

Agentic RAG Chatbot For Multi- Format Document QA Using MCP

A multi-turn document question-answering system supporting PDF, DOCX, PPTX, TXT, and CSV formats. Built using Google Gemini, FAISS, LangChain, and the Model Context Protocol (MCP).



Modular Agent Architecture



IngestionAgent

Parses text and images from documents using PyMuPDF, python-docx, pptx, and OCR.



RetrievalAgent

Searches for relevant chunks using FAISS vector store with OpenAI or SBERT embeddings.



LLMResponseAgent

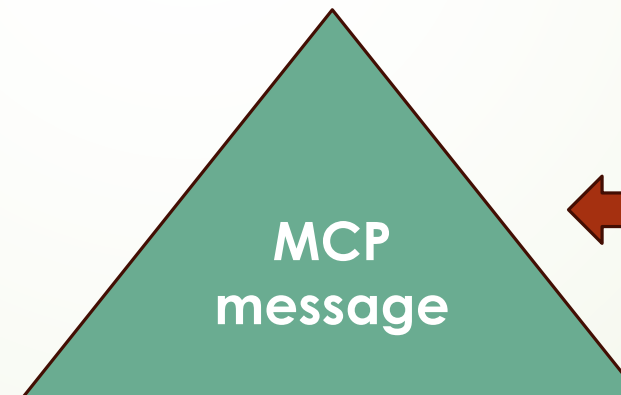
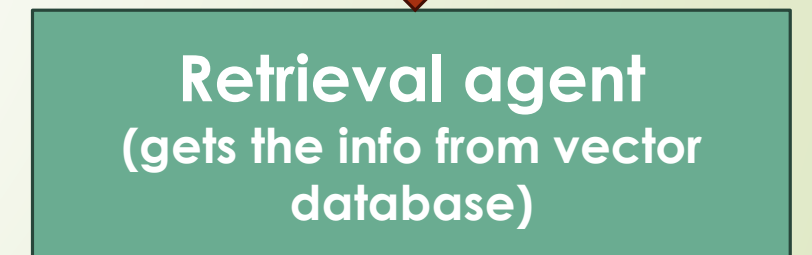
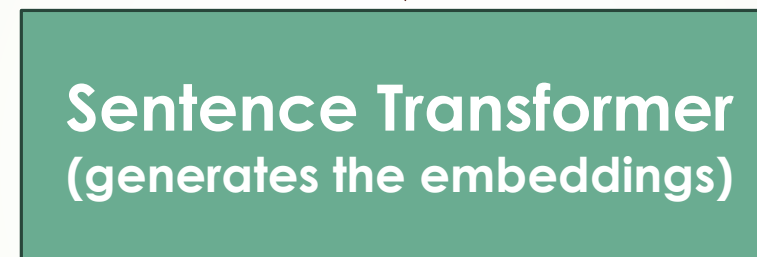
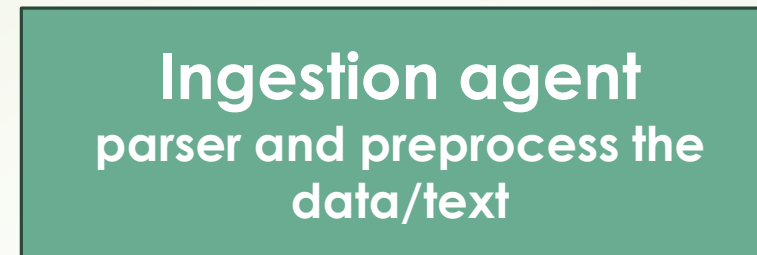
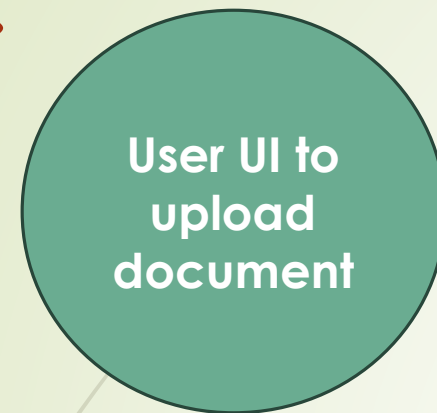
Uses Google Gemini Pro via LangChain to generate answers.



MCP Protocol

Enables message passing between agents for multi-turn dialogue handling.

End to End System flow with MCP Message



Tech Stack

Frontend:

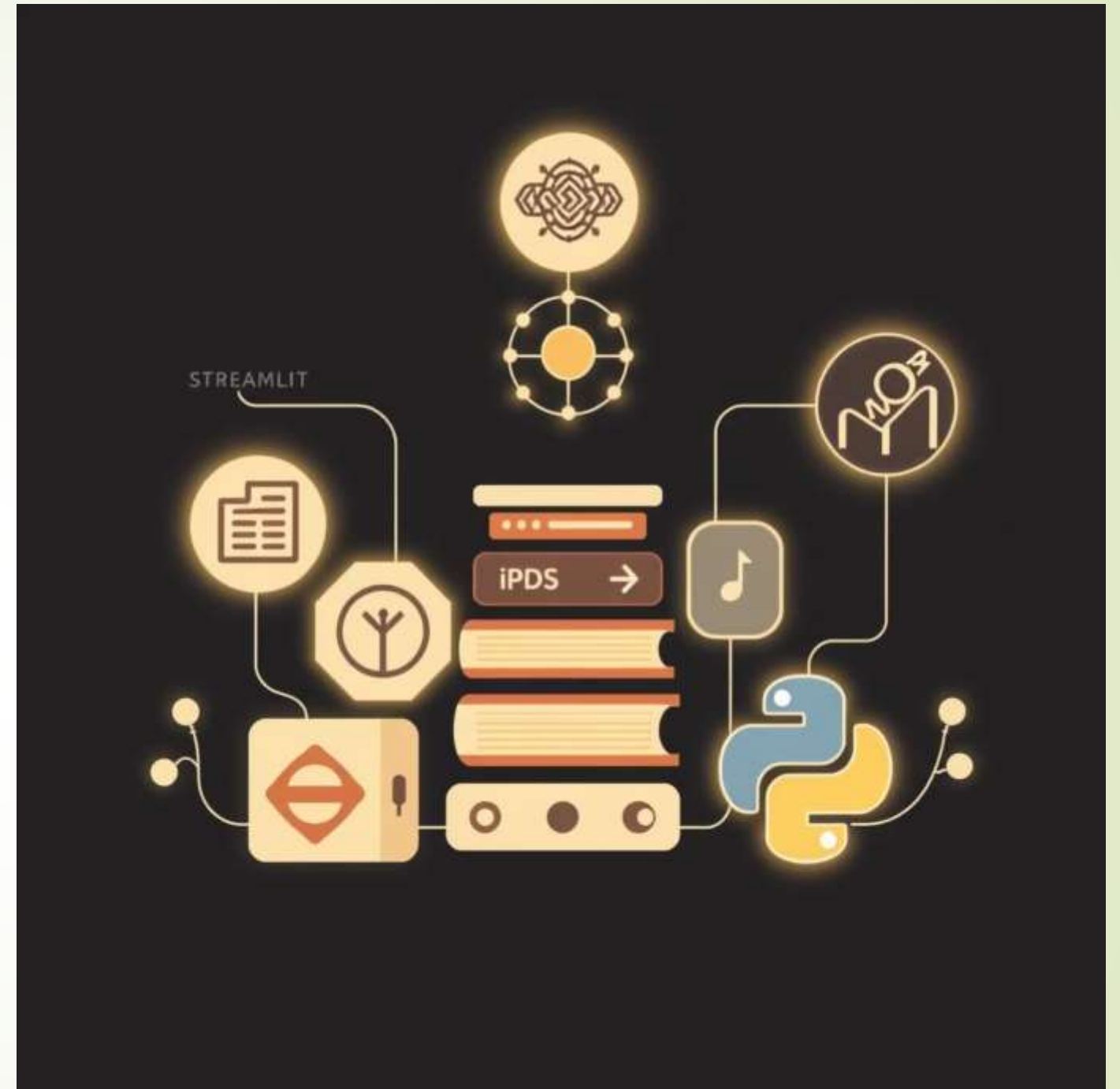
- Streamlit (UI, File Upload, Interaction)

Backend:

- LangChain (Agent Management, Prompt Handling)
- Google Gemini Pro (LLM API)
- FAISS (Vector Similarity Search)
- PyMuPDF, python-docx, pptx (Parsing)
- pytesseract + Pillow (OCR)

Infrastructure:

- Python 3.10+, pip virtual environment
- MCP (Custom Protocol)



UI Screenshots

Agentic RAG Multi-Document Chatbot with MCP

Upload documents (PDF, PPTX, DOCX, TXT, CSV)

Drag and drop files here

Limit 200MB per file • PDF, PPTX, DOCX, TXT, CSV

Browse files

X

Submit and Process Documents

Ask a question from the uploaded documents

explain last question

[Ask Question](#)

Answer:

Based on the document provided, the last question is Q. 32. Here is an explanation of how to solve it.

Question Q. 32

The question asks you to find the value of the following limit:

Step-by-Step Explanation:

Deploy :

1. **Identify the Indeterminate Form:** If you try to substitute $x = a$ directly into the expression, you get $\sqrt{a^2 + a} - a$, which simplifies to $a - a$. This is an "indeterminate form," meaning you cannot determine the limit without rewriting the expression first.
2. **Use the Conjugate Method:** A common technique for limits involving square roots is to multiply the expression by its conjugate. The conjugate of $\sqrt{t^2 + t} - t$ is $\sqrt{t^2 + t} + t$. To keep the expression's value the same, you must multiply and divide by the conjugate:

3. **Simplify the Numerator:** The numerator is now in the form $(a - b)(a + b)$, which equals $a^2 - b^2$.
- Let $a = \sqrt{t^2 + t}$, so $a^2 = t^2 + t$.
 - Let $b = t$, so $b^2 = t^2$.
 - The numerator becomes $(t^2 + t) - t^2 = t$.
4. **Rewrite the Expression:** After simplifying the numerator, the new expression for the limit is: $\lim_{t \rightarrow \infty} \frac{t}{t \cdot \sqrt{(t^2 + t) + t}}$
5. **Divide by the Highest Power of t :** If you substitute $t = \infty$ now, you get $\frac{\infty}{\infty}$, which is another indeterminate form. To solve this, divide both the numerator and the denominator by the highest power of t in the denominator, which is t .

- Numerator: $x / x = 1$

- Denominator: $(\sqrt{t^2 + t} + t) / t = \sqrt{t^2 + t}/t + t/t$

To simplify $\sqrt{(t^2 + 1)/t}$, you can bring t inside the square root as t^2 (since t is positive): $\sqrt{(t^2 + 1)/t} = \sqrt{\{(t^2 + 1)/t^2\}} = \sqrt{\{t^2/t^2 + t/t^2\}} = \sqrt{1 + 1/t}$

So, the denominator becomes: $\sqrt{(1 + 1/n)} + 1$

6. Evaluate the Final Limit: The expression is now: $\lim_{t \rightarrow \infty} (as \ t \rightarrow \infty) [-1 / (\sqrt{1 + 1/t}) + 1]$

Now, substitute $x = \infty$. As x approaches infinity, the term $1/x$ approaches 0. $\frac{1}{\sqrt{4(1 + 0) + 1}} = \frac{1}{\sqrt{4(1) + 1}} = \frac{1}{\sqrt{4 + 1}} = \frac{1}{\sqrt{5}} = 1/\sqrt{5}$

7. Round to One Decimal Place: The value of the limit is $1/2$, which is 0.5 . This is already in the required format.

Answer: The value of the limit is 0.5.

📌 **Source Chunks:**

- drawn from the bag one at a time with replacement. Let S_n denote the total number of black balls drawn in the experiment. The expectation of S_{100} denoted by $E[S_{100}] = \lfloor \text{Round off to one decimal place} \rfloor Q.62$ Consider the following tables, Loan and Borrower, of a bank. Loan loan num branch name amount L11 Banjara Hills 90000 L14 Kondapur 50000 L15 SR Nagar 40000 L22 SR Nagar 25000 L23 Balanagar 80000 L25 Kondapur 70000 L19 SR Nagar 65000 Borrower customer name

- **Numerator:** $t / t = 1$
- **Denominator:** $(\sqrt{t^2 + t} + t) / t = \sqrt{t^2 + t}/t + t/t$

To simplify $\sqrt{t^2 + t}/t$, you can bring t inside the square root as t^2 (since t is positive): $\sqrt{t^2 + t}/t = \sqrt{(t^2 + t)/t^2} = \sqrt{t^2/t^2 + t/t^2} = \sqrt{1 + 1/t}$

So, the denominator becomes: $\sqrt{1 + 1/t} + 1$

6. **Evaluate the Final Limit:** The expression is now: $\lim_{(as\ t \rightarrow +\infty)} [1 / (\sqrt{1 + 1/t} + 1)]$

Now, substitute $t = \infty$. As t approaches infinity, the term $1/t$ approaches 0 . $1 / (\sqrt{1 + 0} + 1) = 1 / (\sqrt{1 + 1} + 1) = 1 / (1 + 1) = 1/2$

7. **Round to One Decimal Place:** The value of the limit is $1/2$, which is 0.5 . This is already in the required format.

Answer: The value of the limit is 0.5.

Source Chunks:

- drawn from the bag one at a time with replacement. Let S_n denote the total number of black balls drawn in the experiment. The expectation of S_{100} denoted by $E[S_{100}] =$ (Round off to one decimal place) Q. 62 Consider the following tables, Loan and Borrower, of a bank. Loan loan num branch name amount L11 Banjara Hills 90000 L14 Kondapur 50000 L15 SR Nagar 40000 L22 SR Nagar 25000 L23 Balanagar 80000 L25 Kondapur 70000 L19 SR Nagar 65000 Borrower customer name loan num Anand L11 Kartek L11
- GATE 2025 Data Science & Artificial Intelligence - DA Q. 14 Let $f(x) = e^x - e^{-x^2}$, $x \in \mathbb{R}$. Let $f^{(k)}(a)$ denote the k th derivative of f evaluated at a . What is the value of $f^{(10)}(0)$? (Note: $!$ denotes factorial) (A) 0 (B) 1 (C) $1/10!$ (D) $2/10!$ Q. 15 Let p and q be any two propositions. Consider the following propositional statements. $S_1 : p \rightarrow q$, $S_2 : \neg p \wedge q$, $S_3 : \neg p \vee q$, $S_4 : \neg p \vee \neg q$, where \wedge denotes conjunction (AND operation), \vee denotes disjunction (OR operation),
- GATE 2025 Data Science & Artificial Intelligence - DA Q. 33 On a relation named Loan of a bank: Loan loan number branch name amount L11 Banjara Hills 90000 L14 Kondapur 50000 L15 SR Nagar 40000 L22 SR Nagar 25000 L23 Balanagar 80000 L25 Kondapur 70000 L19 SR Nagar 65000 the following SQL query is executed. `SELECT L1.loan number FROM Loan L1 WHERE L1.amount > (SELECT MAX (L2.amount) FROM Loan L2 WHERE L2.branch name = 'SR Nagar');` The number of rows returned by the query is (Answer in integer)
- GATE 2025 Data Science & Artificial Intelligence - DA Q. 43 Consider game trees Tree-1 and Tree-2 as shown. The first level is a MAX agent and the second level is a MIN agent. The value in the square node is the output of the utility function. A 2 B x 1 C D E 5 y 2 MAX MIN Tree-1 Tree-2 For what ranges of x and y , the right child of node B and the right child of node E will be pruned by alpha-beta pruning algorithm? (A) $x \in [1, \infty)$ and $y \in (-\infty, 2]$ (B) $x \in (-\infty, 2]$ and $y \in (-\infty, 5]$ (C)
- 3. A ball is drawn at random from the selected box. Given that the ball drawn is white, the probability that it is drawn from Box-2 is (Round off to two decimal places) Q. 32 $\lim_{t \rightarrow +\infty} \sqrt{t^2 + t} - t =$ (Round off to one decimal place) Organizing Institute: IIT Roorkee Page 17 of 34

Challenges Faced

Technical Hurdles:

Installing & configuring OCR
tools

(Tesseract, Poppler)

Handling token length limits
with Gemini

Embedding mismatches
(cosine vs Euclidean)

Multi-turn context management

Inference errors and fallback strategies

Future Scope & Improvements

🔍 What's Next?

- Multi-language OCR support
(Hindi, Telugu, etc.)
- UI-based conversation history saving
- Source highlighting for citations
- Fallback model with GPT-4 when Gemini fails
- Deploy on cloud (GCP/AWS) with GPU acceleration

Thank You!

Thank you for joining our presentation on the Agentic RAG Chatbot. We appreciate your time and interest!

