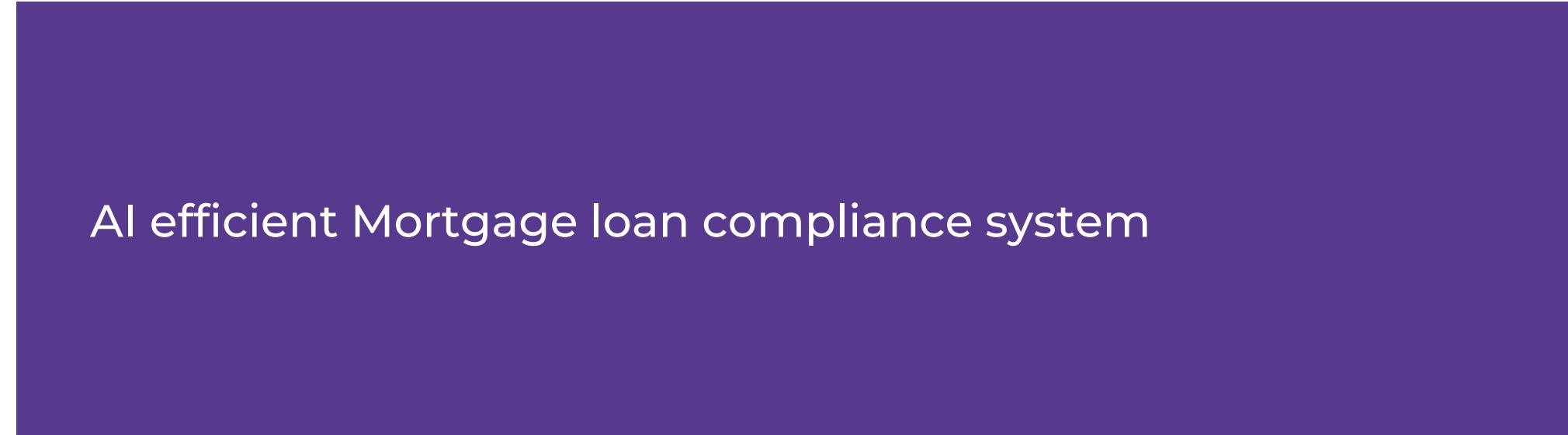


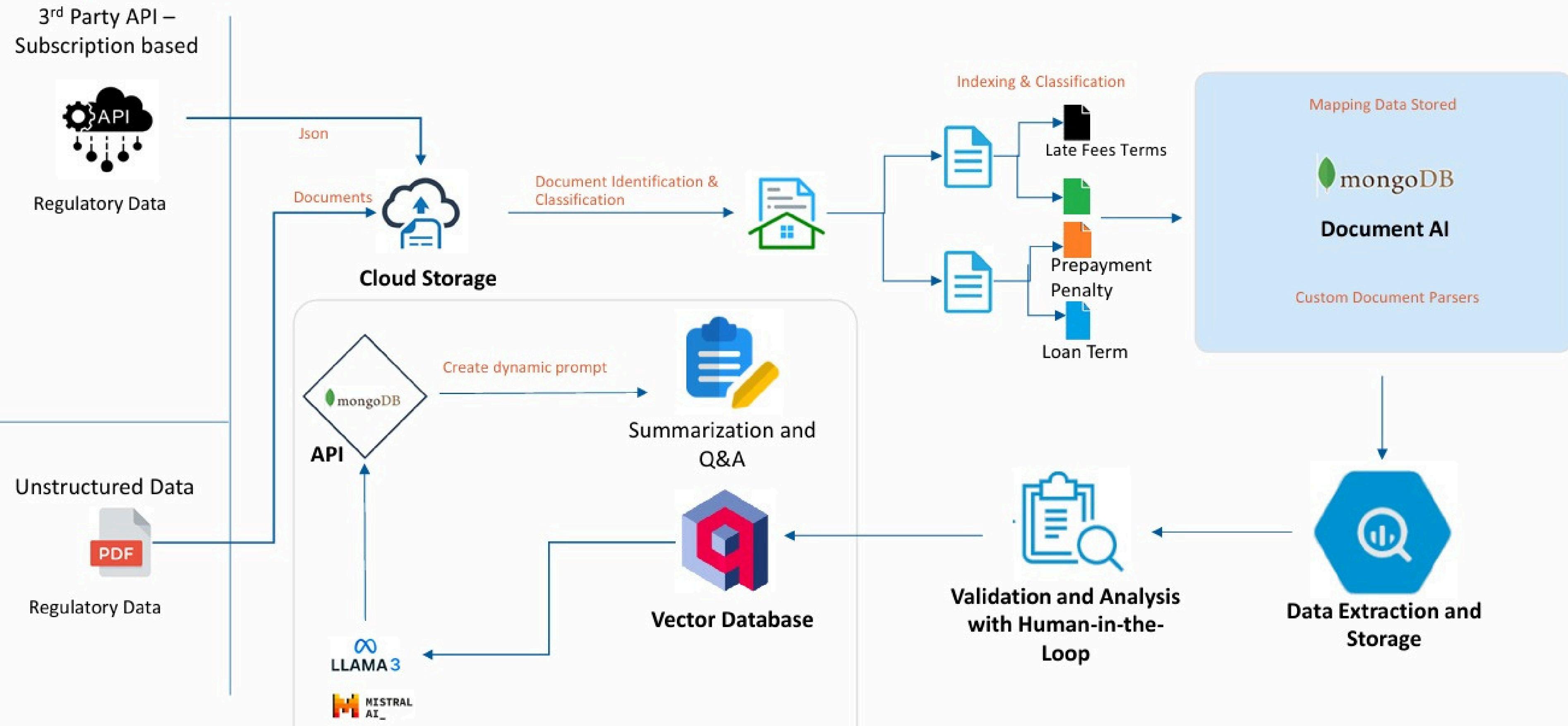


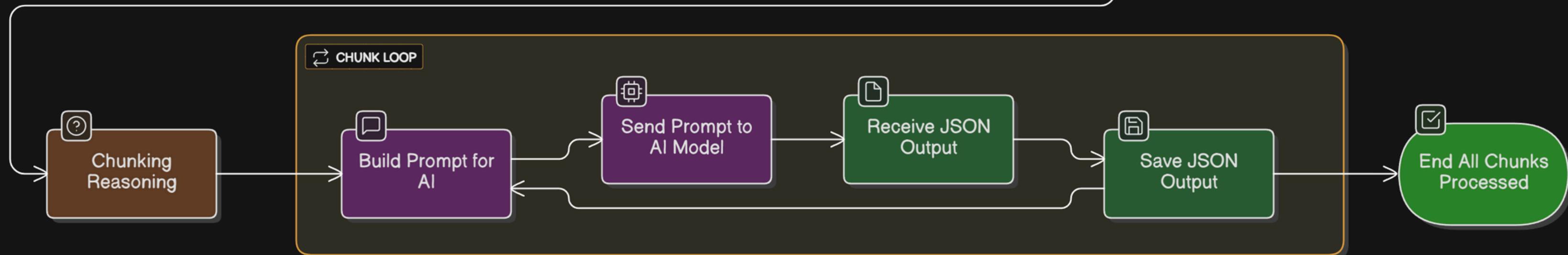
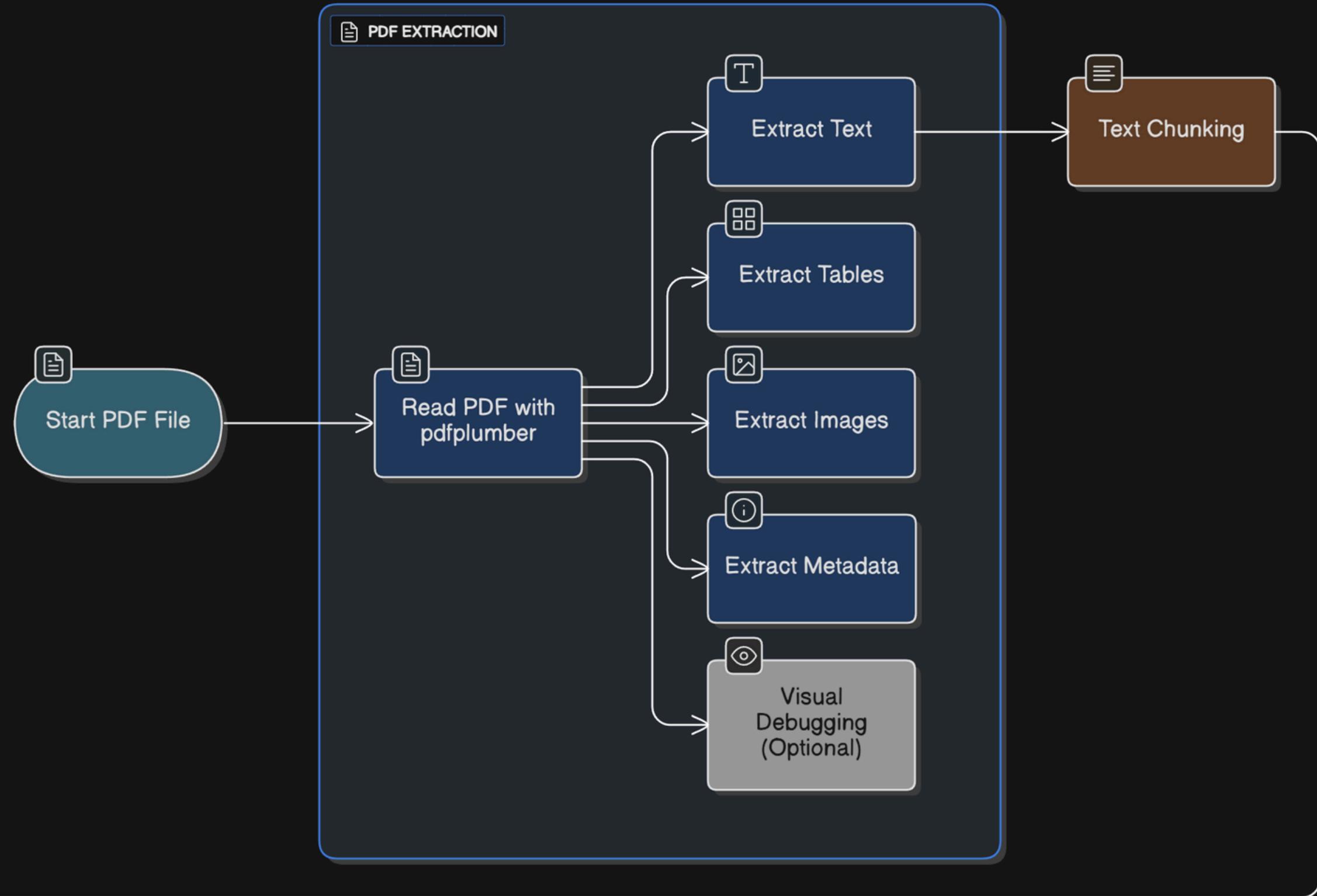
TCS - Mortgage Loan



AI efficient Mortgage loan compliance system

Proposed Gen AI Based Solution Architecture





Key benefits

- 1.Faster lookups
- 2.Efficient search
- 3.Structured RAG (Retrieval-Augmented Generation)

AI models like Mistral or LLaMA can only handle a limited number of tokens (words/characters) at once.



Problems we faced

- **Running the Model Locally**

1. High Latency: Processing a 25-page PDF takes ~3 hours on local hardware.
2. Resource Intensive: Large models consume significant CPU/GPU and RAM.
3. Scalability Issues: Difficult to handle multiple PDFs or real-time processing.

- **JSON Storage & Management**

1. Unclear Storage Strategy: Unsure to use MongoDB, Postgres, or flat files.
2. Scalability Concerns: Need a structure that supports thousands of Q&A entries.
3. Efficient Retrieval: Storing is easy, but indexing and fast querying challenges.
4. Traceability Issues: Metadata needs to be stored consistently for future lookups.



Sample Json Output:

```
"instruction": "Interpret the text about granting loans to Non-Resident Indians.",  
"input": "",  
"output": {  
    "can_banks_grant_loans": true,  
    "conditions": [  
        "Banks can grant term loans to housing intermediary agencies against direct loans  
sanctioned/proposed to be sanctioned by them for NRIs"  
    ]  
},  
"metadata": {  
    "source_file": "mortgagetest.pdf",  
    "page_number": 8,  
    "chunk_id": "8_0"  
}
```

Workflow

Document Ingestion

- Collect regulatory PDFs + mortgage loan forms
- OCR (extract text while preserving structure)
- BeautifulSoup, pdf2image, pytesseract

Text Cleaning & Structuring

- Remove headers/footers
- Maintain sections (e.g., "Section 3.2 – Prepayment Clause")
- Split into chunks (500–1000 tokens)

{
 Instruction: - Question - ,
 Input : - Reference - ,
 Output : - Answer - ,
 Metadata{.....} ,
}

Q&A Generation from Regulations

- Use LLM to generate structured Q&A pairs with:
- Question
- Answer (verbatim from text)
- Section ID
- Citation details
- "iarfmoose/t5-base-question-generator"
- distilbert-base-cased-distilled-squad

Workflow

Knowledge Storage

- Save Q&A in JSON: {question, answer, source, section, citation}
- Convert chunks & Q&A into embeddings
- Store in Vector DB + Metadata DB

Loan Form Processing

- Classify document type (home loan, car loan, etc.)
- Extract fields (interest rate, tenure, foreclosure charges, etc.) using LayoutLM/Donut

Compliance Check

- Query vector DB with extracted field ("prepayment fee")
- Retrieve top-k relevant regulation clauses
- Feed into LLM with form fields → reason over compliance

Alerting & Recommendation

- If non-compliant → generate structured output:

Limitations

Text extraction through poor pdfs

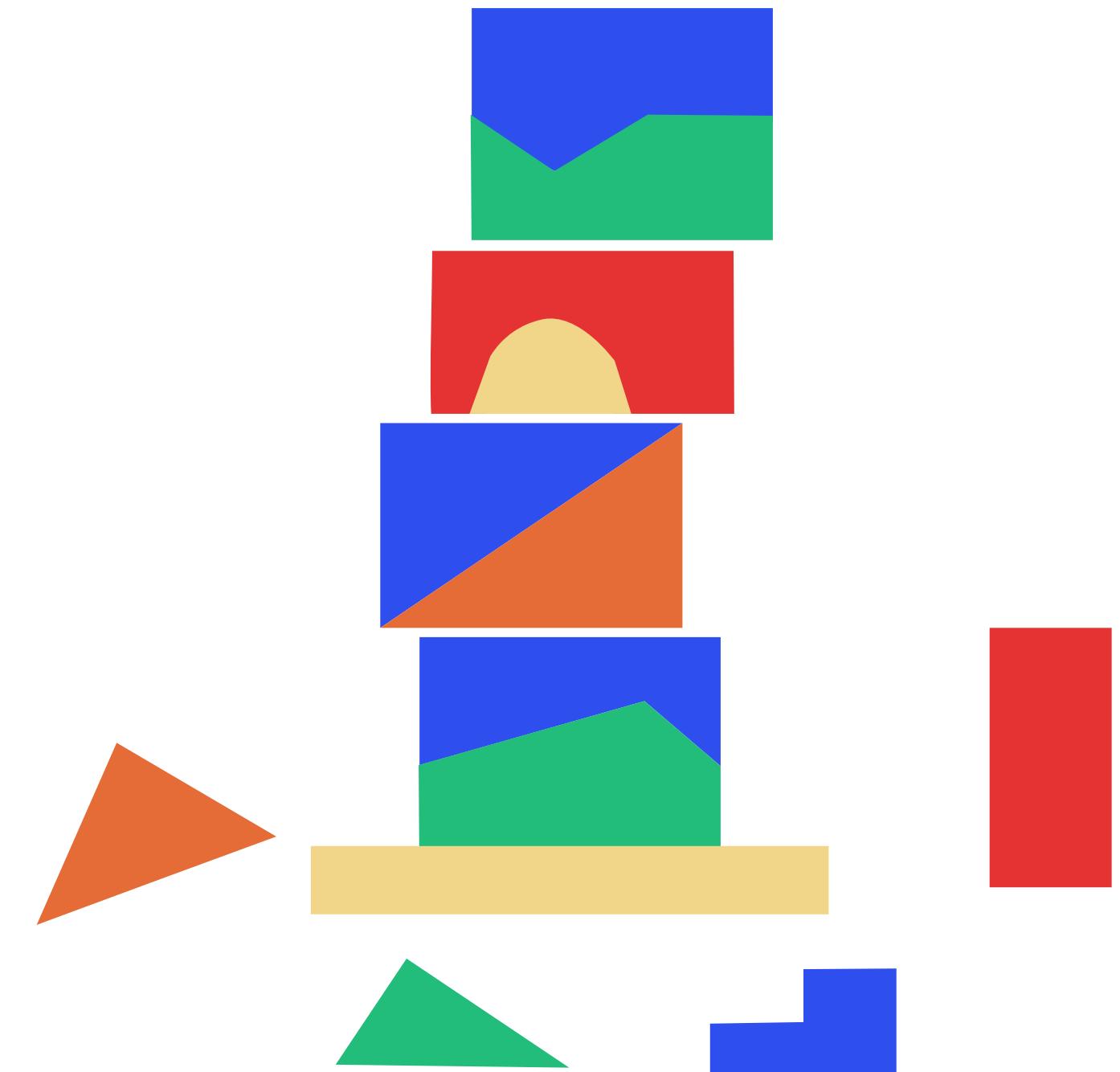
- Quality of pdfs affects the text extraction capabilities.
- Converted pdf through different formats (doc, image etc) leads to poor extraction.

Inconsistency in web scrapping

- Different HTML Structures for different webpages limit the reuse of code.
- Frequent Changes in design and structure overtime.

Time

- LLM Model Size & Hardware Limitations
- Larger inputs are taking more time on the CPUs.



Future Scope

- Use of third party API's for text extraction makes the process fast eg. Pdf rest API
- Use of Google's BERT for efficient text extraction (Predictive approach)
- Auto Trigger → New PDF scraped → Conversion runs automatically
- Database Storage → Store structured Q&A JSON in MongoDB / Postgres
- Classifier Stage → Each Q&A tagged with a Category
- Examples: Eligibility, Loan Terms, Regulatory Penalties, Definitions

Reference Link for Documents:

sr no.	name	reference links
1.	Reserve Bank of India (RBI) circulars & guidelines (PDFs)	https://www.rbi.org.in/commonman/Upload/English/Notification/PDFs/69SR010709_F.pdf?
2.	South Indian Bank	https://www.southindianbank.com/userfiles/forms/Id%202172.pdf?
3.	IBA and government education/loan scheme	https://www.education.gov.in/sites/upload_files/mhrd/files/upload_document/IBAM.pdf?
4.	IREDA (Specialized Public sector loan agreements)	https://www.ireda.in/images/legaldoc/Loan-Agreement%28secured-against-Bank-Guarantee%29.pdf?
5.	BNP Paribas Commercial bank term-loan	https://www.bnpparibas.co.in/app/uploads/sites/25/2024/11/bnpp-term-loan-agreement-india.pdf?