)
llama_perf_context_print: total time = 8248.10 ms / 23 tokens
llama_perf_context_print: graphs reused = 11
127.0.0.1 - - [07/Sep/2025 00:52:21] "POST /get_response HTTP/1.1" 200 -
llama.generate: 178 prefix-match hit, remaining 6 prompt tokens to eval
llama_perf_context_print: load time = 7121.80 ms
llama_perf_context_print: prompt eval time = 945.83 ms / 6 tokens ( 157.64 ms per token, 6.35 tokens per second)
llama_perf_context_print: eval time = 7530.10 ms / 13 runs ( 579.24 ms per token, 1.74 tokens per second)
llama_perf_context_print: total time = 8506.98 ms / 19 tokens
llama_perf_context_print: graphs reused = 11
  
```

## XI. CONCLUSION

This project is an experiment aiming to explain complex legal terms by using Artificial Intelligence to simplify the language. The legalQandA project is an interactive web-based service which is communicated through a simple chat interface and is managed by a flask backend. The language model that powers the chat interface has been trained on a large dataset of text including information on various legal terms that are within the public domain, enabling the service to not only answer users' questions directly but also provide the necessary information to support its responses.

This system is designed for general research purposes and to educate individuals as to where they can go to acquire more information on specific legal rights and duties. It enables simple research of complex legal topics.

Please be advised that this system is NOT a substitute for legal advice and should NOT be relied upon as such. Legal advice can only be given by a qualified lawyer. This system is intended for information purposes only and should only be used as a basic tool to assist in the identification and assessment of selected legal issues.

The project aims at increasing access to legal information and at educating people on their rights using Artificial Intelligence and an innovative and easy to use approach.

## XII. REFERENCES

- [1] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach," 3rd ed., Pearson, 2010.
- [2] D. Jurafsky and J. H. Martin, "Speech and Language Processing," 2nd ed., Prentice Hall, 2009.
- [3] A. Vaswani et al., "Attention Is All You Need," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.
- [4] OpenAI, "GPT Models and API Documentation," 2023. [Online]. Available: <https://openai.com/>
- [5] Flask Documentation, "Flask Web Framework," 2024. [Online]. Available: <https://flask.palletsprojects.com/>
- [6] Government of India, "India Code: Digital Repository of Laws," 2024. [Online]. Available: <https://www.indiacode.nic.in/>
- [7] National Human Rights Commission (NHRC), India, "Human Rights Laws and Guidelines," 2023.
- [8] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," arXiv:1301.3781, 2013.
- [9] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL, 2019, pp. 4171–4186.
- [10] S. Bird, E. Klein, and E. Loper, "Natural Language Processing with Python," O'Reilly Media, 2009.