



OCR

RAG

LLMS

Langchain

SQL



HACK THE FUTURE 2025

# LEGACY HACKERS

An ultimate Legacy Data Processing Tool

>>>>

<<<<

# Our Team

Arunangshu Karrmakar

*2nd Year,  
Mathematics and  
Computing, IIT  
Kharagpur*

Akshat Jiwrajka

*2nd Year,  
Mathematics and  
Computing, IIT  
Kharagpur*

Malyadip Pal

*2nd Year,  
Instrumentation  
Engineering, IIT  
Kharagpur*

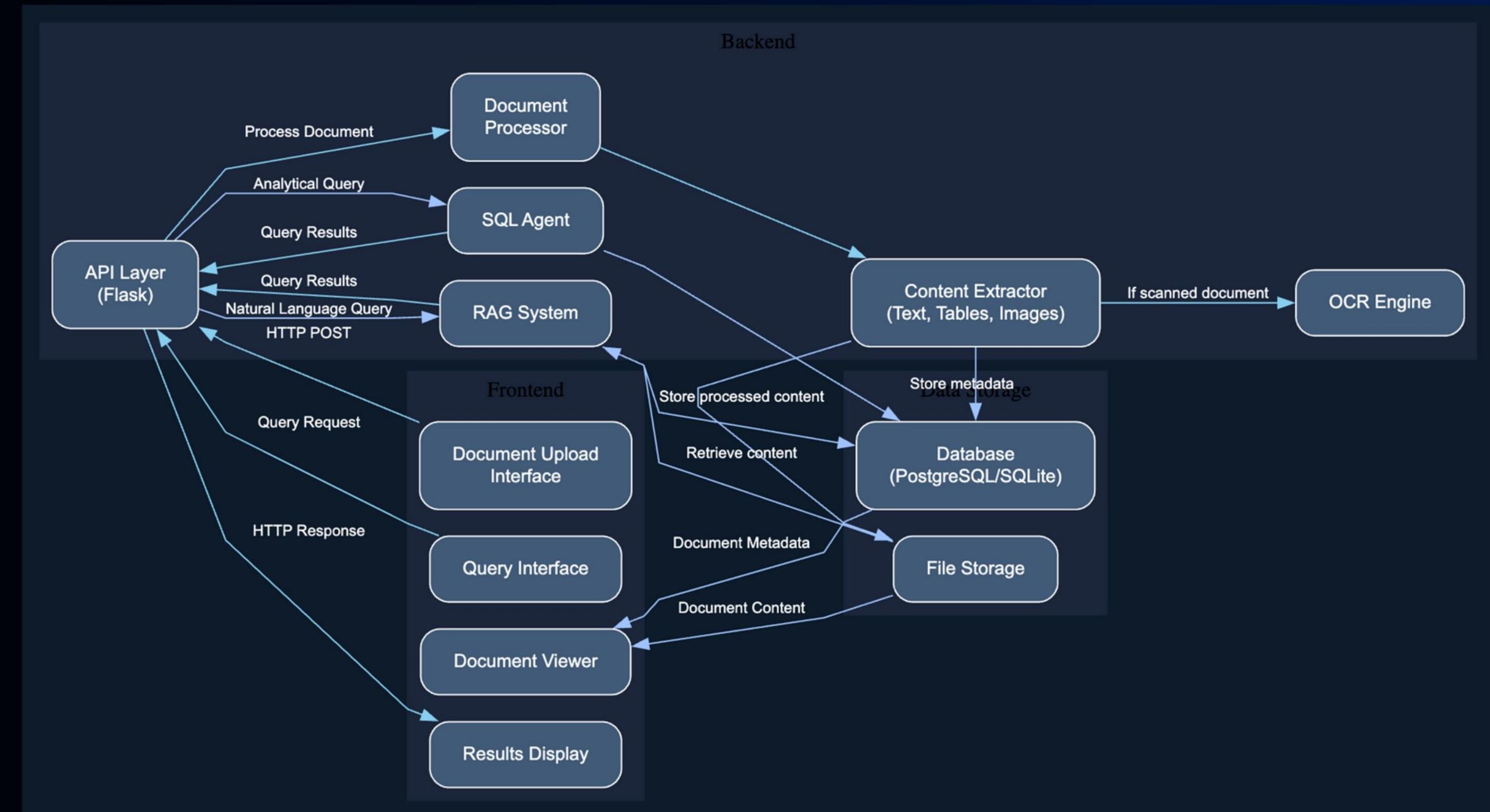
Aditya Debnath

*2nd Year,  
Instrumentation  
Engineering, IIT  
Kharagpur*

Shaunak Majumdar

*2nd Year,  
Mechanical  
Engineering, IIT  
Kharagpur*

# Our Solution : Pipeline Flowchart



# Our Solution : *Novelty and Uniqueness*

**Smart Table Detection:** Clips table boundaries using Microsoft Table-Transformer

**Agentic AI for Queries:** Intelligently finds the right table, generates SQL, executes, and replies in natural language.

**Seamless SQL conversion:** Transforms tables into SQL via markdown middle State .

**15 Stunning Graph Types:** Brings data to life with diverse and dynamic charting

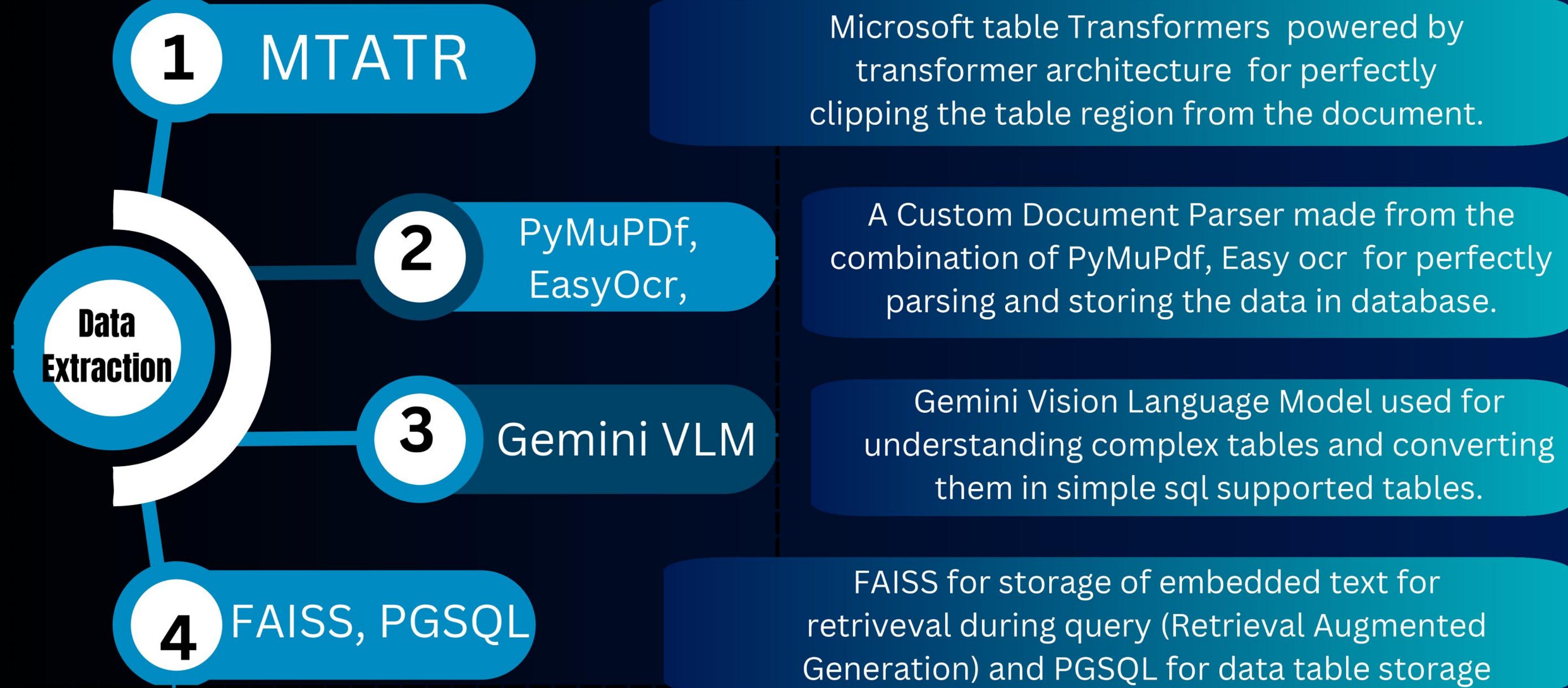
**Multilingual Mastery :** Supports 14 Indian languages and English for querying and graphs

**Zero-Code, Full Control:** User-friendly UI for uploading, managing, querying PDFs

**Blazing-Fast Processing:** with efficient multithreading (even with Free available LLMs)

**Unique PDF Tracking:** Assigns a distinctive Base36 timestamp to every PDF for easy reference.

# Technical Aspects and feasibility



# Technical Aspects and feasibility



# KEY FEATURES

Text, tables, and  
images  
Processing

Tabular data being  
stored in pgsql and  
accessible by user

SQL-based  
analytical queries in  
the analytics mode  
of the chatbot

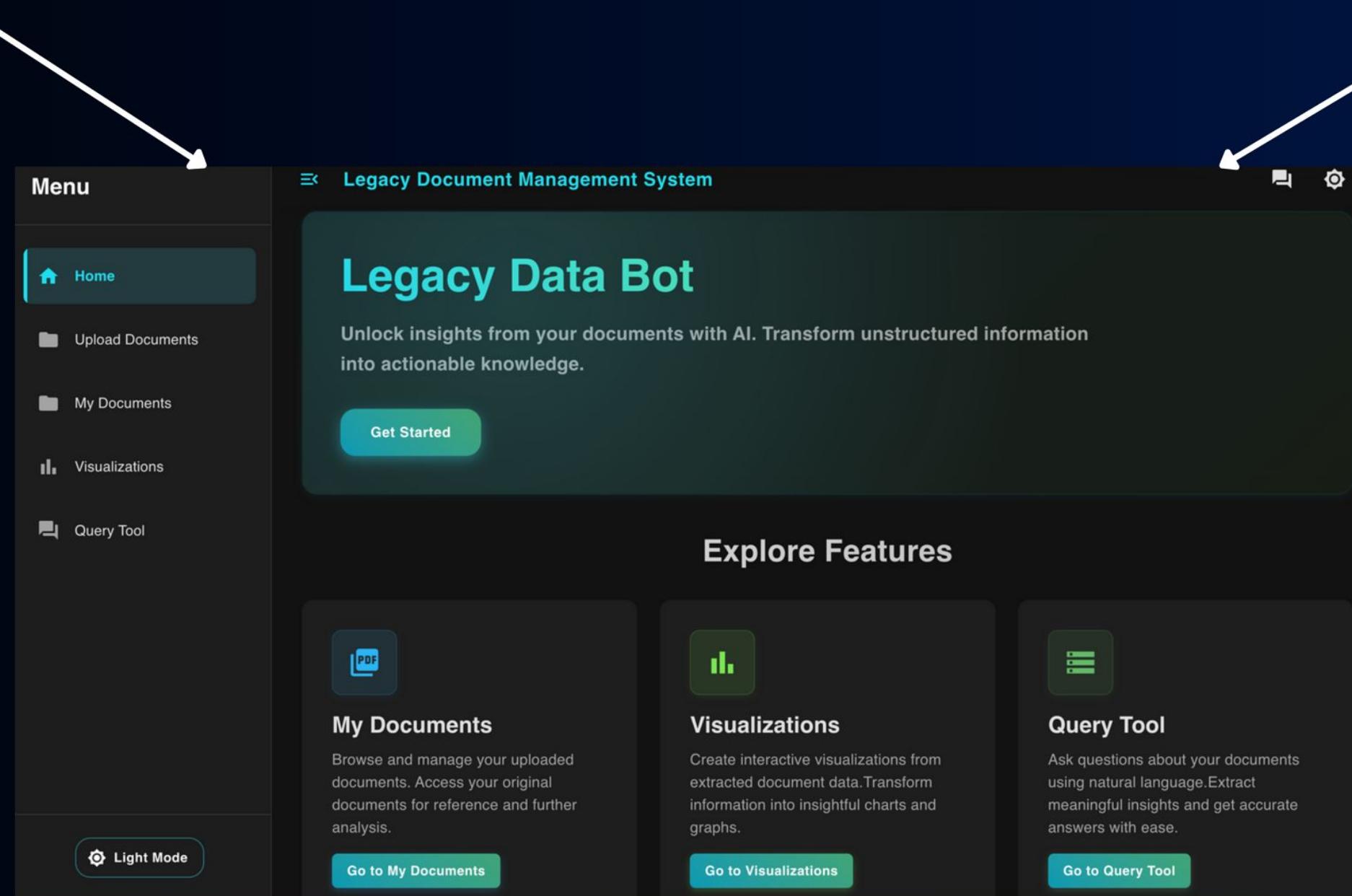
Interactive data  
visualization with just one  
command along with the  
key insights from the  
graph.

Allows users to query the  
data by text or voice using  
in multiple languages AI-  
powered chatbot  
powered by Gemini

User-friendly UI  
accessible to non-  
technical users

# BENEFITS OF OUR SOLUTION

Highly Interactive  
and User Friendly  
UI.



Provides extensive  
multi-lingual support  
during querying and  
visualization

Enhances  
speed by  
applying  
multithreading

Postgres SQL  
access which is  
free and  
accessible to  
everyone

# Screenshots of our UI

(Dark Mode)

Menu

-  Home
-  Upload Documents
-  My Documents
-  Visualizations
-  Query Tool

 Light Mode

Legacy Document Management System

# Legacy Data Bot

Unlock insights from your documents with AI. Transform unstructured information into actionable knowledge.

Get Started

## Explore Features



### My Documents

Browse and manage your uploaded documents. Access your original documents for reference and further analysis.

Go to My Documents



### Visualizations

Create interactive visualizations from extracted document data. Transform information into insightful charts and graphs.

Go to Visualizations



### Query Tool

Ask questions about your documents using natural language. Extract meaningful insights and get accurate answers with ease.

Go to Query Tool

# (Light mode)

Menu

- Home
- Upload Documents**
- My Documents
- Visualizations
- Query Tool

Dark Mode

Legacy Document Management System

## Upload Document

Upload PDF files for automated content extraction and analysis. The system will extract text, tables, and images from your documents.

hackthon\_rejuvenation\_legacy\_data.pdf  
18.79 MB

 **Upload**  **Cancel**

[Home](#)[Upload Documents](#)[My Documents](#)[Visualizations](#)[Query Tool](#)

# Upload Document

Upload PDF files for automated content extraction and analysis. The system will extract text, tables, and images from your documents.



hackthon\_rejuvenation\_legacy\_data.pdf

18.79 MB

Processing...

Cancel

Extracting content... 75%

Dark Mode



Home



Upload Documents



My Documents



Visualizations



Query Tool

### Create Visualization

Try these sample queries:

Bar Chart

Scatter Plot

Pie Chart

Line Chart

Histogram

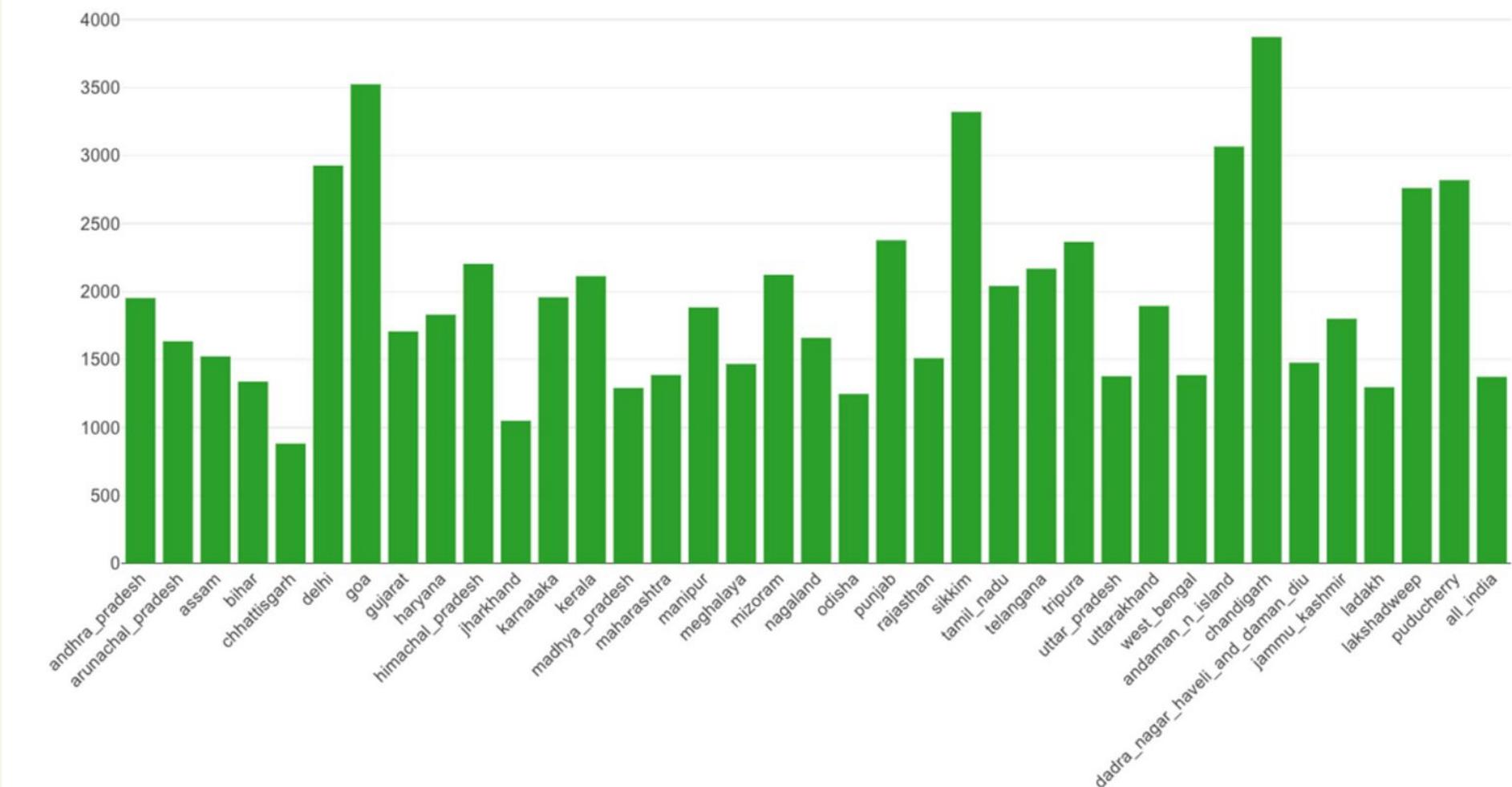
Enter your visualization query

Show me histogram of the data

Visualize

Plot type: bar, X: state\_ut\_all\_india, Y: c\_0\_5

### Show me histogram of the data



### Insights

Analysis of stk3p8\_p11\_1\_average\_mpce\_rural: The visualization shows the relationship between state\_ut\_all\_india and c\_0\_5. This can help understand patterns in your data.

Dark Mode

 Home Upload Documents My Documents Visualizations Query Tool

## My Documents

Browse and manage your uploaded documents and their extracted data.

### rejuvenation\_legacy\_data.pdf

Uploaded: 23/03/2025

Size: 18.8 MB

Status: Processed



[Home](#)[Upload Documents](#)[My Documents](#)[Visualizations](#)[Query Tool](#)

## Document Details

### Document Tables

Page 12

#### average\_mpce\_by\_state\_ut\_all\_india

state_ut_all_india	0_5	5_10	10_20	20_30	30_40	40_50	50_60	60_70	70_80	80_90	90_95	95_100
andhra_pradesh	2187	3017	3676	4297	4858	5451	6136	6911	7903	9806	12589	19784
arunachal_pradesh	2993	3758	4514	5478	6349	7120	7953	8923	10284	12418	15720	24188
assam	2149	2729	3212	3814	4355	4883	5497	6281	7326	9005	11548	17585
bihar	1739	2174	2585	3018	3425	3825	4266	4851	5677	6972	8749	13451
chhattisgarh	1350	1819	2207	2615	3042	3499	3949	4463	5232	6640	9134	14059
delhi	2540	3311	4008	4643	5236	5889	6636	7662	9361	12518	17634	28965
	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

< < 1 ... 11 12 13 ... 21 > >|



## Data Visualization Explorer

This tool helps you explore and visualize your database tables using natural language queries. Select a table and ask questions about your data!

Select a table to analyze

stk3p8\_p11\_1\_average\_mpce\_rural



### Table: stk3p8\_p11\_1\_average\_mpce\_rural

Total Rows

37

Total Columns

14

Memory Usage

8 KB

### Column Information

Column Name	Data Type	Non-null Count	Unique Values
state_ut_all_india	string	37	37
c_0_5	number	37	37
c_5_10	number	37	37
c_10_20	number	37	37
c_20_30	number	37	36
c_30_40	number	37	36



Home

Upload Documents

My Documents

Visualizations

Query Tool



### Assistant

Hello! I'm your document assistant. You can select documents by typing @ in the query box or choose a mode at the bottom left. Ask me anything about your documents!



#### Chat Mode

Ask questions about your documents



#### Analytics Mode

Query tables from your documents

Chat Mode: Type your message... Use @ to reference documents



Dark Mode



Home

Upload Documents

My Documents

Visualizations

Query Tool

**Assistant**

Hello! I'm your document assistant. You can select documents by typing @ in the query box or choose a mode at the bottom left. Ask me anything about your documents!

**Select Table References**

Please specify which tables to query. Each table is identified by its page number and table index within that page.

*Tables are stored with the format:*

*{timestamp}\_p{page\_number}\_{table\_index}\_{table\_name}*

Page Number  
p 11

Table Index  
1



+ Add Another Table

Cancel

Confirm

hackthon\_rejuvenation\_legacy\_data.pdf X

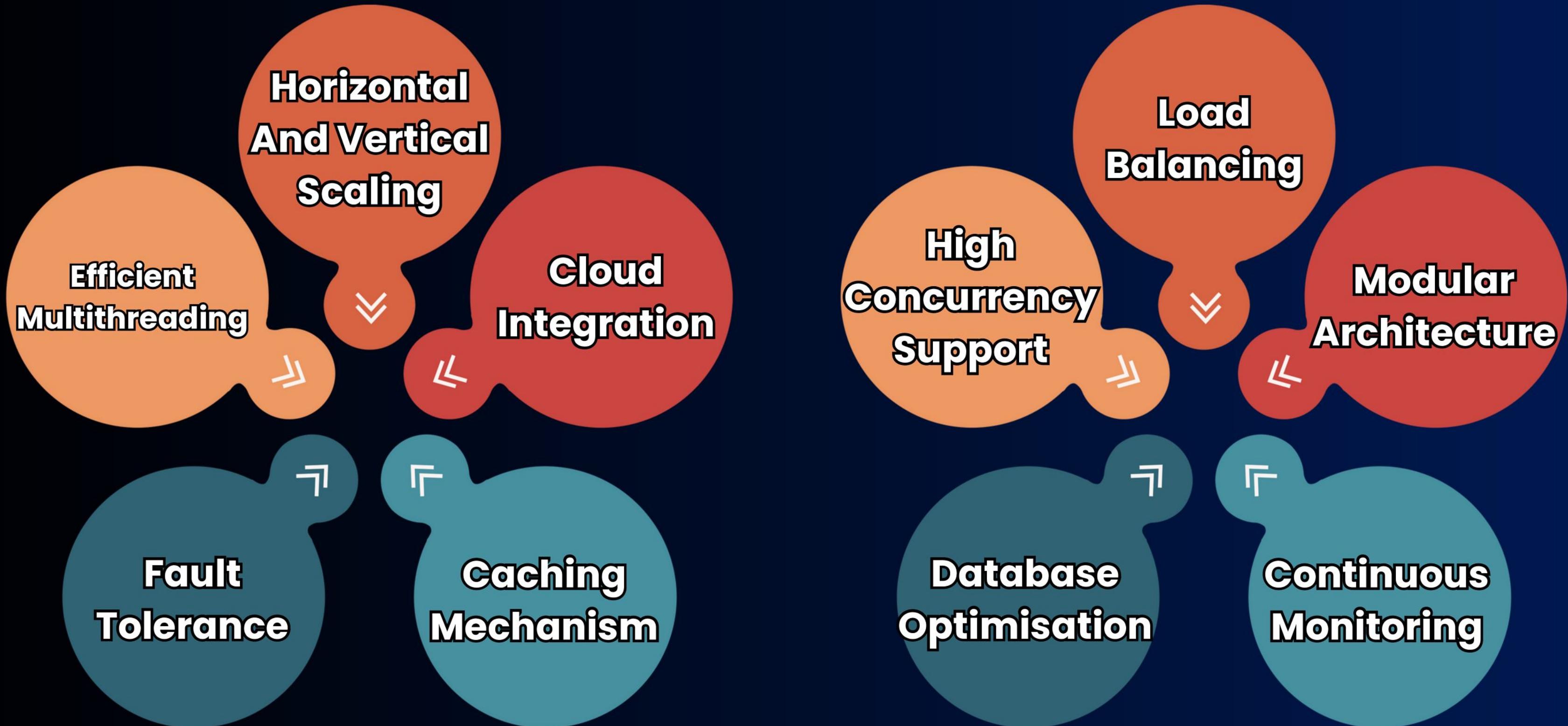
/



EN

Dark Mode

# Scalability of Solution



# Impact

Empowering Users

Enhanced Accessibility

Time Efficiency

Minimal User Effort

Business Value

Scalable Solution

Community Impact

Environmental Benefit

# Role of Team Members

- ◆ Arunangshu Karmakar - Pipeline, Data Extraction and Databasing
- ◆ Malyadip Pal - Data Extraction from files and Backend Integration
- ◆ Aditya Debnath - RAG implementation and SQL Agent
- ◆ Akshat Jiwrajka - Frontend implementation and Backend Development
- ◆ Shaunak Majumdar - Ideation of RAG and agent pipeline, SQL Visualization and Its UI

# Potential Future Plans

- ◆ We have several ideas to implement accurate extraction of tables spanning across several pages as a single table
- ◆ We have multiple ideas to improve our queries and accuracy by using paid LLM APIs like OpenAI which would greatly increase the quality of our models
- ◆ We have plans to make our system more personalised by making a sign up, login mechanism and deploying our system at a large scale using several cloud services

# THANK YOU

