

# **Agentic RAG Chatbot for Multi-Format Document QA using Model Context Protocol (MCP)**

**Candidate Name :** Dhowfeek Hasan  
**E-mail :** Dhowfeekhasan@gmail.com  
**Phone :** 7708072882  
**Linkedin :** <https://www.linkedin.com/in/dhowfeek-hasan/>

## Agent-Based Architecture with MCP

In this task, Deployed 4 Agents:

- Agent 1 : Ingestion Agent(Handles file parsing)
- Agent 2 : RetrievalAgent(Embeds & retrieves relevant text)
- Agent 3 : Interacts with LLM to answer
- Agent 4 : Log Monitoring and Saves in CSV file

### MCP:

- **Model Context Protocol**
- MCP is a lightweight communication protocol that enables structured, traceable message passing between agents in a modular system.
- **Components** : "type" , "sender", "receiver","trace\_id","payload" , "query"

## Tech Used:

Components	Tech Used
UI	Streamlit
Coordination.py file	Python Classes, MCP Protocol
File Handling	PyMuPDF, docx2txt, pandas
Vector DB	FAISS + sentence-transformers
LLM	meta-llama/Llama-3-70B-Instruct free
Logging	CSV logs in /logs/ folder
Containerzation	Docker

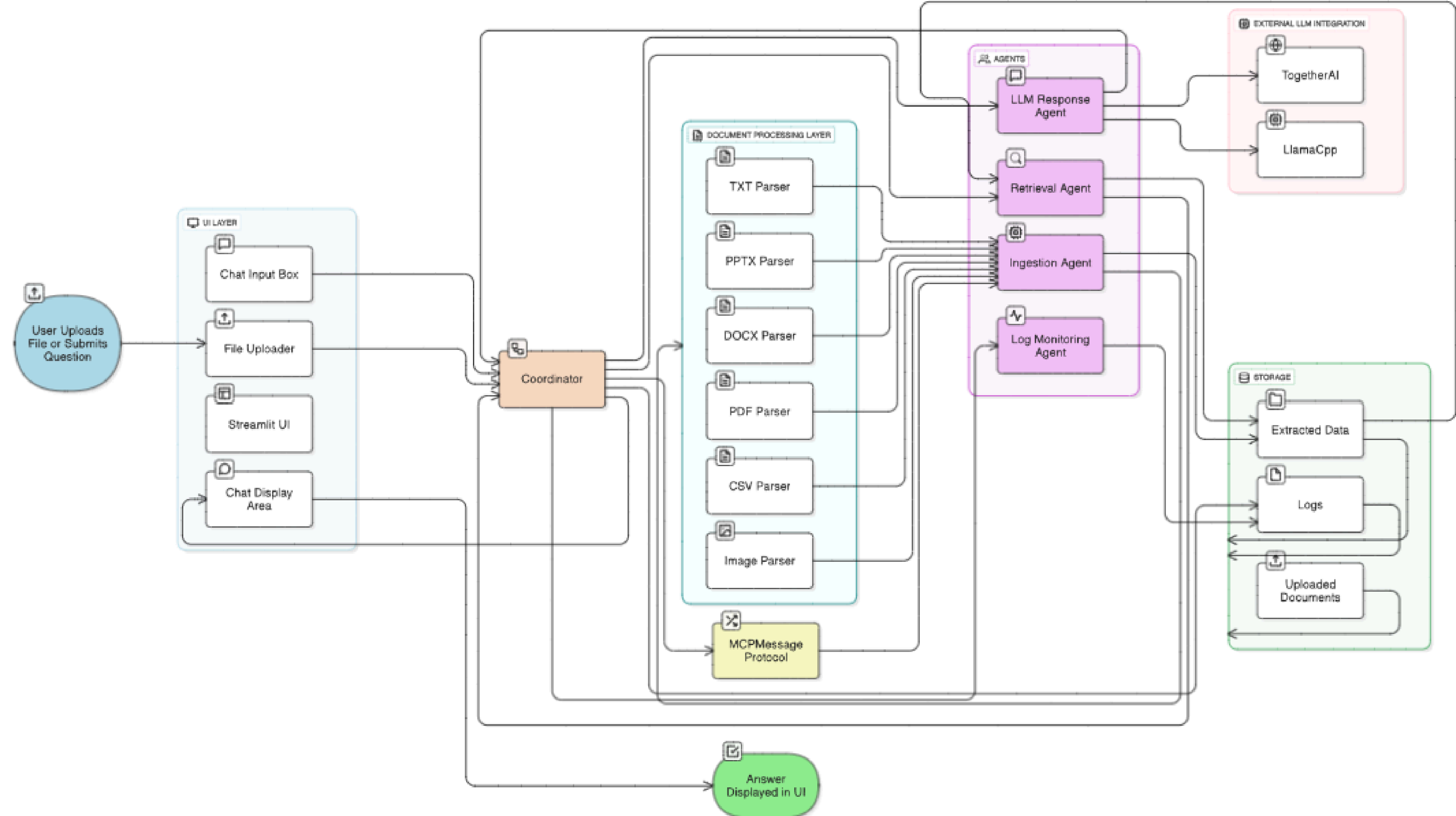
## Extra Features :

- It cleans old file automatically when the new file enters.
- Added Agent 4 for logs monitoring which help us to improve the models accuracy and train the model in future.
- also Accepts the image formats and maintains the original tabular format.
- This Extracts the images from the file (ppt,pdf,word,excel) and save those images in extracted data folder.
- created docker file which helps to deploy easily.
- can able to see the relevant chunks for each answers.

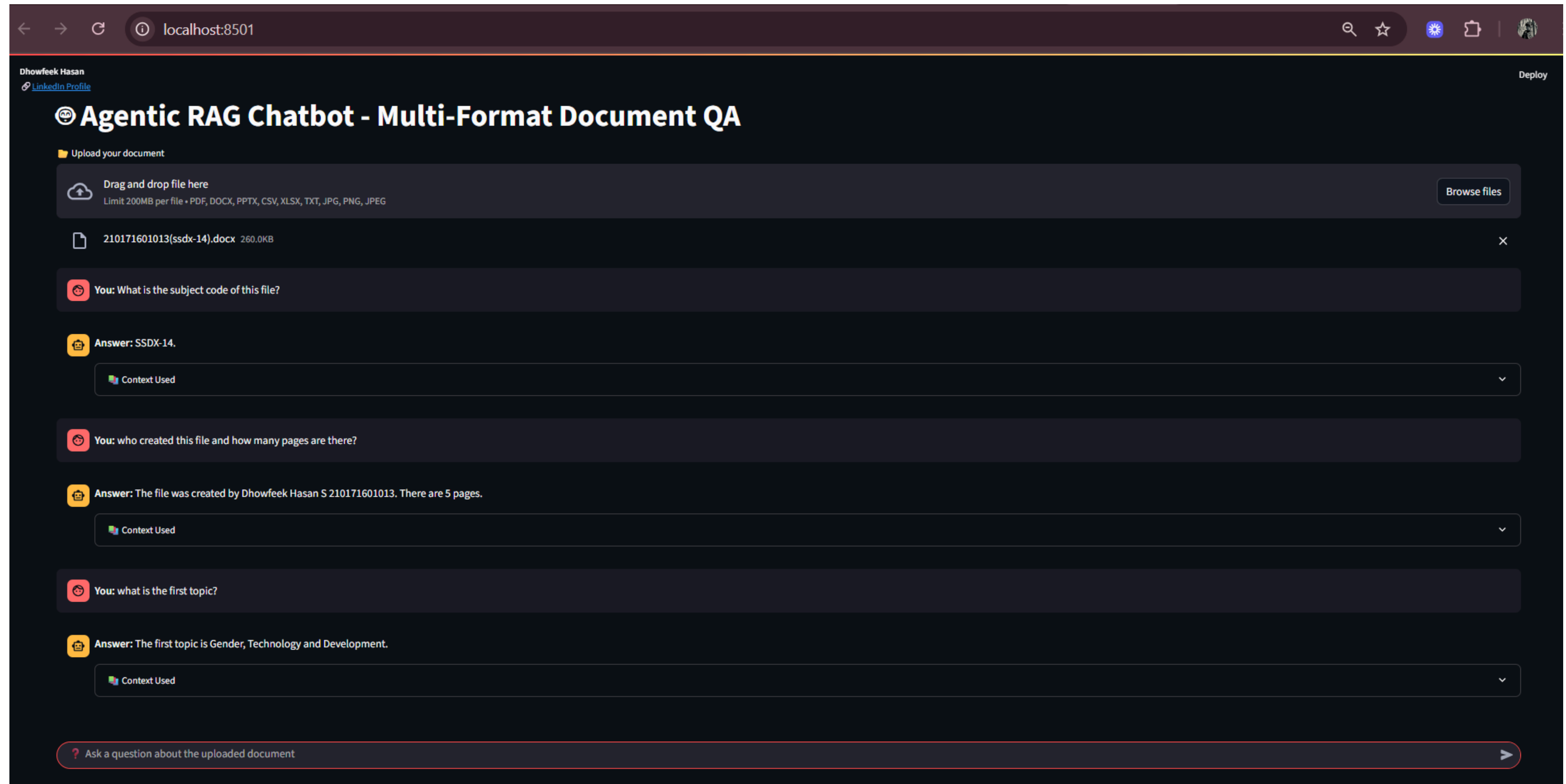
## Comman line to run the project :

- **UI** : streamlit run ui/app.py
- **Run in Terminal** : python -m main.coordinator
- **Run the Docker** : docker build -t rag-chatbot(creates a docker image)  
docker run -p 8501:8501 rag-chatbot

# System Flow Diagram:



# UI screenshots:



Agentic RAG Chatbot for Multi-Format Document QA using Model Context Protocol (MCP)

## UI screenshots:

## Saved Logs:

1	trace_id	timestamp	filename	question	answer	sources	document	status	error_message
2	125c0aae-	05-07-2025 02:43	Coding Ro	First five p	1. CSV	--- Page 1	pdf	success	
3	a7969ec1-	05-07-2025 02:52	WhatsApp	what are t	There are	programm	jpeg	success	
4	52eda8c7-	05-07-2025 02:53	WhatsApp	how many	There are	programm	jpeg	success	
5	c0d53d15-	05-07-2025 02:56	WhatsApp	command	SELECT.	programm	jpeg	success	
6	8bab6f9a-	05-07-2025 02:56	WhatsApp	name the s	SQL stands	SQL Mind I	jpeg	success	
7	fadcb715-	05-07-2025 03:04	Coding Ro	what is the	The first	--- Page 1	pdf	success	
8	a34448dd-	05-07-2025 03:05	Coding Ro	what is the	The second	--- Page 1	pdf	success	
9	a77b82bf-	05-07-2025 03:31	receipt.jpg	what is the	There is no	NS KALE S	jpg	success	
10	77e1a6b3-	05-07-2025 03:31	receipt.jpg	what are t	They	AYMENT C	jpg	success	
11	1dfe29e3-	05-07-2025 03:32	receipt.jpg	what the t	There is no	SLICED hi	jpg	success	
12	3ba2ba48-	05-07-2025 04:03	Load balar	show the f	Least	Least Conr	pptx	success	
13	f071f0f0-8	06-07-2025 23:11	Coding Ro	What is th	The task ti	--- Page 1	pdf	success	
14	4a7cc81e-	06-07-2025 23:15	Coding Ro	task Name	The task n	--- Page 1	pdf	success	
15	ccc375b1-	06-07-2025 23:15	Coding Ro	What is th	The Proble	--- Page 1	pdf	success	
16	985b8a41-	06-07-2025 23:16	Coding Ro	list the all	1. ðŸ”	â—‹ âœ…	pdf	success	
17	4e7c6507-	06-07-2025 23:17	Coding Ro	Task Title	UI screens	--- Page 1	pdf	success	
18	126f9277-	06-07-2025 23:19	Coding Ro	What is th	sales_revi	â—‹ âœ…	pdf	success	
19	e0e050d3-	06-07-2025 23:21	Owais Fin	Student na	Mr.Muhan	APPENDIX	docx	success	
20	044c6684-	06-07-2025 23:22	Owais Fin	Which coll	B.S. Abdur	APPENDIX	docx	success	
21	91c96470-	06-07-2025 23:23	Owais Fin	What cour	M.Tech IT	APPENDIX	docx	success	
22									
23									



## Challenges:

- Firstly, I download the TinyLlama from the huuging face in gguf format but i dont have enough RAM to run the model. Then I use meta-llama/Llama-3.3-70B-Instruct-Turbo-Free through cloud.
- Most faced challenges is attribute error.
- Very difficult to find the proper prompt to get the more accurate results.
- and LLM (gave poor results without a good prompt.
- Multiple extracted\_data/ and uploaded\_docs/ folders were created unexpectedly.

## Future Enhancements:

- Fine tuning the model by using the log data to improve the accuracy.
- Add login-based access for users and admins





Thank you