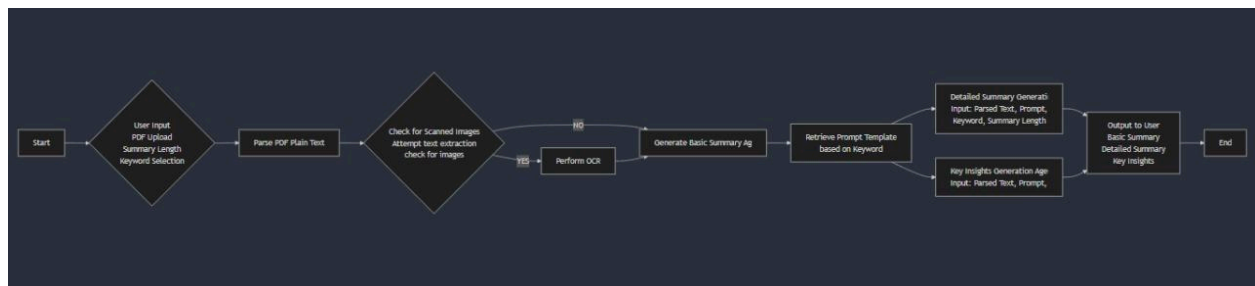


## Problem statement:

Develop a platform that processes and analyzes competitor-related documents such as financial reports, market research papers, and business strategy documents in PDF format. The system must leverage Generative AI to extract key insights, summarize findings, and identify emerging trends in competitor strategies. Key functionalities should include document parsing, AI-driven summarization, and entity recognition to highlight crucial aspects such as market positioning, strengths, weaknesses, and business opportunities.

## Flowchart:



## Methodology:

The pipeline first begins by taking a user input (keywords from a list, summary expectations as well as pdf input). This input is put through a PDF parser (PyPDF2) or through an OCR library (to parse scanned documents). Based on the parsed pdf as plain text. An agent is used to make a basic summary that outlines the entire document without any optimization that improves user experience while they are waiting for detailed analysis to generate.

Based on the keywords selected by the user (These keywords include classification of documents by the user in categories like "Market research paper", "Financial report", "Business strategy"), a series of pre-made prompt templates are selected that are used further for summary generation and insights generation.

This summary generated is catered to the user's original preferences and it is a detailed summary that uses keywords as well as prompt templates to generate a summary based on the document type.

A similar logic is used for key-insight generation but here the prompt template is dedicated to analysis instead of summarization.

1	2	3	4	5	6
15	7	9	10	123	125
2138	1948123	213	293	1010	55
66	18	19	20	21	22

## Keywords

Once the keywords are chosen, they are paired with a set of predefined prompt templates. These prompts are designed to elicit specific types of information or analysis related to the keywords. For example, a prompt like "Use SWOT analysis" can be used to generate analysis for keyword "Business strategy" or a prompt like "Give aid in budget and planning" can be used to generate analysis for keyword "Financial report". The keyword is used to map the prompt template to create a specific, targeted prompt. This combination of pre-defined prompts and chosen keywords allows for the generation of insightful and relevant summaries

## Key insights and summarization.

- LLM Agents are used for both tasks.
- **Detailed Summary:** Generated using the parsed text, selected keyword, retrieved prompt template, and summary length selection. Provides a comprehensive overview of the document.
- **Key Insights:** Generated using the parsed text, selected keyword, and retrieved prompt template. Focuses on extracting the most important findings and trends.
- The use of prompt templates is used to guide the LLM's output.
- The prompt templates are different for detailed summary as well as key insights that guide the LLM to its respective tasks with good accuracy.
- Basic summary is displayed to the user instantly for a better UX.

## Impacts and Benefits

- **Efficiency:** Automates the analysis of competitor documents, saving significant time and effort.
- **Insights:** Provides valuable insights into competitor strategies, strengths, and weaknesses.
- **Informed Decisions:** Enables data-driven decision-making.
- **Competitive Advantage:** Helps businesses stay ahead of the competition.
- **Scalability:** The platform can handle a large volume of documents.
- **Security:** Controlled user input minimizes security risks.

## Future Scope:

- **Table Extraction:** We will integrate table extraction capabilities to analyze tabular data within PDFs using libraries like Tabula-py or Camelot.
- **Free-Form User Inputs :** We will carefully explore free-form user input for more customized analysis, while maintaining security. This would require robust prompt validation.
- **Dynamic Keyword Generation:** Instead of only predefined keywords, we will explore having the system suggest or generate keywords based on the uploaded document's content.
- **Enhanced Visualization:** We can add visualizations (charts, graphs) to present the insights in a more engaging way.



## Methodology:

The pipeline first begins by taking user input (text as well as pdf input). This input is then processed (to parse scanned documents) to generate a basic summary that outlines the user's experience while they are working with the system.

Based on the keywords selected by the user in the "Keywords" section, a series of pre-defined prompts are generated and used to analyze the documents by the user in the "Documents" section.