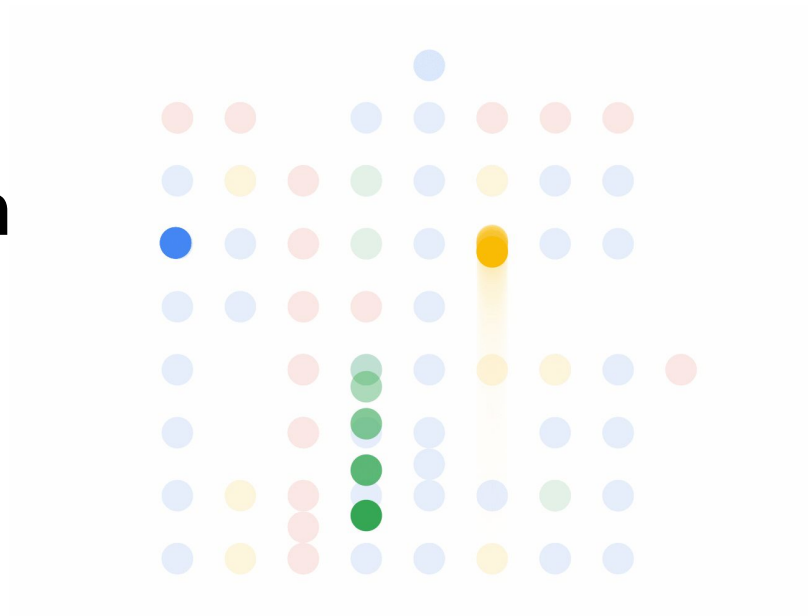


Intelligent Data Extraction from PDF

-Powered by Gemini-Pro & Langchain



1

Problem Statement:

FS enterprises often **struggle to quickly manage and extract valuable insights from vast amounts of in-house data**. This poses a **significant challenge** for providing **timely responses** to customer inquiries.

2

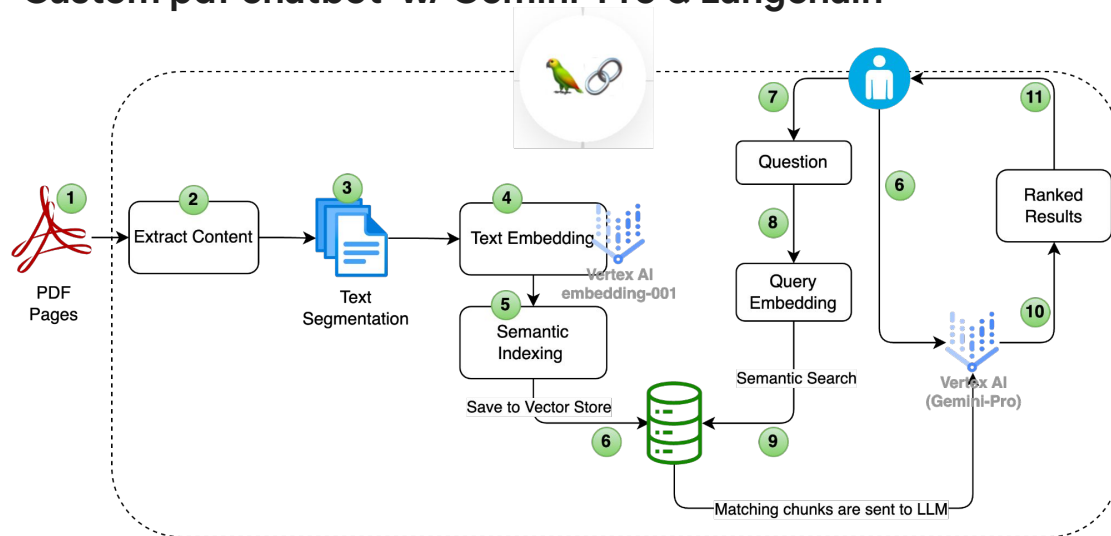
Use Case

By leveraging the power of **Large Language Models** (LLMs) with technologies like **Langchain** and techniques like **RAG**, enterprises can revolutionize how they interact with data and create production ready solution quickly. .

- **Intelligent Data Extraction from Unstructured Documents**
 - Custom pdf chatbot w/ Gemini-Pro & Langchain
 - Pdf chatbot w/ Vertex Search & Langchain

Custom pdf chatbot w/ Gemini-Pro & Langchain

Proprietary + Confidential



- **Gemini-Pro** as the LLM
- **Vertex AI** (textembedding-gecko@001) for text embedding
- **Chroma db** as vector database
- **Langchain** as the overall orchestrator for document extraction, text splitting etc
- **GCS bucket** to store the uploaded files
- **Colab Enterprise** for the notebook
- **Streamlit** for end user interface
- **Ngrok** for UI hosting

Highlights:

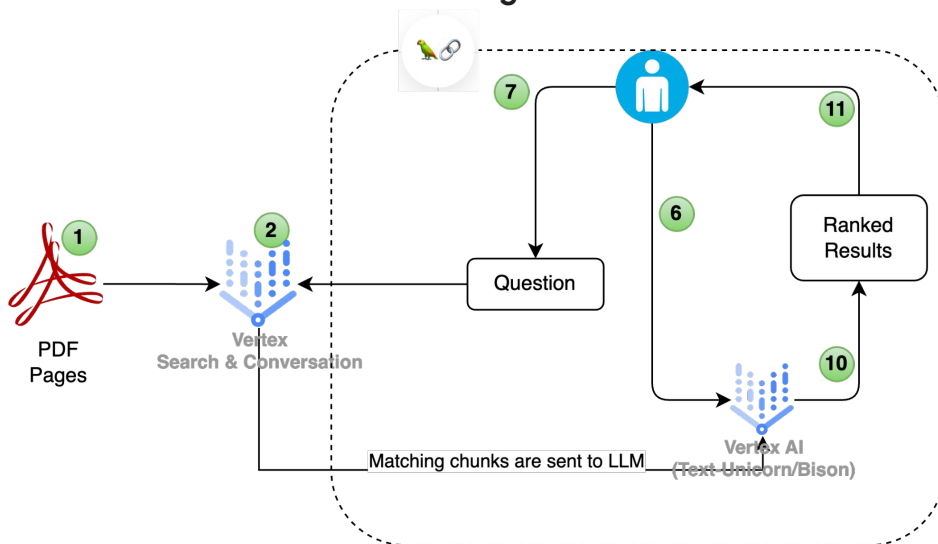
- **Build vs Buy:** Choice for customization
- **Greater flexibility:**
 - **Model Selection:** Choose the best fit & specific to the need
 - **Retrieval Control:** Fine-tune search results
 - **LLM Parameters:** Optimize responses
 - **Prompt Updates:** Improve grounding
- **Open Source w/ custom libraries:** Freedom and flexibility to build unique features
- **Gaining Enterprise Acceptance/Popularity :** Easier adoption

Lowlights:

- Not a Managed Solution
- Longer time to Production = More Developer time = more cost
- Potential Security vulnerability if running w/ outdated versions
- Management overhead

Pdf chatbot w/ Vertex Search & Langchain

Proprietary + Confidential



- **Unicorn** as the LLM
- **Vertex AI** for Search
- **Langchain** as the overall orchestrator for document extraction, text splitting etc
- **GCS bucket** to store the uploaded files
- **Colab Enterprise** for the notebook
- **Streamlit** for end user interface
- **Ngrok** for UI hosting

Highlights:

- **Vertex Search:**
 - Fully Managed Solution
 - Powered by Google Search Technologies
 - Fast & Accurate information retrieval at scale
 - No need to write additional code for embeddings, or storing indexes into vector store
- **Flexibility**
 - Model Selection, Retrieval Control, LLM Parameters, prompt templates etc.

Lowlights:

- Retrieval part is not managed
- Potential Security vulnerability if running w/ outdated versions
- Management overhead