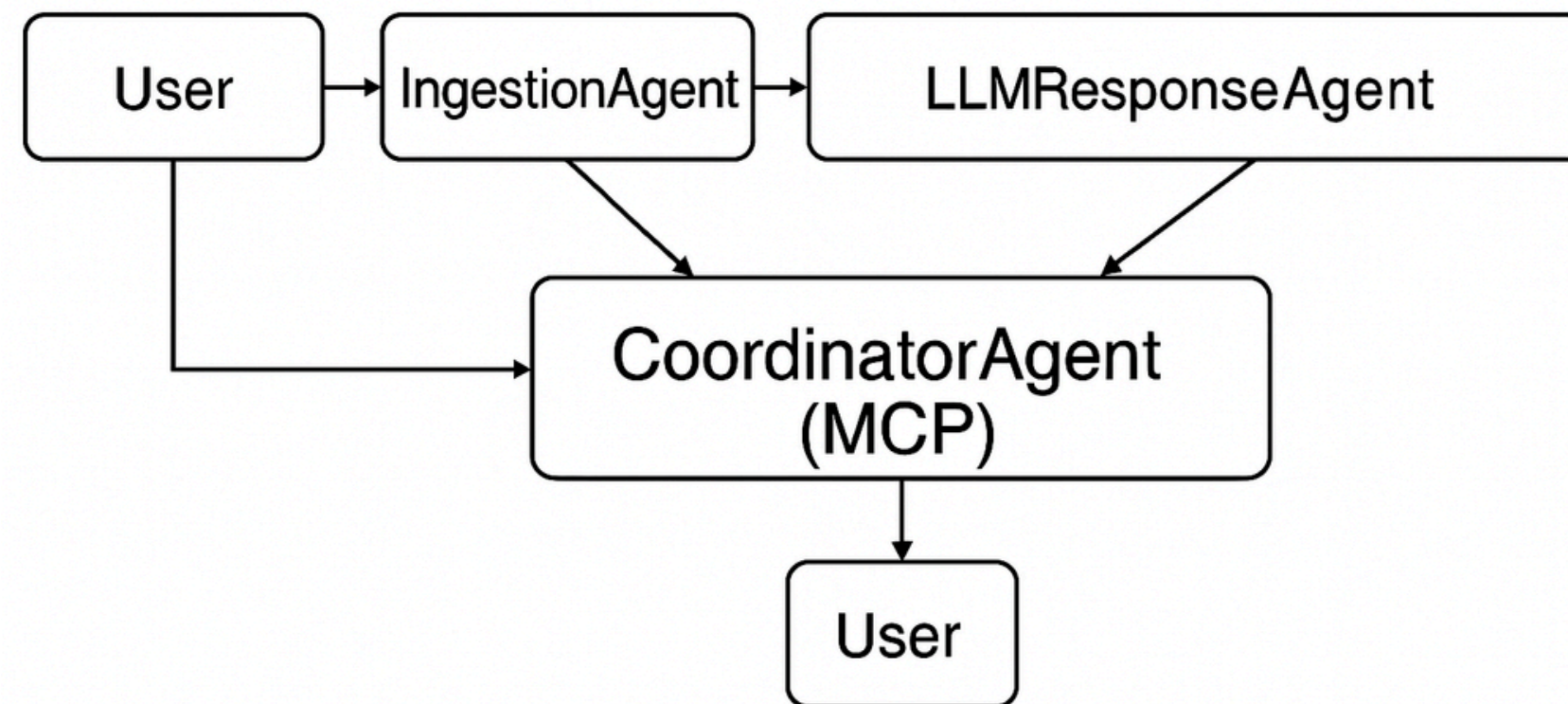


Agentic RAG Chatbot with MCP Messaging

Project Overview + Agent Architecture

Agentic RAG Chatbot that answers user queries from uploaded documents using GPT-4. Documents are parsed, chunked, vectorized, retrieved, and answered, all via coordinated agents.



Key Modules:

- **`IngestionAgent`**: File parsing + smart chunking
- **`RetrievalAgent`**: Embedding + ChromaDB retrieval
- **`LLMResponseAgent`**: GPT-4 based answer generation
- **`CoordinatorAgent`**: Manages workflow via **`MCPMessage`**

System Flow & MCP Message Passing

System Workflow:

1. User uploads files and inputs question.
2. `IngestionAgent` chunks the docs.
3. `RetrievalAgent` stores + fetches top chunks.
4. `LLMResponseAgent` forms response using GPT-4.
5. `CoordinatorAgent` orchestrates everything via messages.

****Example MCPMessage:**

```
```json
{
 "sender": "RetrievalAgent",
 "receiver": "LLMResponseAgent",
 "msg_type": "RETRIEVAL_RESULT",
 "trace_id": "abc-123",
 "payload": {
 "query": "What is a FITS file?",
 "top_chunks": ["chunk 1...", "chunk 2..."]
 }
}
```

## Tech Stack:

1. **UI (User Interface):** The UI is built using Streamlit, it allows for the easy creation of interactive and visually appealing web applications.
2. **LLM (Large Language Model):** The application uses OpenAI GPT-4, one of the most advanced language models available. GPT-4 can understand and generate human-like text, making it ideal for tasks such as text completion, question answering, and more.
3. **Embeddings:** For embedding tasks, the stack uses sentence-transformers. This library is highly effective for generating embeddings for sentences and paragraphs, which are useful for applications like semantic search and clustering.
4. **Vector DB (Vector Database):** ChromaDB is employed as the vector database. It provides efficient storage and retrieval of vector embeddings, which is crucial for applications involving similarity search.
5. **File Parsing:** PyMuPDF for PDF files, python-docx for Word documents, and pptx for PowerPoint presentations.

## Challenges:

1. Chunking too small → poor context
2. Managing state across agents and UI
3. Clean vector store handling per query

## Future Scope:

- Add chat memory (multi-turn conversation)
- Switch to local LLM (e.g., Mistral, Falcon)
- Embed persistence for faster queries
- Visual trace viewer for message flow

# 🧠 Agentic RAG Chatbot with MCP

## 📁 Upload & Process Documents

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

Drag and drop files here

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

Browse files

📄

s11214-021-00852-5.pdf

6.5MB

×

📄

fits\_standard40aa-le.pdf

1.1MB

×

📄

EXI\_Data\_Product\_Guide.pdf

1.2MB

×

Showing page 1 of 3

📁 Process Uploaded Documents

## 💬 Ask a Question

Type your question here:

What does Martian atmosphere consist of?

🔍 Get Answer

## 💬 Ask a Question

Type your question here:

Why Fits format is suitable for astronomical image data?

🔍 Get Answer

## 🧠 Answer

The FITS format is suitable for astronomical image data because it is a standard archival data format for such data sets. It was originally designed for transporting image data and has expanded to accommodate more-complex data structures. The role of FITS has grown from simply a way to transport data between different analysis software systems into the preferred format for data in astronomical archives, as well as the online analysis format used by many software packages. This standard defines the organization and content of the header and data units for all standard FITS data structures. It also specifies minimum structural requirements and general principles governing the creation of new extensions.

📄 Source Chunks Used

^

1. Technical Records of the Federal Government, [US] National Research Council, National Academy Press 1995. This document, hereafter referred to as the 'Standard', de- scribes the Flexible Image Transport System (FITS), which is the standard archival data format for astronomical data sets. Although FITS was originally designed for transporting image data on magnetic tape (which accounts for the 'I' and 'T' in the name), the capabilities of the FITS format have expanded to ac- commodate more-comple...

2. formal codification of the FITS format, which has been endorsed by the International Astronomical Union (IAU) for the interchange of astronomical data (IAU 1983). It is fully consistent with all actions and en- dorsements of the IAU FITS Working Group (IAUFWG), which was appointed by Commission 5 of the IAU to oversee further development of the FITS format. In particular, this standard de- fines the organization and content of the header and data units for all standard FITS data structures: the pr...

3. so that it can be run on future computer systems, then the information encoded in those data files could be lost. 1.1. Brief history of FITS The FITS format evolved out of the recognition that a

Deploy ⋮

## 🧠 Answer

The Martian atmosphere consists of argon, trace levels of water vapor, oxygen, carbon monoxide, ozone, and noble gases. The average surface pressure is 6.1 hPa. The atmosphere can be divided into three layers: the lower atmosphere (up to 50 km), the middle atmosphere (50-100 km), and the upper atmosphere (above 100 km).

📄 Source Chunks Used

^

1. Kerzendorf, Y.P. Bach, B.A. Weaver, F. D'Eugenio, H. Souchereau, L. Ferreira, astropy/photutils: 1.0.0 (2020). [https:// doi.org/10.5281/zenodo.4044744](https://doi.org/10.5281/zenodo.4044744) B.A. Cantor, P.B. James, W.M. Calvin, MARCI and MOC observations of the atmosphere and surface cap in the North polar region of Mars. Icarus 208(1), 61–81 (2010). <https://doi.org/10.1016/j.icarus.2010.01.032> P.R. Christensen, Martian dust mantling and surface composition: interpretation of thermophysical properties. J. Geophys. Res. 87, 9985–9998...

2. and argon. Trace levels of water vapor, oxygen, carbon monoxide, ozone and noble gases are also present with an average surface pressure of 6.1 hPa which is much lower than Earth's average surface pressure of 1000 hPa. Mars has an axial tilt similar to Earth's axial tilt which means the intensity of sunlight varies depending on the location during the year giving Mars seasons. Martian atmosphere can be divided into three layers: lower atmosphere (up to 50 km), middle atmosphere (50-100 km) ...

3. 3 from lower atmosphere, and from above by solar ultraviolet (UV) radiation heating and heliospheric charged particles and magnetic field environment (Almatroushi et al., 2021). In the exosphere collisions are very rare hence particles travel on ballistic paths and escape of particles occur (Haberle, 2003). A deeper understanding of the processes in the thermosphere and exosphere (upper atmosphere) improves our understanding of how the Martian atmosphere has evolved with time from a thic...

Deploy ⋮

## 💬 Ask a Question ↔

Type your question here:

What is a dog?

🔍 Get Answer

## 🧠 Answer

The answer is not available in the context.

📄 Source Chunks Used

^

1. char...] [Constraint: The total number of characters in a FITS commentary keyword record must be exactly equal to 80.] FITS value keyword record := keyword field value indicator [space...] [value] [space...] [comment] [Constraint: The total number of characters in a FITS value keyword record must be exactly equal to 80.] [Comment: If the value field is not present, the value of the FITS keyword is not defined.] long string keyword := initial kwd record [continuation kwd record...] last continuation...

2. HDU Header and Data Unit. A data structure consisting of a header and the data the header describes. Note that an HDU may consist entirely of a header with no data blocks. Header A series of keyword records organized within one or more header blocks that describes structures and/or data that follow it in the FITS file. Header block A 2880-byte FITS block containing a sequence of thirty-six 80-character keyword records. Heap The supplemental data area following the main data table in a binary-tabl...

3. field value indicator [space...] [partial string value] [space...] [comment] [Constraint: The total number of characters in an ini- tial kwd record must be exactly equal to 80.] continuation kwd record := CONTINUE keyword [space...] [partial string value] [space...] [comment] [Constraint: The total number of characters in a continua- tion kwd record must be exactly equal to 80.] last continuation record := CONTINUE keyword [space...] [character string value] [space...] [comment] [Constraint: The ...

Deploy ⋮

Deploy ⋮