**MULTI AGENT AI SYTEM FOR FILE CLASSIFIACTION , DETECTION & EXTRACTION:**

**AIM:**  
The main goal of this project is to build a robust, multi-agent AI system that can handle incoming documents in various formats—PDFs, JSON files, and email texts. It is used to classify each document based on its format and intent (like Invoice, RFQ, Complaint, or Regulation), and then route it to the right processing agent. Also, I’ve kept a shared log of all actions—source, type, timestamps, extracted fields, thread IDs—so everything stays traceable and organized across agents.

**System Architecture**

| **Agent** | **Functionality** |
| --- | --- |
| Classifier Agent | Detects file format (PDF/JSON/Email) & intent (keywords + LLM). Routes accordingly. |
| JSON Agent | Validates JSON structure. Flags missing or strange fields. |
| Email Agent | Pulls out sender, subject, body, urgency—formats it for CRM-style use. |
| PDF Handler | Uses OCR (pytesseract) to extract text from PDFs. |
| LLM Integration | Advanced intent detection via OpenAI GPT-3.5-turbo. |
| Shared Memory | Logs source, type, timestamp, extracted fields, and thread ID for full traceability. |

**Key Features**

* Detects document formats: PDF, JSON, Email
* Intent detection by:  
   • Keyword-based classification (Invoice, RFQ, Complaint, Regulation)  
   • LLM classification using OpenAI GPT-3.5-turbo
* Routes to the right agent (JSON or Email)
* Field extraction:  
   • JSON: Checks id, date, amount  
   • Email: Grabs sender, subject, body, urgency  
   • PDF: OCR text extraction
* Shared memory logs for source, type, timestamp, extracted fields, thread ID
* Modular design for easy upgrades and future features

**Sample Workflow**  
1️ User uploads a file via Colab (PDF, JSON, Email)  
2️. Classifier Agent spots the format and intent  
3️. It sends the file to the right agent (JSON or Email)  
4️. Extracted data gets saved in shared memory  
5️.Final shared memory log gives a full, traceable picture of what happened

**Tech Stack**

* Python 3.x
* OpenAI GPT-3.5-turbo API for intent detection
* pytesseract + pdf2image for PDF OCR
* Google Colab as the execution environment
* pandas for shared memory and logs and for display

**Outputs**

* Final shared memory logs (in DataFrame & CSV format)
* Extracted fields from each file
* Logs of both LLM and keyword-based intent detection