

MULTI-MODAL RAG APPLICATION

Building Essence Towards Personalized Knowledge Model (PKM)

Team Members:

Ch. G. S. V. Chennaiah

! Problem Statement

Retrieving relevant images along with text for user queries is challenging due to traditional methods failing and language model hallucinations leading to misinformation.





Current Solution

Traditional RAG uses vector databases for text retrieval but LLMs often make errors and hallucinate. Visuals like images or tables help minimize these inaccuracies.

Table 4: State-wise details of expenditure incurred for conservation, preservation and environmental development of monuments of national importance for the years 2018-19, 2019-20 and 2020-21

Sl. No	State/UT	Number of Monuments	Expenditure (in crores)		Allocation (till Jan- 2021, in crores)
			2018-19	2019-20	
1	Uttar Pradesh	745	16.98	15.95	22.35
2					
3	Karnataka	506	21.14	28.55	21.1
4					
5	Odisha	80	5.9	5.3	2.75
6	West Bengal & Sikkim	139	9.68	8.1	6.71
7	Tamil Nadu & Puducherry	420	8.09	9.6	5.75
8	Punjab & Haryana	124	2.89	1.57	1
9	Himachal Pradesh	40	1.51	2.15	1.25
10	NCT Delhi	174	30.66	30.25	18.5
11	Goa	21	1.59	2.2	2.5
12	Assam	55	3.15	3.4	3
	Manipur	1			
	Meghalaya	8			
	Mizoram	1			
	Nagaland	4			
	Tripura	8			
	Arunachal Pradesh	3			
13	Rajasthan	163	7.05	10.28	10
14	Andhra Pradesh & Telangana	137	3.7	5.35	3
15	Telangana	--	5.29	5.15	4
16	Bihar	70	1.75	1.45	1
17	Jammu & Kashmir	69	2.44	1.83	1.2
18	Kerala	28	2.27	4.5	2.52
19	Gujarat & Daman and Diu	215	9.61	7	7.75
20	Uttarakhand	42	0.82	3.58	3.5
21	Chhattisgarh	47	2.09	3	2.6
22	Jharkhand	13	0.86	1.6	1
	Chemical Preservation	--	5.17	3.34	3.36
	Horticultural Activity	--	45.62	38.1	38.5
	DG office	--	190.57	202.23	68.81
	Total	405.64	428.96	266.4	

Source: Information provided in response to a parliamentary question in the Lok Sabha on 15 March, 2021

The answer to the question based on the table RAG architecture sometimes becomes hallucinated. Therefore, the availability of visual information is essential for a better retrieval process

Disclaimer: This project is built based on the attached PDF in the sidebar. If you wish to replicate it, please follow the tutorial links provided in the sidebar using your own resources.

Please feel free to ask me any questions related to the content of the PDF linked on the sidebar to your left.

Discuss the fiscal challenges mentioned for maintaining Monuments of National Importance and provide the budgetary allocation for Karnataka from the table?

Submit

Discuss the fiscal challenges mentioned for maintaining Monuments of National Importance and provide the budgetary allocation for Karnataka from the table?

The fiscal challenges for maintaining Monuments of National Importance include inadequate and geographically skewed expenditure, imbalance in allocation of funds among states, lack of utilization of revenue generated at monuments, and the absence of well-defined principles for selecting monuments. The budgetary allocation for Karnataka in the table is as follows:

- Expenditure (in crores):
 - 2018-19: ₹21.14 crores
 - 2019-20: ₹28.55 crores
 - 2020-21 (till Jan): ₹12.5 crores

⭐ Desired Solution

Multi-Modal RAG integrates text, images, and video frames to provide comprehensive answers, reducing LLM hallucinations and enhancing user trust and understanding.

The screenshot shows a web browser window with a dark theme. The address bar displays "multi-modal-rag-chabot.streamlit.app". The main content area has a dark background with white text. At the top, there is a search bar with the placeholder "Jump to...". Below it, a dropdown menu is open, showing "MMR-PDF" and "Monuments-of-National-Importance PDF". A message encourages users to "Wanna create your own Multi-Modal RAG using your PDF resource? 📄". Another message invites users to "Dive into our detailed Google Colab Notebook and make your project truly unique! ✨".

In the center, a question is displayed: "Discuss the fiscal challenges mentioned for maintaining Monuments of National Importance and provide the budgetary allocation for Karnataka from the table?"

Below the question, a text block states: "The text highlights the fiscal challenges in maintaining Monuments of National Importance in India, including inadequate budget allocation, geographical imbalance in distribution of funds, lack of revenue utilization from monument visits, and a lack of defined criteria for monument selection and preservation. Specifically, the budgetary allocation for Karnataka from the table is as follows:"

- For 2018-19: ₹21.14 crores
- For 2019-20: ₹28.55 crores
- For 2020-21 (till Jan): ₹10.28 crores.

A green banner at the bottom says "Relevant Image Found....". Below it is a table:

Sl. No	State/UT	Number of Monuments	Expenditure (in crores)		Allocation (till Jan- 2021, in crores)
			2018-19	2019-20	
1	Uttar Pradesh	745	16.98	15.95	22.35
2	Maharashtra	286	17.89	20.98	17.75
3	Karnataka	506	21.14	28.55	21.1
4	Madhya Pradesh	292	8.09	12.5	15
5	Odisha	80	5.9	5.3	2.75

Multi-Modal RAG retrieved relevant text and avail., image for avoiding hallucinations....

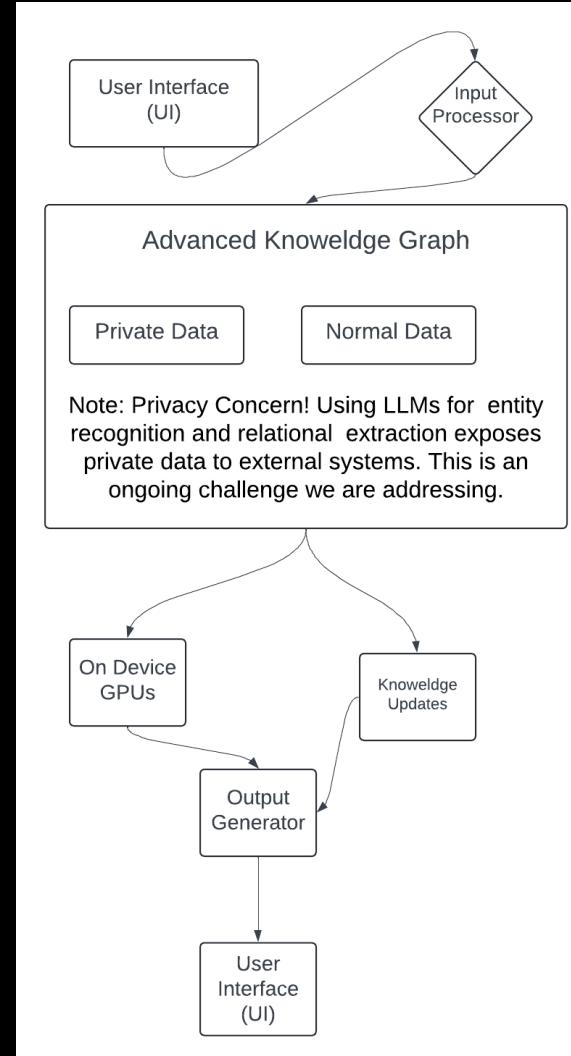


🎯 Current Focus

For the prototype, static resources were used to create a multimodal vector database, requiring fewer computational resources than dynamic methods. Detailed tutorials for creating your own database are provided in this [GitHub link](#).

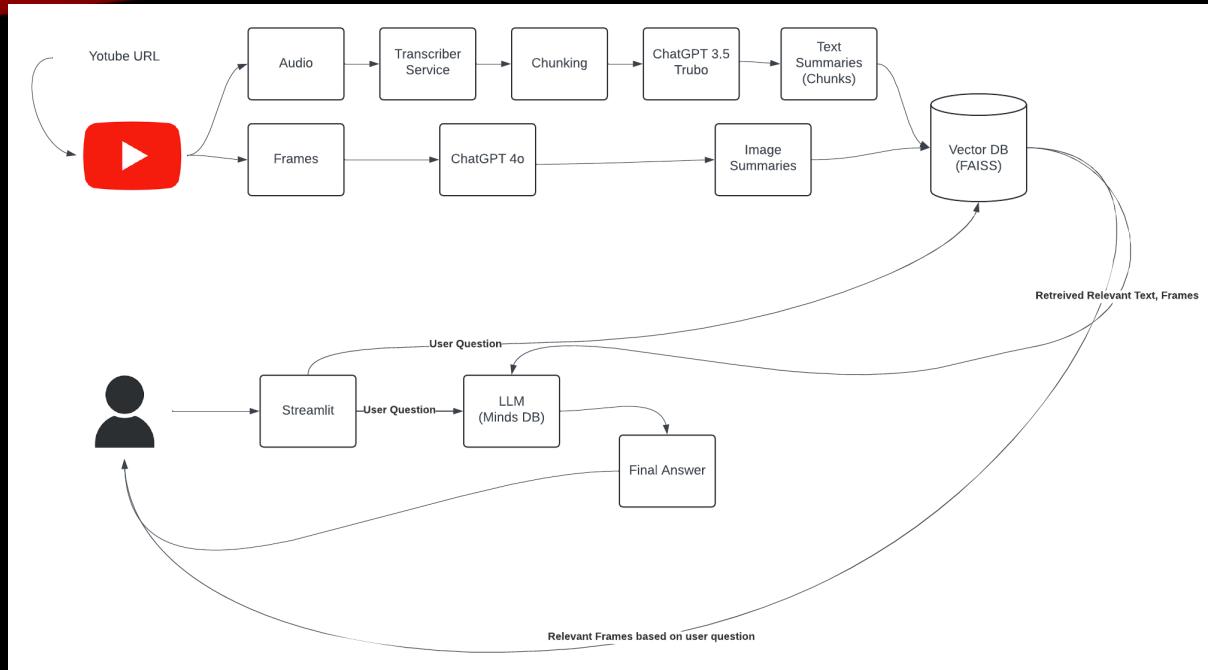
Architectures

- Basic PKM Architecture: Uses device GPUs for on-device knowledge graph building.



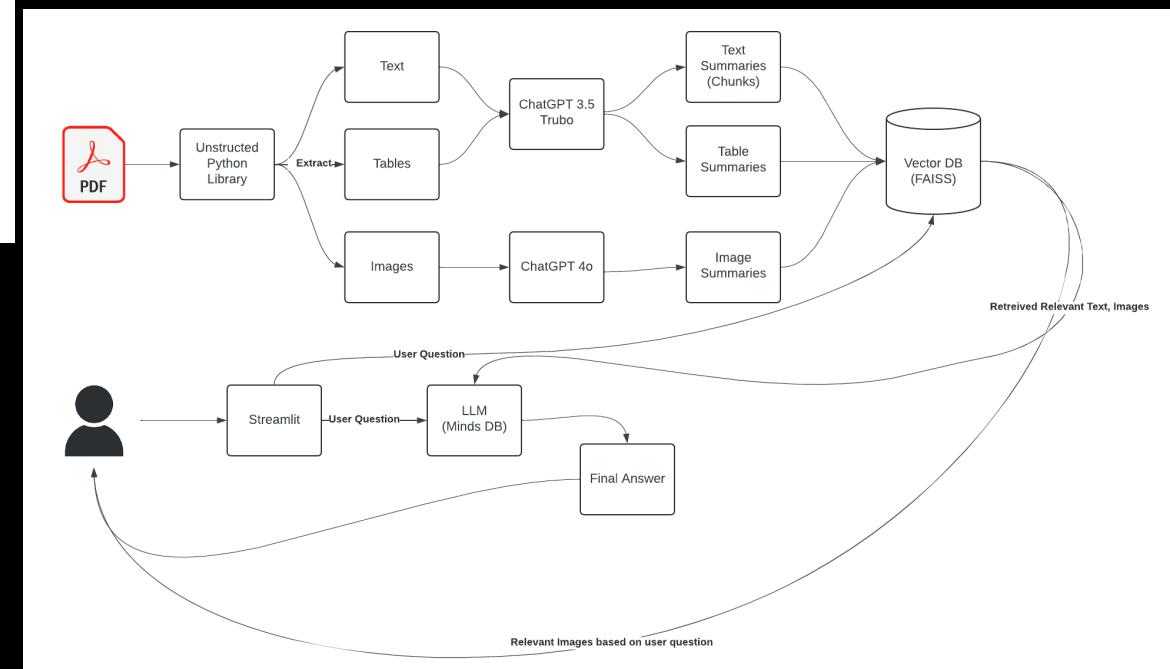
Architectures Cont.,

- MMR-PDF and MMR-Video Architectures: Static models for handling PDFs and video content.



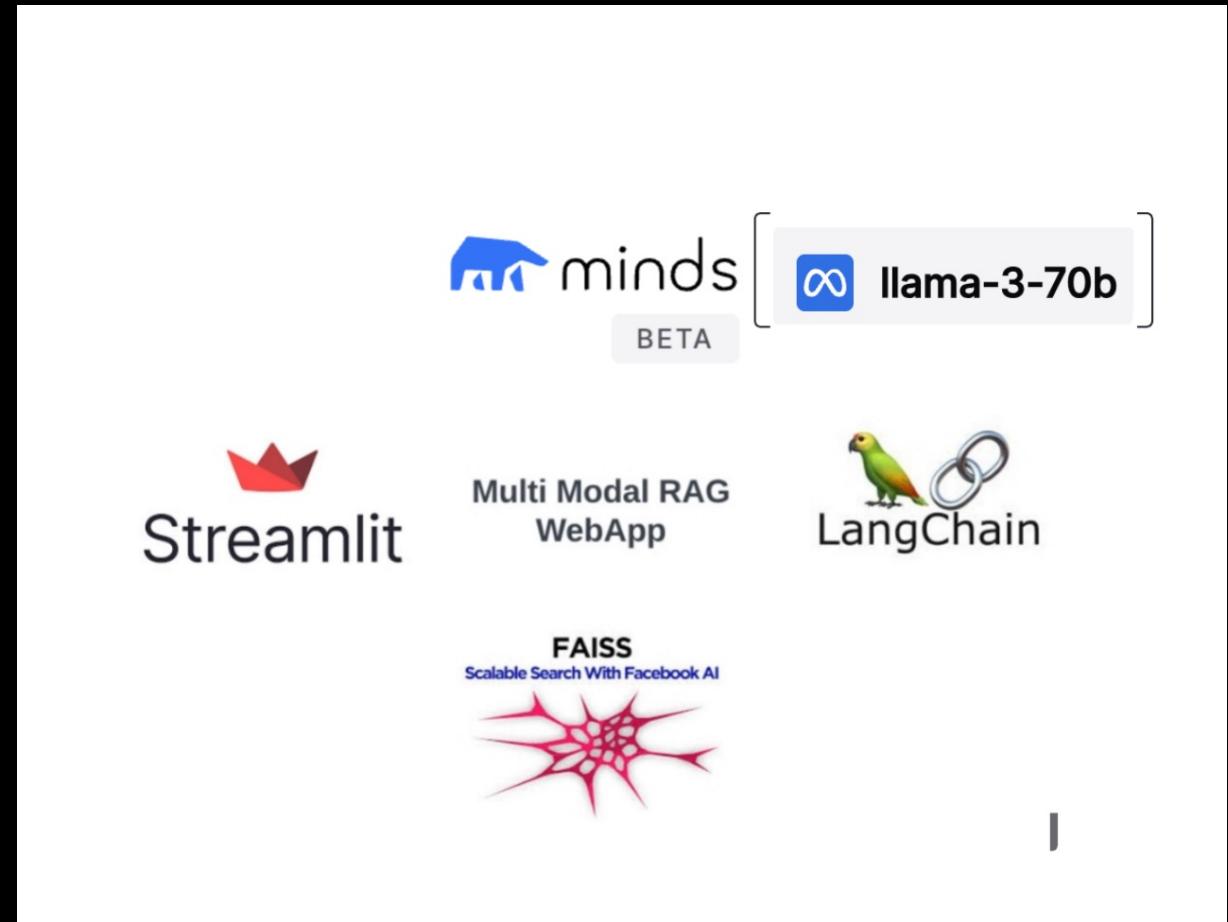
Multi - Modal RAG (Video – Static) ➡

➡ Multi – Modal RAG (Video – Static)



🔧 Tech Stack

- 🧠 MindsDB Endpoints (mdb.ai)
- 🧩 Langchain
- 💻 Streamlit
- 📁 FAISS



Demo

Live Prototype:- <https://multi-modal-rag-chabot.streamlit.app/>

The screenshot shows the Streamlit app interface. On the left sidebar, there are links for 'About' (with icons for mindsdb, Streamlit, FADS, LangChain, Meet the Author, LinkedIn, X, GitHub), 'Multi-Modal RAG WebApp' (with icons for Images, Tables, Text), and 'Multimodal Embedding' (with icons for Documents, Images, Tables, Text). The main content area features a pink header with the title 'Multi-Modal RAG WebApp'. Below it is a diagram illustrating the system architecture. It shows 'Documents' being processed into 'Images', 'Tables', and 'Text'. These are then converted into 'Multimodal Embedding' (Image Summary, Table Summary, Text Summary) and 'Text Embedding' (Image Summary, Table Summary, Text Summary). These embeddings are fed into a 'Multi-Vector Retriever' (represented by a cylinder). The retriever outputs three options for generating an 'Answer':

- Option 1: Retrieve raw image from a 'Raw Table & Text' database.
- Option 2: Retrieve image summary from a 'Raw Table & Text' database.
- Option 3: Retrieve image summary but pass raw image to LLM for synthesis.

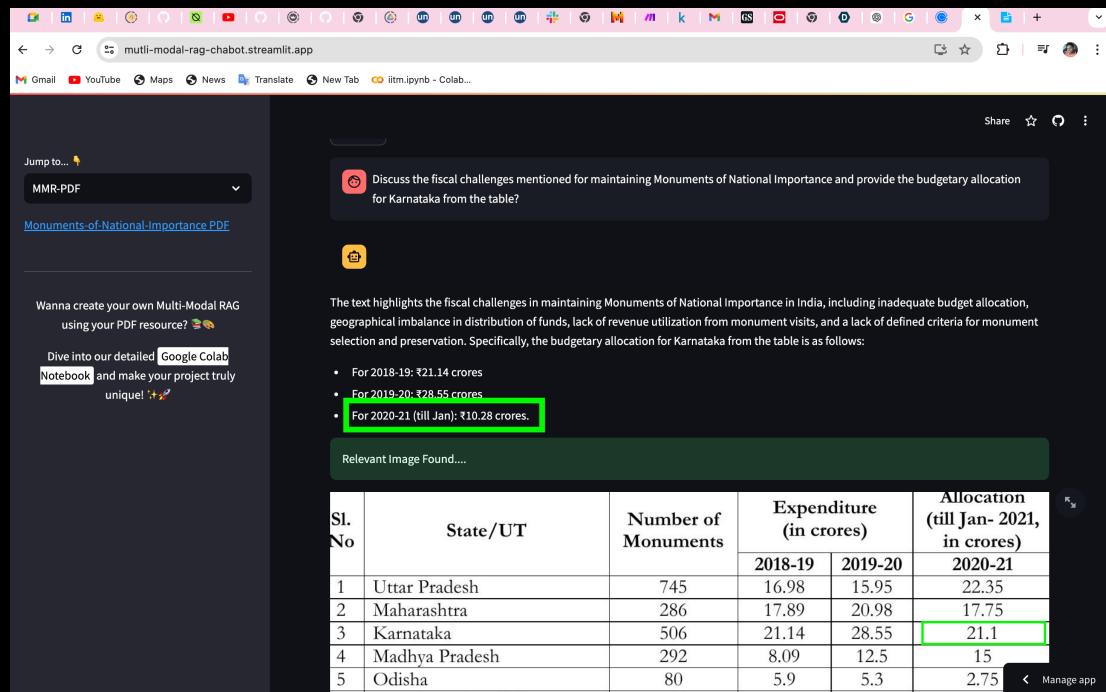
A yellow hand icon points to the 'Multi-Vector Retriever' component.

Live Web Application Home Page

The screenshot shows the Streamlit app interface. The sidebar on the left lists 'Jump to...' (MMR-PDF), 'Monuments-of-National-Importance PDF', 'Wanna create your own Multi-Modal RAG using your PDF resource?', 'Dive into our detailed Google Colab Notebook and make your project truly unique!', and 'Submit'. The main content area has a dark background with white text. It displays a question from a user asking about the fiscal challenges for maintaining Monuments of National Importance in Karnataka. A text box below provides a detailed answer, mentioning inadequate budget allocation, geographical imbalance in distribution of funds, lack of revenue utilization from monument visits, and a lack of defined criteria for monument selection and preservation. It also includes a note about the budgetary allocation for Karnataka from the table.

User asking a question from the attached pdf related to the text and table data

Demo Contd.,



Discuss the fiscal challenges mentioned for maintaining Monuments of National Importance and provide the budgetary allocation for Karnataka from the table?

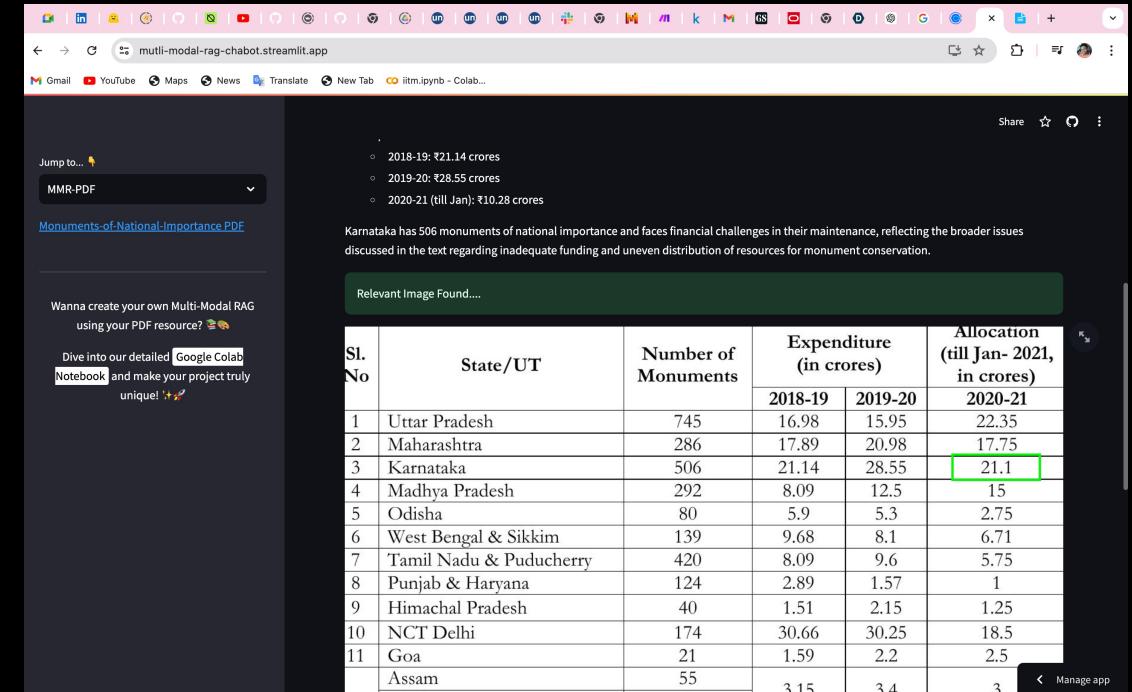
The text highlights the fiscal challenges in maintaining Monuments of National Importance in India, including inadequate budget allocation, geographical imbalance in distribution of funds, lack of revenue utilization from monument visits, and a lack of defined criteria for monument selection and preservation. Specifically, the budgetary allocation for Karnataka from the table is as follows:

- For 2018-19: ₹21.14 crores
- For 2019-20: ₹28.55 crores
- For 2020-21 (till Jan): ₹10.28 crores.

Relevant Image Found....

Sl. No	State/UT	Number of Monuments	Expenditure (in crores)		Allocation (till Jan- 2021, in crores)
			2018-19	2019-20	
1	Uttar Pradesh	745	16.98	15.95	22.35
2	Maharashtra	286	17.89	20.98	17.75
3	Karnataka	506	21.14	28.55	21.1
4	Madhya Pradesh	292	8.09	12.5	15
5	Odisha	80	5.9	5.3	2.75

As I mentioned, LLM hallucinates in the answer



2018-19: ₹21.14 crores
2019-20: ₹28.55 crores
2020-21 (till Jan): ₹10.28 crores

Karnataka has 506 monuments of national importance and faces financial challenges in their maintenance, reflecting the broader issues discussed in the text regarding inadequate funding and uneven distribution of resources for monument conservation.

Relevant Image Found....

Sl. No	State/UT	Number of Monuments	Expenditure (in crores)		Allocation (till Jan- 2021, in crores)
			2018-19	2019-20	
1	Uttar Pradesh	745	16.98	15.95	22.35
2	Maharashtra	286	17.89	20.98	17.75
3	Karnataka	506	21.14	28.55	21.1
4	Madhya Pradesh	292	8.09	12.5	15
5	Odisha	80	5.9	5.3	2.75
6	West Bengal & Sikkim	139	9.68	8.1	6.71
7	Tamil Nadu & Puducherry	420	8.09	9.6	5.75
8	Punjab & Haryana	124	2.89	1.57	1
9	Himachal Pradesh	40	1.51	2.15	1.25
10	NCT Delhi	174	30.66	30.25	18.5
11	Goa	21	1.59	2.2	2.5
	Assam	55	3.15	3.4	3

Clear cut information from the retrieved image (Power of Multi Modal RAG)

Demo Contd.,

The screenshot shows a web browser window with the URL multi-modal-rag-chabot.streamlit.app. The main content area features a title "Multi-Modal RAG ChatBot (Video)" with two small robot icons. Below it is a "Disclaimer" box stating: "This project is built based on the attached Video in the sidebar. If you wish to replicate it, please follow the tutorial links provided in the sidebar using your own resources." A yellow hand icon points from the left towards the "Disclaimer" box. The sidebar on the left contains sections for "MMR-Video", "Zero-Shot Prompting...", and "Wanna create your own Multi-Modal RAG using your Video resource?". The main input field asks, "According to the author, what is meant by zero-shot prompting?", and the "Submit" button is highlighted with a red border. At the bottom, there is a detailed explanation of zero-shot prompting.

User asking a question from attached YouTube video

The screenshot shows the same web browser window after the user has submitted their question. A green box highlights the "Relevant Frame Found..." section, which displays the text: "Zero-shot prompting is the simplest prompting technique used for chatbots. It involves asking the chatbot to answer a question or perform a task without providing any examples. This technique relies on clear instructions or questions without the need for specific examples. While zero-shot prompting is easy to implement, it may not yield optimal results for complex tasks as it requires careful crafting of prompts to achieve desired outcomes. Overall, zero-shot prompting offers advantages such as minimal effort, accessibility for everyone, and versatility in various tasks, but it may have limitations in terms of performance and prompt dependence." Below this, another green box highlights the retrieved answer: "It involves asking the chatbot to answer a question or to do a task without giving any examples". At the bottom right, there is a signature "Kalyan KS @ Akmmus AI Labs".

Web Application retrieves the relevant information and relevant frame image for better answer

Business Value

- Multi-modal chatbots aid decision-makers, developers, and researchers by providing quick summaries, answers, and visual explanations from large sources.
- Currently using static resources, but planning to develop an on-device vector database using users' device resources.
- Dedicated app in development for easy attachment of data sources and notifications when the multi-modal vector database is created.
- Revenue model involves collecting a platform fee and charging users for training resources.
- Potential to revolutionize the way users interact with data by providing comprehensive and trustworthy answers to user queries.



🔗 Links

- GitHub:- <https://github.com/chakka-guna-sekhar-venkata-chennaiah/Mutli-Modal-RAG-ChaBot-lablab.ai-llama3-hackathon>
- Live Prototype:- <https://mutli-modal-rag-chabot-lablabai-llama3-hackathon.streamlit.app/>



OO Future Vision

- Developing a dedicated mobile app for integrating data from multiple resources, enabling users to easily attach and manage their data sources.
- Creating a multi-modal vector database using on-device resources, enhancing privacy and reducing reliance on cloud services. The app will notify users when the vector database is created and ready for use.
- Expanding the vector database on a regular basis (weekly, daily, or monthly) based on user interest, ensuring that the application remains up-to-date and relevant to users' needs. This will enable users to ask questions from an ever-growing knowledge base, making the application an indispensable tool for decision-makers, developers, and researchers.
- Continuously improving the user experience and functionality of the application based on user feedback and emerging technologies.



THANK YOU