

CS699 Project - Legal Today

Aditya Gaikwad (25M0745)

Madhur Thakkar (25M0742)

Monil Desai (25M0744)

Contents

Title	Page No
Problem Statement	3
Scope & Limitation	4
Architecture Diagram	5
Tech Stack	6
SQLite DB Schema	7
Team Contributions	8-9
Git Logs	10-12
Implementation Screenshots	13-18
Challenges	19
Future Scope	20

Problem Statement

1. **Inaccessibility of Legal Information:** Legal judgements and court documents are typically lengthy, dense, and filled with complex jargon, making them difficult for the general public to comprehend without expensive expert assistance.
2. **Inefficiency of Traditional Search:** Conventional keyword-based search engines often fail to capture the semantic nuance of legal queries, making it frustratingly difficult to find relevant case precedents when the exact legal terminology is unknown.
3. **Difficulty in Finding Specialized Representation:** Users often struggle to identify and connect with lawyers who have specific expertise relevant to their unique legal situation (e.g., finding a "divorce" lawyer vs. a "corporate" lawyer).

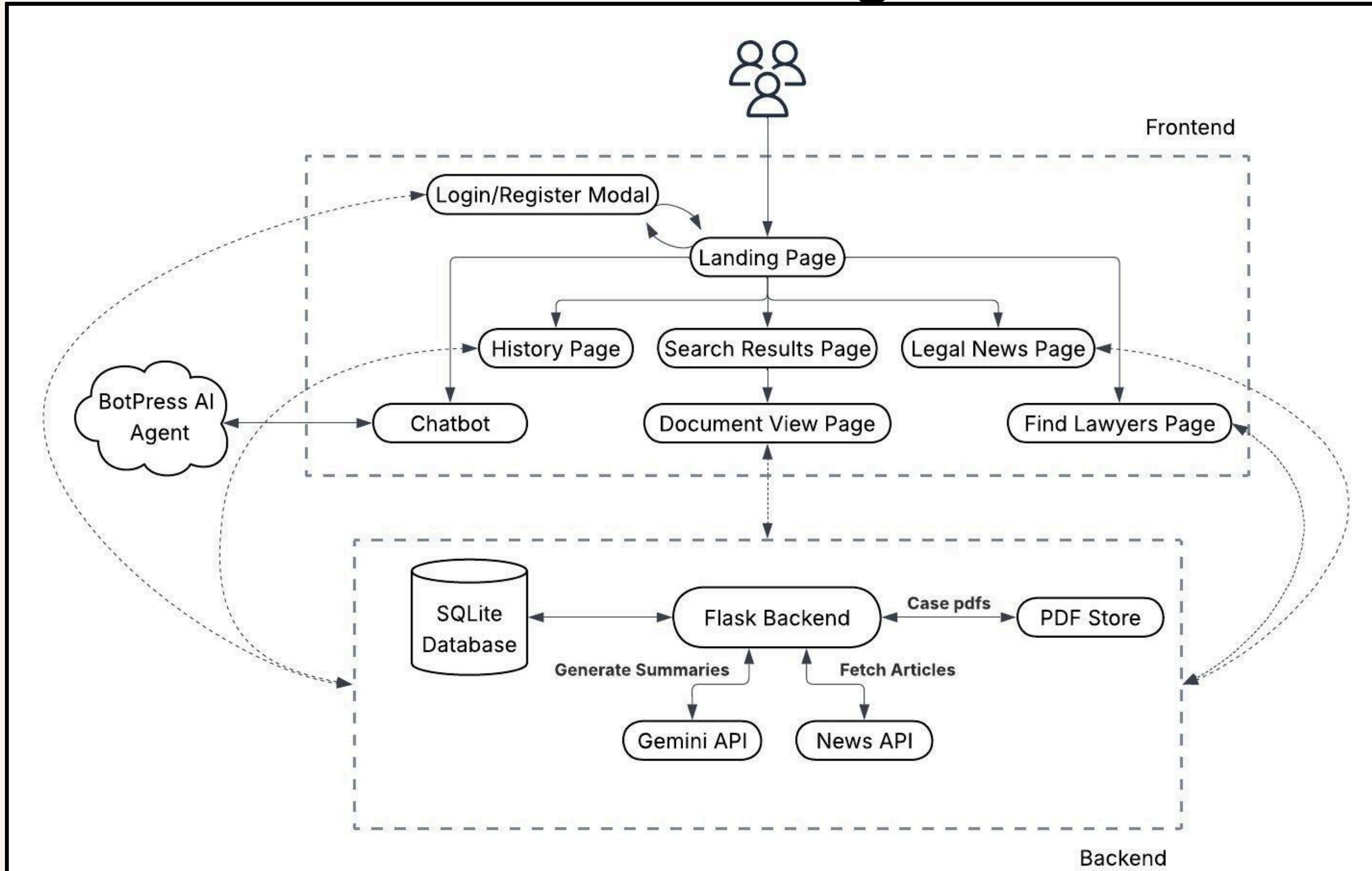
Scope

- **Semantic search:** Court Judgements are searched from the scraped judgements via semantic vector cosine similarity.
- **AI Summarization:** Summary of court judgements is generated via Gemini API.
- **Lawyers Search:** Various lawyers data is scraped and a lawyer can be searched based on name, city, their practice area, etc.
- **Real time legal news:** News article w.r.t. the legal world are displayed in real time, and the articles can be searched by keywords.
- **Legal Chatbot:** Chatbot capable of answering queries related to the Indian legal system, acts, sections, lawyer recommendations, legal situational queries, etc.

Limitations

- Dependence on Gemini API and News API quotas.
- Lack of multi-lingual support.
- Court Judgements cannot be scrapped in real-time due to heavy anti-scrape measures. Needs some level of human intervention.

Architecture Diagram



Tech Stack

- **Python**
- **Flask**: Web framework for building the application.
- **SQLite**: Database for storing user data, case information, and search history.
- **pandas**: Used for data manipulation, especially when loading initial datasets from CSV.
- **bcrypt**: For secure password hashing.
- **python-dotenv**: Manages environment variables for API keys and configurations.
- **requests**: HTTP client for making external API calls (e.g., to NewsAPI).
- **Gemini Flash & Embeddings**: Used for semantic search, case summarization, and extracting practice area keywords.
- **NumPy**: Essential for numerical operations, particularly with vector embeddings in search.
- **Flask-Session**: Manages server-side sessions.
- **concurrent**: For concurrent execution of tasks (e.g., generating multiple summaries).
- **Selenium**: For emulating browser to perform scraping.
- **BeautifulSoup4**: For extracting data from scraped webpages.
- **Frontend**: HTML, CSS, JS.
- **Botpress**: External AI chatbot integrated via JavaScript.

SQLite Database Schema

users	
email	TEXT
name	TEXT NN
pwd	TEXT NN
dob	TEXT

lawyers	
id	INTEGER
name	TEXT
url	TEXT
image_url	TEXT
city	TEXT
state	TEXT
address	TEXT
specialization	TEXT
experience	TEXT
rating	TEXT

cases	
id	INTEGER
case_title	TEXT NN
citation	TEXT
judges	TEXT
judgement_date	TEXT
case_id	TEXT
bench	TEXT
pdf_path	TEXT
snippet	TEXT
embedding	TEXT

history	
id	INTEGER
email	TEXT NN
query	TEXT NN
created_at	DATETIME

Team Contributions

Aditya Gaikwad

- Scraped 110 pdfs corresponding to 15 crime-related keywords from <https://scr.sci.gov.in/scrsearch/> and data of 50 lawyers from <https://lawrato.com/lawyers>.
- Create a chatbot using BotPress by providing it by a knowledge base enabling it to answer Indian legal queries.
- Implemented the Lawyer search interface and conducted the final phase of frontend refinement and optimization.

Madhur Thakkar

- Established the foundational Flask backend architecture and designed the SQLite database schema.
- Engineered the secure user authentication system using Bcrypt hashing and session management.
- Implemented the full-stack search history module for persistent user data tracking.
- Optimized performance by parallelizing AI summary generation using multithreading.

Monil Desai

- Architected the responsive UI/UX and executed the comprehensive frontend implementation.
- Engineered the core semantic search engine using cosine similarity on vector embeddings to enable context-aware retrieval.
- Integrated Google Gemini API to build a generative AI pipeline that produces instant, parallelized legal case summaries.
- Developed the real-time legal news module, incorporating external APIs with relevance sorting and custom search filtering.

Team Contributions - Key Commits

Member	
Madhur Thakkar	<code>feat: pagination and parallelized summary generation</code> <code>backend: setup users db, registration and login</code>
Monil Manish Desai	<code>feat: AI search and summary features</code> <code>frontend: history page and search result logic</code>
Aditya Gaikwad	<code>ui: major UI overhaul and lawyer scraping</code> <code>data: collecting supreme court cases via scraping</code>

Git Logs - Madhur Thakkar

Hash	Date	Message	Author
167a004	28-Sep-2025	Added flask code and required folder structure	madhur247
a9ab417	29-Sep-2025	Flask backend attached for scrapping dynamically	madhur247
1e885e2	29-Sep-2025	gitignore functioning correctly	madhur247
e494b10	19-Oct-2025	Added flask backend, users db setup and also registration and login functionality	madhur247
1e03ec6	02-Nov-2025	added backend for user history	madhur247
f8ea3c6	08-Nov-2025	Added recent history	madhur247
296720a	08-Nov-2025	testing	madhur247
f57c1f6	09-Nov-2025	added scraped data	madhur247
713e1b2	21-Nov-2025	pagination done and summary generation parallelized	madhur247

Git Logs - Monil Desai

Hash	Date	Message	Author
5acfb60	16-Oct-2025	frontend development by 25M0744	monil2003
d828192	19-Oct-2025	form setting	monil2003
48d0d83	22-Oct-2025	added databse functionality for search-result and doc-view :: backend complete (25M0744)	monil2003
9fb014c	24-Oct-2025	logout added	monil2003
04ecae8	07-Nov-2025	History Page added	monil2003
689d9c2	07-Nov-2025	search page updated	monil2003
414115b	08-Nov-2025	corrected tiny error	monil2003
238d613	20-Nov-2025	AI search and summary features	monil2003

Git Logs - Aditya Gaikwad

Hash	Date	Message	Author
3ed0763	06-Nov-2025	WIP: Collecting supreme court cases data via scraping.	Flameberry
fe0ee40	06-Nov-2025	Merge branch 'master' into scraping	Flameberry
22a4504	06-Nov-2025	Added a script to merge all scraped csv and added a snippet column to the data.	Flameberry
23b36ce	07-Nov-2025	Fixed a bug in the merging logic, which created spurious rows.	Flameberry
744d487	23-Nov-2025	Added lawyers scraping.	Flameberry
d110a2d	23-Nov-2025	Slightly updated UI.	Flameberry
c10ab44	23-Nov-2025	Major UI overhaul.	Flameberry
d03b897	23-Nov-2025	Scraped 50 lawyers.	Flameberry
5c80969	25-Nov-2025	Added chatbot made via botpress.	Flameberry
ccba06c	25-Nov-2025	Merge branch 'master' into aditya	Flameberry
af39e95	25-Nov-2025	Added chatbot option in all pages.	Flameberry
8145f28	25-Nov-2025	Fixed lawyer recommendation logic in the search results page.	Flameberry

Implementation Screenshots - Landing Page

The screenshot displays the landing page of LegalToday. At the top left is the logo "LegalToday". To its right are links for "Legal News" and "Find Lawyers". On the far right, there is a user profile placeholder with the name "madhur". The main heading "Legal Research Simplified ..." is centered above a descriptive subtitle: "Access a vast database of legal judgements from Supreme Court and access summarised views and search over them". Below this is a search bar with the placeholder text "Search using case titles, citations, keywords...". To the right of the search bar is a blue "Search" button. Further down, a section titled "Recent Searches:" lists three items: "Robbery", "robbery", and "state", each enclosed in a small button-like box.

LegalToday

Legal News Find Lawyers

madhur

Legal Research Simplified ...

Access a vast database of legal judgements from Supreme Court and access summarised views and search over them

Search using case titles, citations, keywords...

Search

Recent Searches:

Robbery

robbery

state

13

Implementation Screenshots - Search Results

LegalToday Legal News Find Lawyers madhur

Recommended Experts for your Case



Advocate Balaji T
★ 5.0

Anticipatory Bail, Cheque Bounce, Civil, Consumer Court, Court Marriage, Criminal,...

[View Profile](#)



Advocate Kriteka Sheokand
★ 5.0

Anticipatory Bail, Child Custody, Civil, Court Marriage, Criminal, Divorce, Domestic Violence,...

[View Profile](#)



Advocate Preeti Manderna
★ 5.0

Anticipatory Bail, Arbitration, Cheque Bounce, Child Custody, Civil, Consumer Court, Corporat...

[View Profile](#)

Search Results for "Money Laundering"

1 2 3 4 5 6 7 8 9 10 11 Next »

Pradeep Nirankarnath Sharma v. Directorate of Enforcement & Anr.

CRIMINAL APPEAL No. 1314/2025 • [2025] 4 S.C.R. 71 • [Hide Summary](#)

Prevention of Money Laundering Act, 2002 – The allegation against the appellant was that he was involved in financial transactions related to proceeds of crime, generated through fraudulent activities causing significant financial losses to the State – The appellant has contended that the alleged acts do not constitute an offence under the PMLA as the same was not in force during the relevant period, or the predicate offences as alleged were not included in the schedule to the PMLA at the relevant time and, therefore, cannot be subject to proceedings under the PMLA:Held: The contention is untenable – It is well established that offences under the PMLA are of a continuing nature, and the act of money

14

Implementation Screenshots - Explore Lawyers Page

LegalToday Legal News Find Lawyers madhur

Find Expert Lawyers

Search by name or specialization...

City...

Filter Results



Advocate Rajesh K.S
★ 4.7

📍 Subedar Chatram Road (SC Road), Bangalore / Bengaluru
⚖️ 19 years Experience

Practice: Anti Corruption, Anticipatory Bail, Arbitration, Bankruptcy / Insolvency, Breach of Contract, Cheque Bounce, Child Custody, [See More](#)

[View Profile](#)



Advocate Wg Cdr Ajit Kakkar (Retd)
★ 4.8

📍 Dwarka, Delhi
⚖️ 25 years Experience

Practice: Anticipatory Bail, Armed Forces Tribunal, Criminal, High Court, NRI, Pmla, Pocso Act, Supreme Court [See More](#)

[View Profile](#)



Advocate H Gouri Shankar
★ 4.9

📍 Banjara Hills, Hyderabad
⚖️ 29 years Experience

Practice: Anti Corruption, Arbitration, Banking / Finance, Cheque Bounce, Child Custody, Civil, Corporate, Criminal, Customs & Central Excise, [See More](#)

[View Profile](#)



Advocate Suksham Aggarwal



Advocate Rajeev Nigam
★ 4.9



Advocate Bala Janaki
★ 4.9

15

Implementation Screenshots - Legal News Feed

LegalToday Legal News Find Lawyers madhur

Legal News Feed

Latest updates on "blast"

blast Find News



The Indian Express
Ghaziabad bus blast case: 29 years an accused, 12 years a convict, and now a free man
How the case against Mohd Ilyas, charged in the 1996 Ghaziabad bus blast, was thrown out by High Court
2025-11-23T12:19:01Z [Read Full Article →](#)



The Indian Express
Shah Rukh Khan pays tribute to soldiers, civilians who lost their lives in terror attacks in Mumbai, Pahalgam and Delhi blast: 'Nothing can break the spirit...'
Shah Rukh Khan attended the Global Peace Honours in Mumbai to honour the lives lost in recent terror attacks across India.
2025-11-23T04:07:10Z [Read Full Article →](#)



The Indian Express
One handler behind Red Fort blast suspected to be based in Afghanistan, another linked to ISI
Digital footprints leading to Afghanistan, Pakistan, efforts on to determine real IDs, locations
2025-11-22T22:23:48Z [Read Full Article →](#)



Implementation Screenshots - Chatbot Conversation #1

The screenshot shows the LegalToday website interface on the left and a chatbot conversation on the right.

LegalToday Website:

- Header:** LegalToday, Legal News, Find Lawyers.
- Main Title:** Legal Research Simplified ...
- Text:** Access a vast database of legal judgements from Supreme Court and access sun views and search over them
- Search Bar:** Search using case titles, citations, keywords... (Search button)
- Recent Searches:** Money Laundering, Robbery, robbery, state

Chatbot Conversation:

- User (NyayBandhu):** Suggest me some indian lawyers with high rating for my divorce case.
- Bot:** Here are some highly rated Indian lawyers who specialize in divorce cases:
 - **Advocate Ankit Sharma:** Located in Gurgaon, with 13 years of experience and a rating of 4.8. [Profile](#).
 - **Advocate Sunil Kumar Bakshi:** Based in Faridabad, having 38 years of experience and a rating of 4.8. [Profile](#).
 - **Advocate Madhuri Bakshi:** Practicing in Faridabad, with 10 years of experience and a rating of 4.8. [Profile](#).
 - **Advocate Balaji T:** Located in Chennai, with 4 years of experience and a perfect rating of 5.0. [Profile](#).
 - **Advocate Jia Kapur:** Based in Delhi, with 7 years of experience and a rating of 4.7. [Profile](#).
- User (NyayBandhu):** What does article 370 state?
- Bot:** Article 370 of the Indian Constitution granted special autonomy to the State of Jammu and Kashmir.
- User:** Type your message... (Microphone icon)
- Bot:** by Botpress

Implementation Screenshots - Chatbot Conversation #2

The screenshot shows the LegalToday website interface. At the top, there's a navigation bar with 'LegalToday' (in purple), 'Legal News', and 'Find Lawyers'. Below this is a main header 'Legal Research Simplified ...' with a sub-instruction: 'Access a vast database of legal judgements from Supreme Court and access sun views and search over them'. A search bar below the header contains the placeholder 'Search using case titles, citations, keywords...'. To the right of the search bar is a blue button labeled 'Search'. Underneath the search bar, there's a section titled 'Recent Searches:' with four recent search terms: 'Money Laundering', 'Robbery', 'robbery', and 'state'. On the right side of the page, a chatbot window is open. The chatbot is identified by a blue circular icon with a white 'N' and the name 'NyayBandhu'. It has sent two messages: one explaining Article 370 of the Indian Constitution and another asking about the difference between Section 370 and Article 370. The user has responded with a question about Section 370. The chatbot has also responded with an explanation of the difference, mentioning that Article 370 pertains to constitutional provisions for a state, while Section 370 is a penal provision addressing human trafficking. At the bottom of the chat window, there's a text input field with the placeholder 'Type your message...' and a small '0' indicating no attachments. Below the input field is the text '⚡ by Botpress'. In the bottom right corner of the page, there's a large blue circular icon with a white 'P' and the number '18'.

LegalToday Legal News Find Lawyers

Legal Research Simplified ...

Access a vast database of legal judgements from Supreme Court and access sun views and search over them

Search using case titles, citations, keywords...

Recent Searches:

Money Laundering Robbery robbery state

N NyayBandhu

to have its own Constitution and limited the Indian Parliament's legislative powers in the region to matters agreed upon in the Instrument of Accession, such as defense, foreign affairs, and communications, unless the state's government consented to extend other laws.

what is section 370 and what is it's difference from the article 370?

Article 370 of the Indian Constitution provided special autonomy to the State of Jammu and Kashmir, allowing it to have its own Constitution and limiting the Indian Parliament's legislative powers in the region to certain matters. In contrast, Section 370 of the Indian Penal Code deals with the trafficking of persons and prescribes punishment for the same. The primary difference lies in their domains: Article 370 pertained to constitutional provisions for a state, while Section 370 is a penal provision addressing human trafficking.

Type your message... 0

⚡ by Botpress

18

Challenges

- **AI Latency:** Generating summaries for 10 results sequentially took 15+ seconds.
 - Implemented Python's ThreadPoolExecutor to fire 10 API requests in parallel. Reduced time to < 2s.
- **Frontend State:** History dropdown was closing before clicks registered due to 'blur' event.
 - Added a setTimeout delay in main.js to allow event propagation.
- **Data Consistency:** Scraping disparate legal sources resulted in duplicates and "spurious rows".
 - Engineered a robust Pandas cleaning script to sanitize and normalize dataset structure.
- **Scraping <https://scr.sci.gov.in/scrsearch/>:**
 - Because of various anti-scraping methods, this was the hardest part. The website not only had captchas, we cannot control the search results via the URL, we had to rely on user input, hence we used selenium to simulate user clicks.
 - PDF links weren't static and a temporary URL was assigned on every access of a PDF, hence via selenium we had to emulate clicking on the PDF URL to extract it.
 - When a PDF was opened in a new tab via selenium, the original tab redirected to the PDF, making it harder to return to the search results page. Hence we had to simulate back redirection via selenium.
 - The web page attributes were inconsistent hence attributes like Court Judgement title couldn't be easily extracted. We extracted the title via reading the PDF and applying a regex expression.

Future Scope

- Multi-lingual support can be added, as a big portion of the target audience (i.e. common Indian citizen with little to no expertise in law related areas) would prefer conversing in their native language.
- Optical character recognition for scanned legal documents, enabling easy book-keeping for legal documents.
- Chat with PDF functionality.