



Vidhi.ai : The RAG Based AI Agent for Indian Legal System

Ayush Raj

Kashish Tiwari

Groclake Agentathon

RAG : Retrieval Augmented Generation

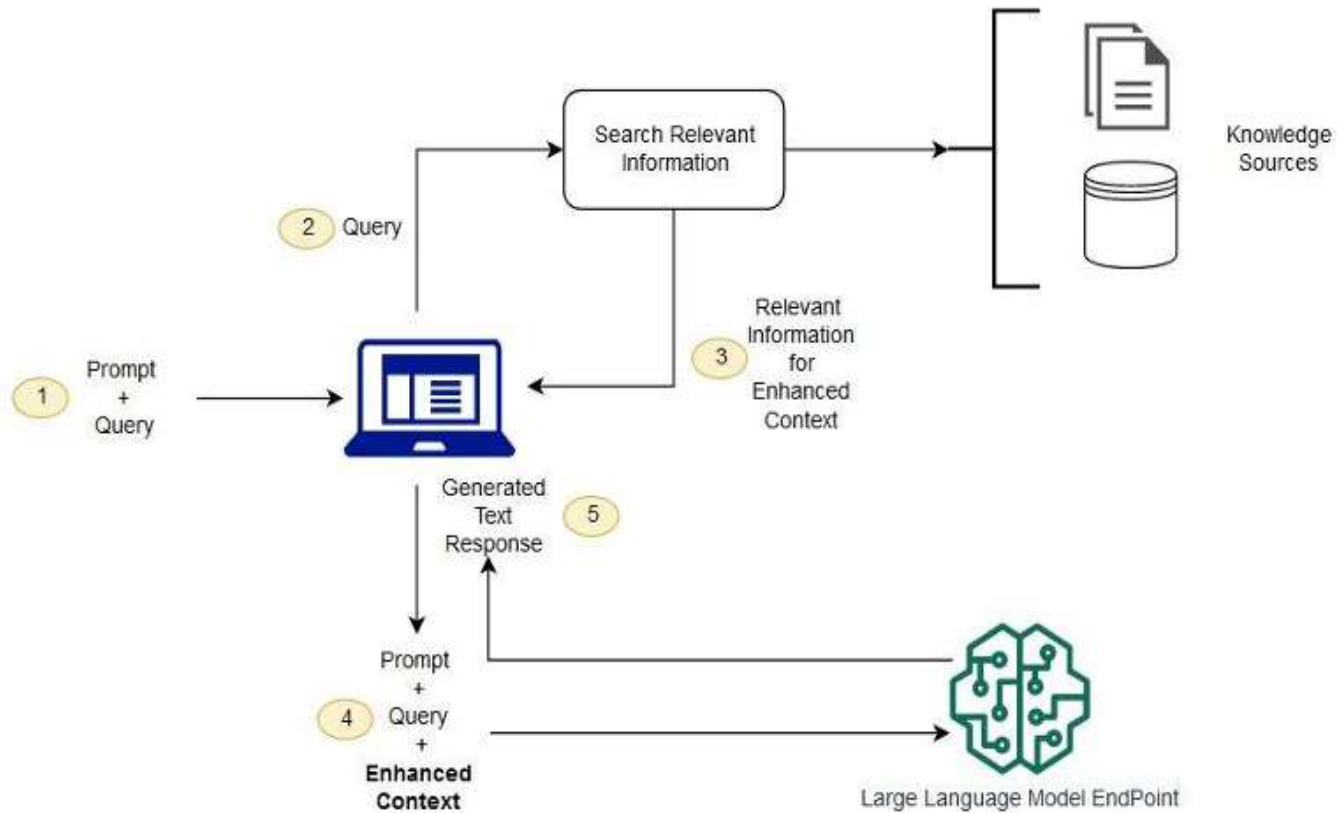
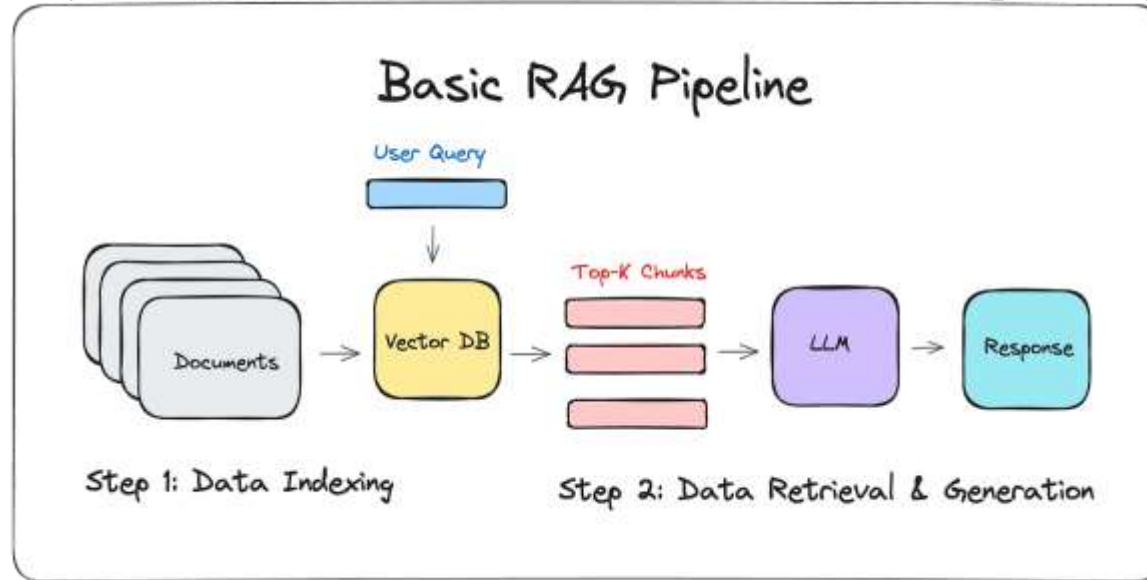


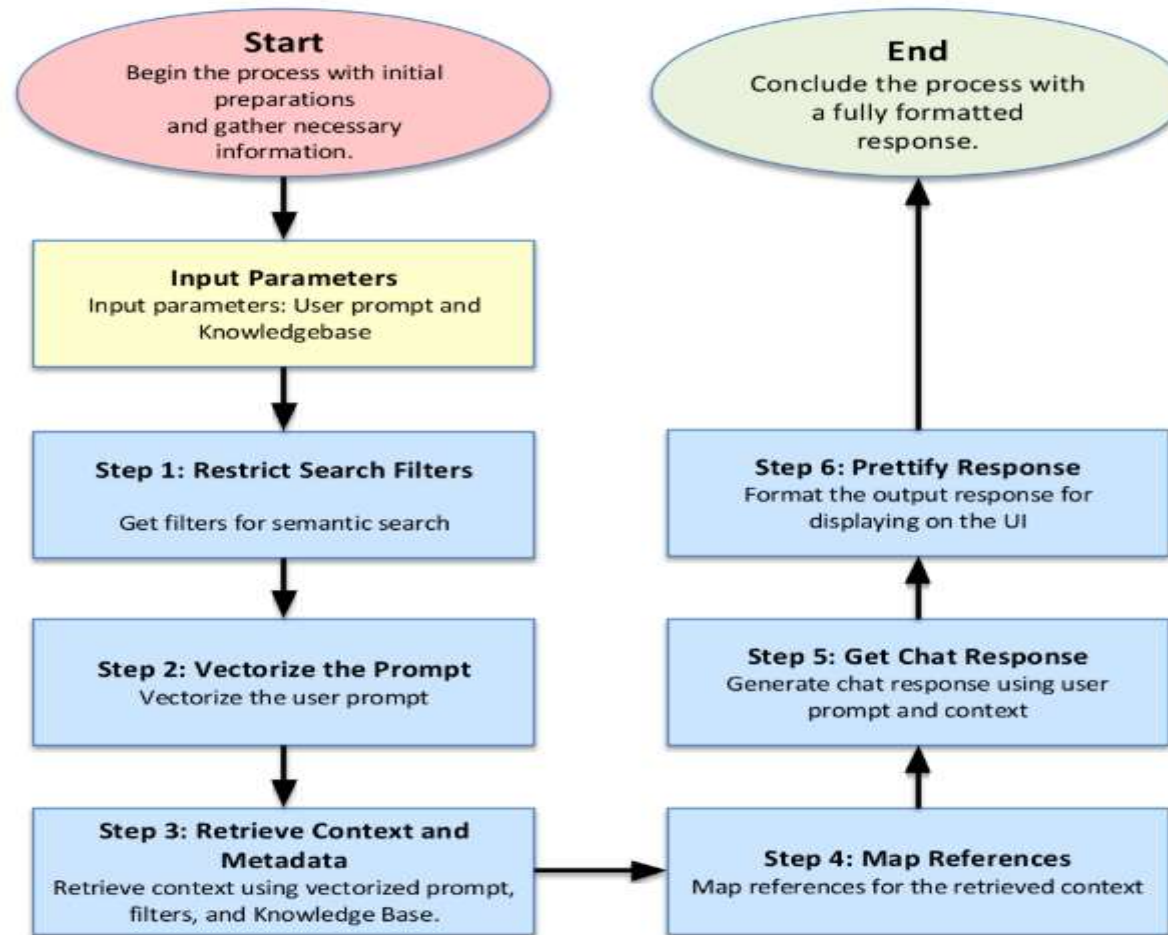
Fig.1. RAG Pipeline [1]

Introduction to RAG:

- A hybrid AI model that combines retrieval-based search with generative models to enhance response quality.
- Retrieves relevant documents, integrates them with an LLM, and generates informed responses.
- Key Components : Retrieval mechanism (vector databases, embeddings), generative model (LLMs like GPT), and external knowledge sources.
- Improves factual accuracy, reduces hallucinations, and enhances adaptability for domain-specific tasks.

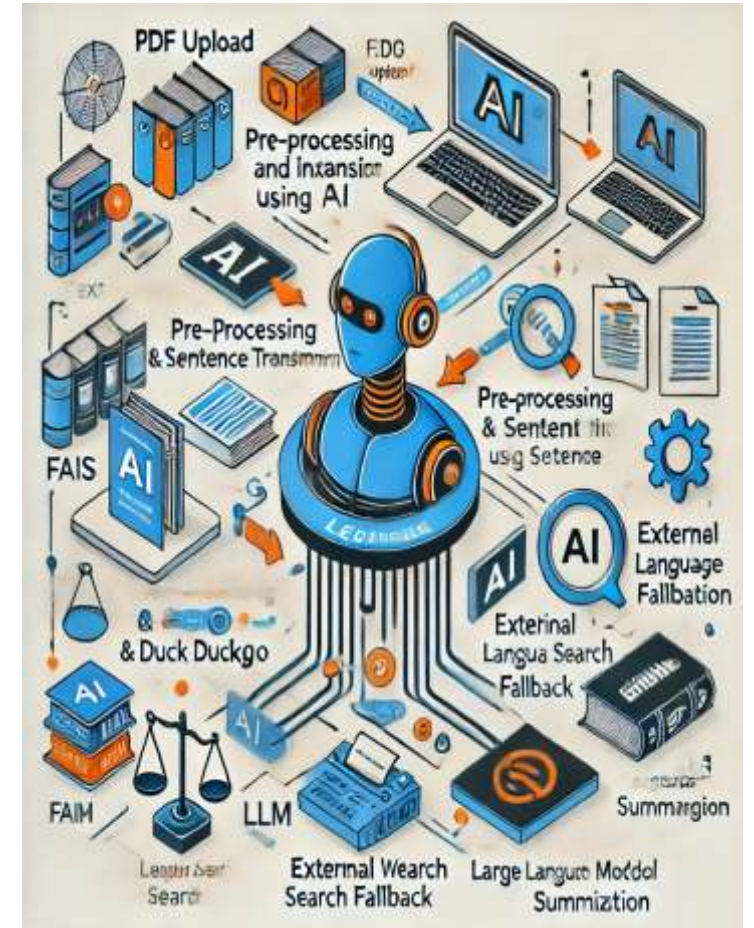


Process Flowchart :



End-to-End Process:

- **Document Upload & Text Extraction**
 - Users upload PDFs containing legal information.
 - PyPDF2 extracts text, which is stored in a knowledge base.
- **Knowledge Base Creation**
 - Text is pre-processed, cleaned, and split into chunks.
 - Indexed using FAISS & Sentence Transformers for efficient retrieval.
- **Semantic Search & Web Search Fallback**
 - Queries trigger a semantic search on stored text.
 - If no relevant results, DuckDuckGo search provides external context.
- **Text-Summarization & Response Generation**
 - LLM summarizes and prompt tuning ensures key details for legal responses.
- **Gradio Chat Interface**
 - User-friendly interface for PDF upload & chatbot interaction.
 - Features include document summarization & chat history.
- **Error Handling & Deployment**
 - Robust error handling for text extraction, search, and response failures.
 - Deployed via Gradio for multi-user access.



Integration of Groclake Tools

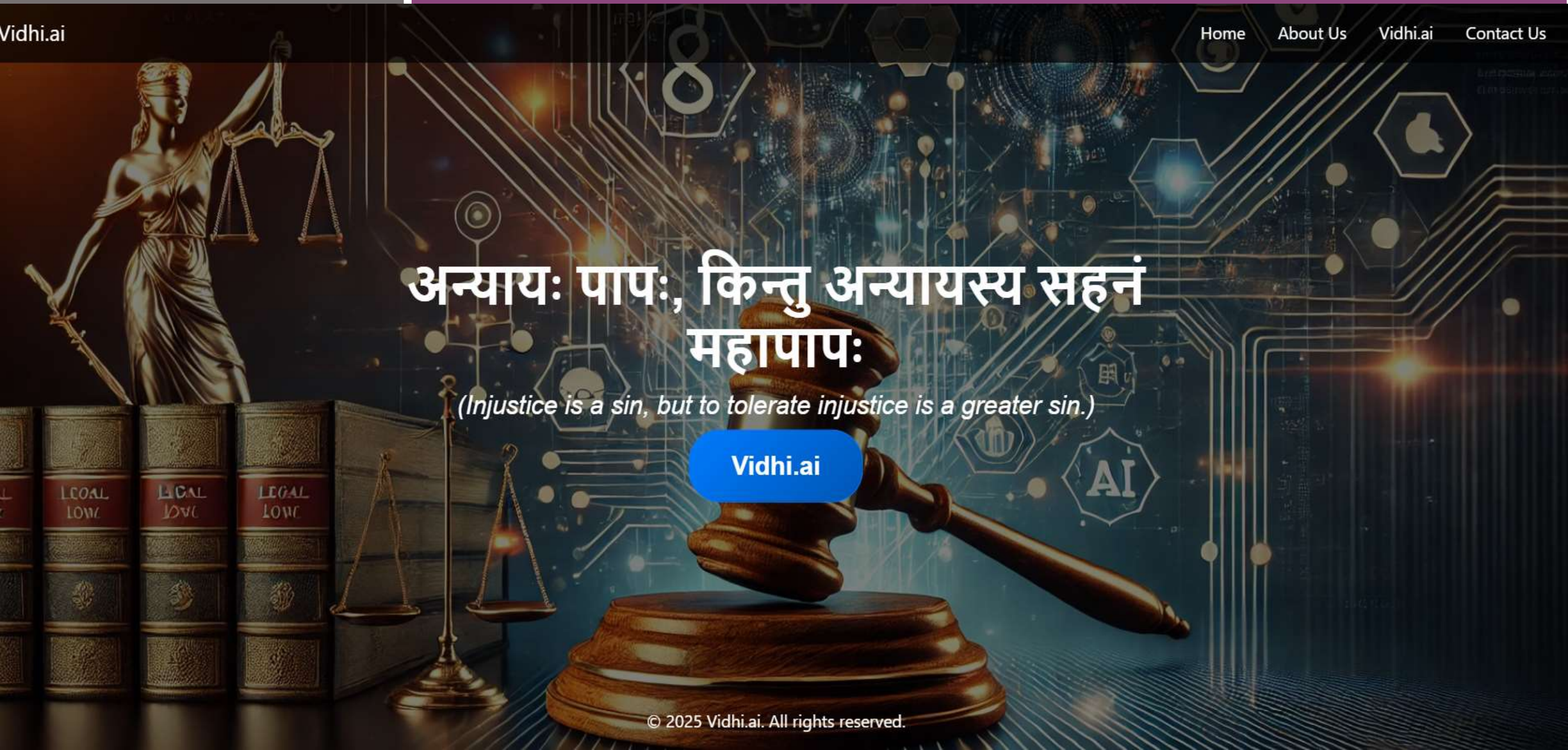
➤ ModelLake:

- ModelLake serves as the core foundation of our project, enabling seamless integration, management, and utilization of GPT-4.

➤ Datalake & Vectorlake:

- We initially aimed to integrate DataLake as a data warehouse and VectorLake for storing our vectorized knowledge base.
- However, due to our current level of proficiency, we relied on open-source libraries to complete the project. While this was a necessary compromise, we are eager to expand our expertise in these areas.
- We are confident that mastering DataLake and VectorLake will significantly enhance the efficiency and capability of our project in the future.

Getting Started With : Home Page



Vidhi.ai

Home

About Us

Vidhi.ai

Contact Us

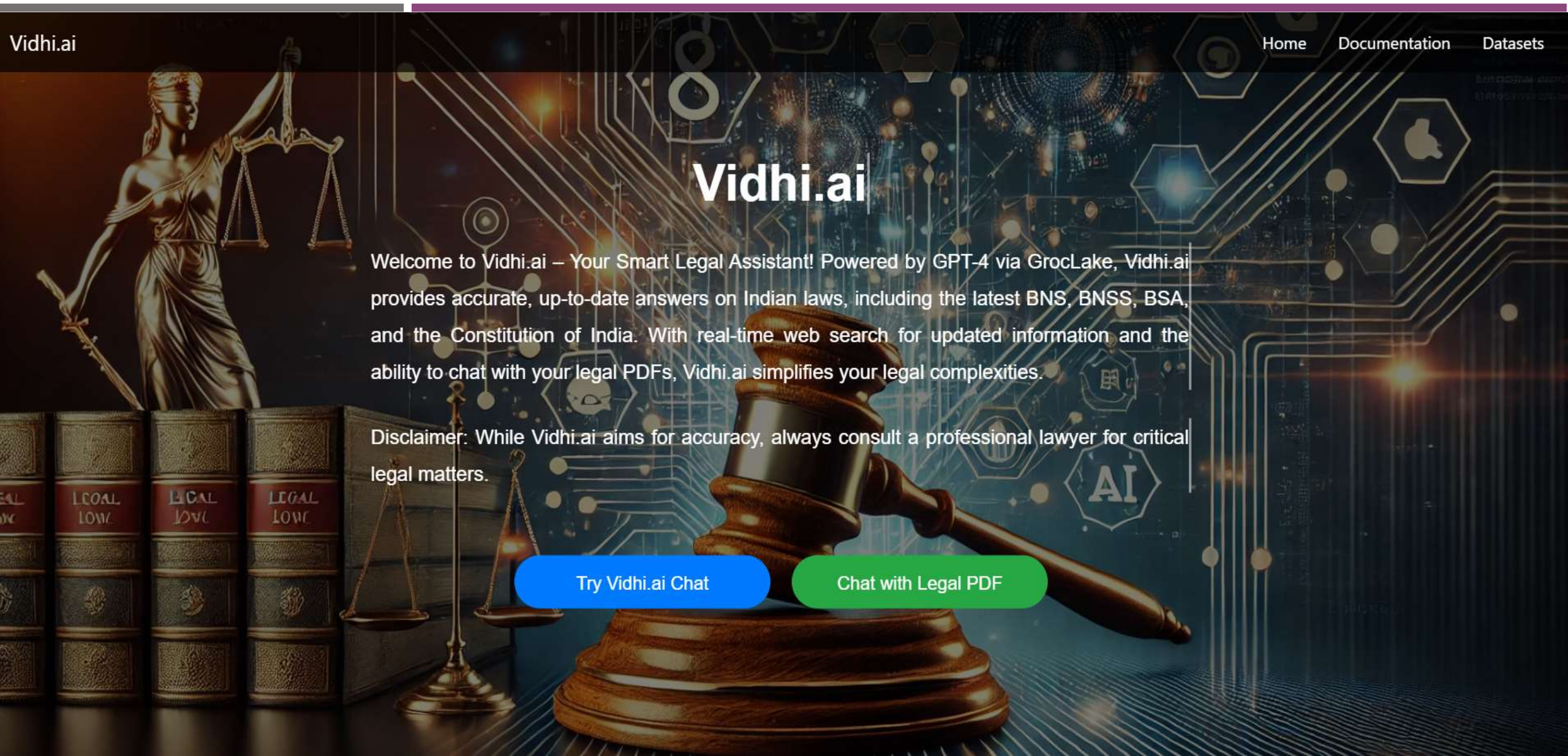
अन्यायः पापः, किन्तु अन्यायस्य सहनं
महापापः

(Injustice is a sin, but to tolerate injustice is a greater sin.)

Vidhi.ai

© 2025 Vidhi.ai. All rights reserved.

Getting Started With : Welcome Page



Vidhi.ai

[Home](#)

[Documentation](#)

[Datasets](#)

Vidhi.ai

Welcome to Vidhi.ai – Your Smart Legal Assistant! Powered by GPT-4 via GrocLake, Vidhi.ai provides accurate, up-to-date answers on Indian laws, including the latest BNS, BNSS, BSA, and the Constitution of India. With real-time web search for updated information and the ability to chat with your legal PDFs, Vidhi.ai simplifies your legal complexities.

Disclaimer: While Vidhi.ai aims for accuracy, always consult a professional lawyer for critical legal matters.

[Try Vidhi.ai Chat](#)

[Chat with Legal PDF](#)

Getting Started With : ChatApp

Vidhi.ai

Home

Vidhi.ai

Documentation

Datasets

Vidhi.ai: Indian Legal Assistance Agent

Ask any legal question related to Indian law, and I'll try to provide concise advice.

Chatbot

what is indian constitution

The Indian Constitution, the paramount legal document of India, outlines the political code, structure, procedures, powers, and responsibilities of governmental entities. It encapsulates the fundamental values and principles of India, also referred to as Bharat, a Union of States. The Constitution provides the guiding framework for the country's governance and functioning, ensuring the rule of law, democratic principles, and protection of fundamental rights. It also establishes the separation of powers among the legislative, executive, and judiciary branches, fostering checks and balances. The Constitution is amendable, allowing it to evolve with changing societal needs, but its basic structure remains inviolable.

what does article 16 of indian constitution states?

Article 16 of the Indian Constitution states that there shall be equality of opportunity for all citizens in matters relating to employment or appointment to any office under the state. It prohibits discrimination on grounds of religion, race, caste, sex, or place of birth. It also empowers the state to make provisions for reservation of appointments or posts in favor of any backward class of citizens which, in the opinion of the state, is not adequately represented in the services under the state. It emphasizes on ensuring equal opportunities to every citizen.



|

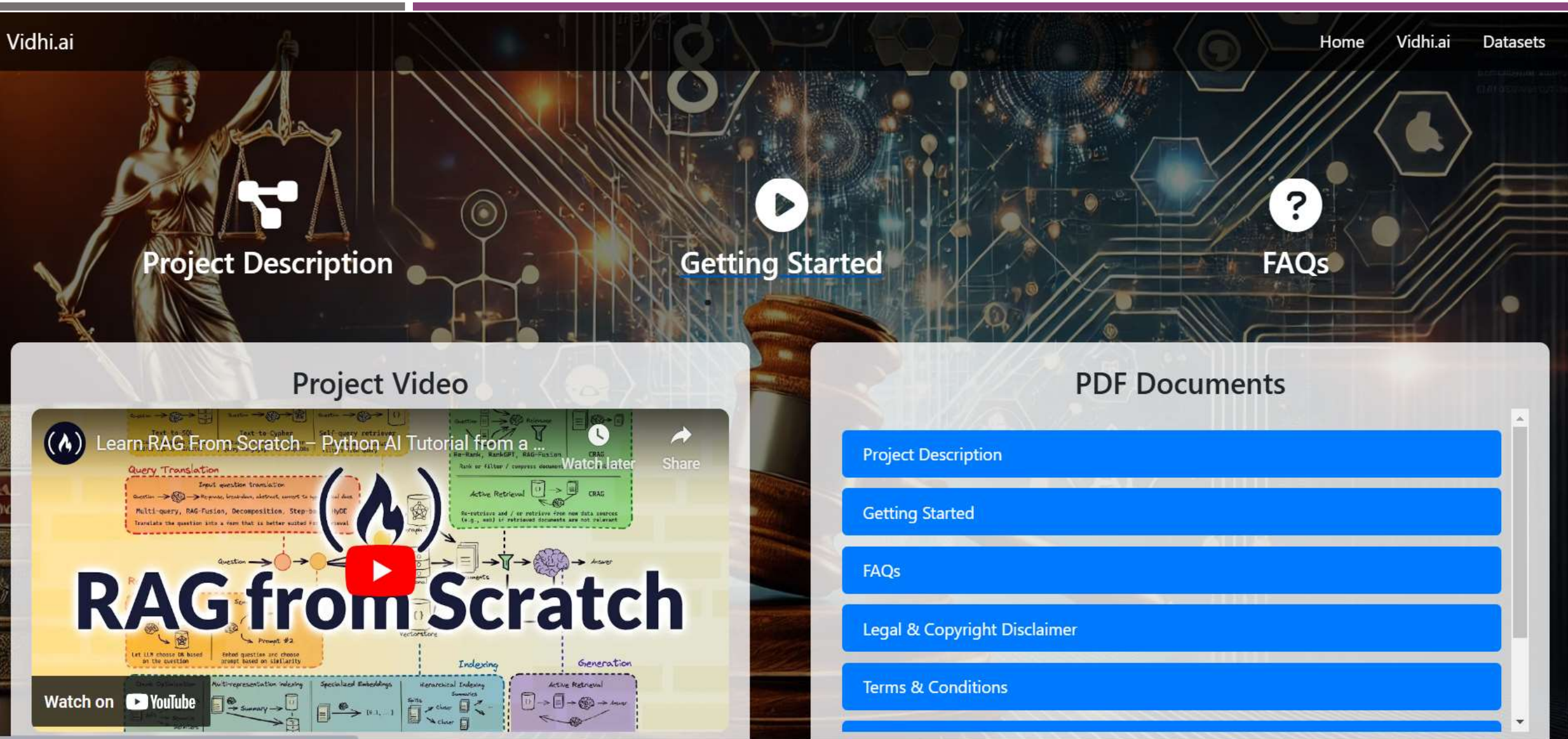


Getting Started With : Vidhi.ai

Vidhi.ai provides instant answers to legal queries, whether about recent reforms or fundamental laws.

- **Response Time:** Currently deployed on Hugging Face Spaces using CPU, leading to an average response time of **~10 seconds** per query. Future upgrades to GPU/TPU will significantly reduce this.
- **Concise & Informative Answers:** To handle GPT-4's token limits, responses are optimized to be brief yet comprehensive, ensuring all essential details are included.
- **Powered by RAG:** Uses **Retrieval-Augmented Generation (RAG)** and an up-to-date legal knowledge base for accurate responses.
- **Future Enhancements:** Based on user feedback, we plan to integrate models with higher token limits (e.g., DeepSeek models) for even more detailed answers.

For detailed documentation, refer this :



Use Cases

1. India-Specific Legal Knowledgebase:

- Problem: Most general-purpose chat models are trained on vast, non-specific corpora, making them less reliable for Indian legal inquiries.
- Solution: Vidhi.ai was trained specifically on Indian laws, the Constitution of India, and relevant legal texts, creating a focused knowledgebase.

2. Real-Time Updates with Web Search:

- Problem: Indian laws are constantly evolving through amendments, judicial pronouncements, and changes.
- Solution: Vidhi.ai integrates a web search feature that ensures access to the most current legal information.

3. Continuous Learning and Improvement:

- Problem: Users' legal queries become more specific and complex over time, demanding continuous model improvements.
- Solution: Since Vidhi.ai uses a GPT-based model, it improves with each legal query by adapting to specific legal language and context.








4. Future Vision: Fine-Tuning a Custom LLM

- Problem: While Vidhi.ai's current model is effective, there's potential for even greater performance with a fully customized legal model.
- Solution: Vidhi.ai plans to fine-tune a custom Legal Language Model (LLM) based on India-specific legal datasets.

Some Suggestions :

- Requirement of detailed documentation on VectorLake and DataLake
 - **Insufficient documentation:** Limited guidelines on API integration and query execution.
 - **Query optimization:** Lack of clear strategies for efficient vector-based retrieval.
 - **Indexing methods unclear:** No detailed explanation of indexing algorithms used (e.g., FAISS, Annoy, HNSW).
- Kindly increase the token limit of GPT-4 via ModelLake
- Facilitation to connect different other LLMs integration through ModelLake.
- More detailed You-tube video tutorials to make it more student and new user friendly.

Overall RAG updates : Visit RAG update page

Vidhi.ai				Home	Vidhi.ai	Documentation	Datasets
Category	Status	Details	Next Steps				
PDF Text Extraction from Legal documents for creating Knowledgebase	 Green	Text extraction from PDFs is fully functional. Handles most formats well.	Next Steps				
Semantic Search Implementation	 Green	Semantic search works as expected. Retrieves relevant chunks from documents.	Next Steps				
DuckDuckGo Web Search Integration	 Green	Web search fallback is operational and provides useful results.	Next Steps				
AI-Powered Summarization	 Amber	Summarization works but needs fine-tuning for better brevity and accuracy.	Next Steps				
Gradio Chat Interface	 Green	Chat interface is fully functional and user-friendly.	Next Steps				
Indian Law Specialization	 Amber	System is trained on Indian laws but needs more data for better accuracy.	Next Steps				
Error Handling & Fallbacks	 Green	Robust error handling is in place for most edge cases.	Next Steps				

Thank-You

```
{  
  "message": "Thank you!",  
  "details": {  
    "language": "English",  
    "context": "Expression of gratitude",  
    "usage": "General"  
  }  
}
```