

# PDF QUERY TOOL

Team Ciise

# AGENDA

Introduction

High Level Architecture

Project Timeline

Short Demo

# INTRODUCTION

With the growth and adoption of AI and ML around us, I was interested to take on a project in this space.

This project is focused on creating an MVP: an app that takes in a collection of PDFs and coupled with a LLM, provide a chat-like interaction for querying documents.

Essentially, ChatGPT for your PDFs.

The bigger goal after this project is to take what I've learned and wrap this question answering bot in something like an Azure function so we can provide our customers with an easy to integrate solution....unless there's something out there that does it already cheaply...

## PROJECT: PDF Query Tool

Team Cise

Cabdulqadir Cise

Project start date: 9/10/2023

Scrolling increment: 5

Legend:

On track

Low risk

Med risk

High risk

Unassigned

September

October

November

15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	1	2	3	4	5	6	7	8	9
F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T

Milestone description	Category	Assigned to	Progress	Start	Days
-----------------------	----------	-------------	----------	-------	------

Exploring Existing Solutions

Explore off the Shelf Solutions	On Track	Cabdul	100%	9/13/2023	3
---------------------------------	----------	--------	------	-----------	---

What is LangChain?	On Track	Cabdul	100%	9/18/2023	10
--------------------	----------	--------	------	-----------	----

What is Streamlit?	On Track	Cabdul	100%	9/20/2023	1
--------------------	----------	--------	------	-----------	---

Create UML Artifacts

System Request	On Track		100%	9/13/2023	6
----------------	----------	--	------	-----------	---

System Proposal	On Track		100%	9/15/2023	13
-----------------	----------	--	------	-----------	----

Functional Diagrams	On Track		100%	9/23/2023	6
---------------------	----------	--	------	-----------	---

Structural Diagrams	On Track		100%	10/1/2023	5
---------------------	----------	--	------	-----------	---

System Request Specification	On Track		100%	10/10/2023	5
------------------------------	----------	--	------	------------	---

Traceability Links	Low Risk		0%	11/3/2023	3
--------------------	----------	--	----	-----------	---

Change Management Plan	Low Risk		0%	11/10/2023	3
------------------------	----------	--	----	------------	---

Codebase

Setup Repository	On Track		100%	9/13/2023	1
------------------	----------	--	------	-----------	---

Create Project Site	On Track		100%	10/3/2023	2
---------------------	----------	--	------	-----------	---

Setup Environment	On Track		100%	10/7/2023	3
-------------------	----------	--	------	-----------	---

Implement PDF Processing	On Track		100%	10/6/2023	7
--------------------------	----------	--	------	-----------	---

Build UI for PDF Upload	On Track		100%	10/13/2023	1
-------------------------	----------	--	------	------------	---

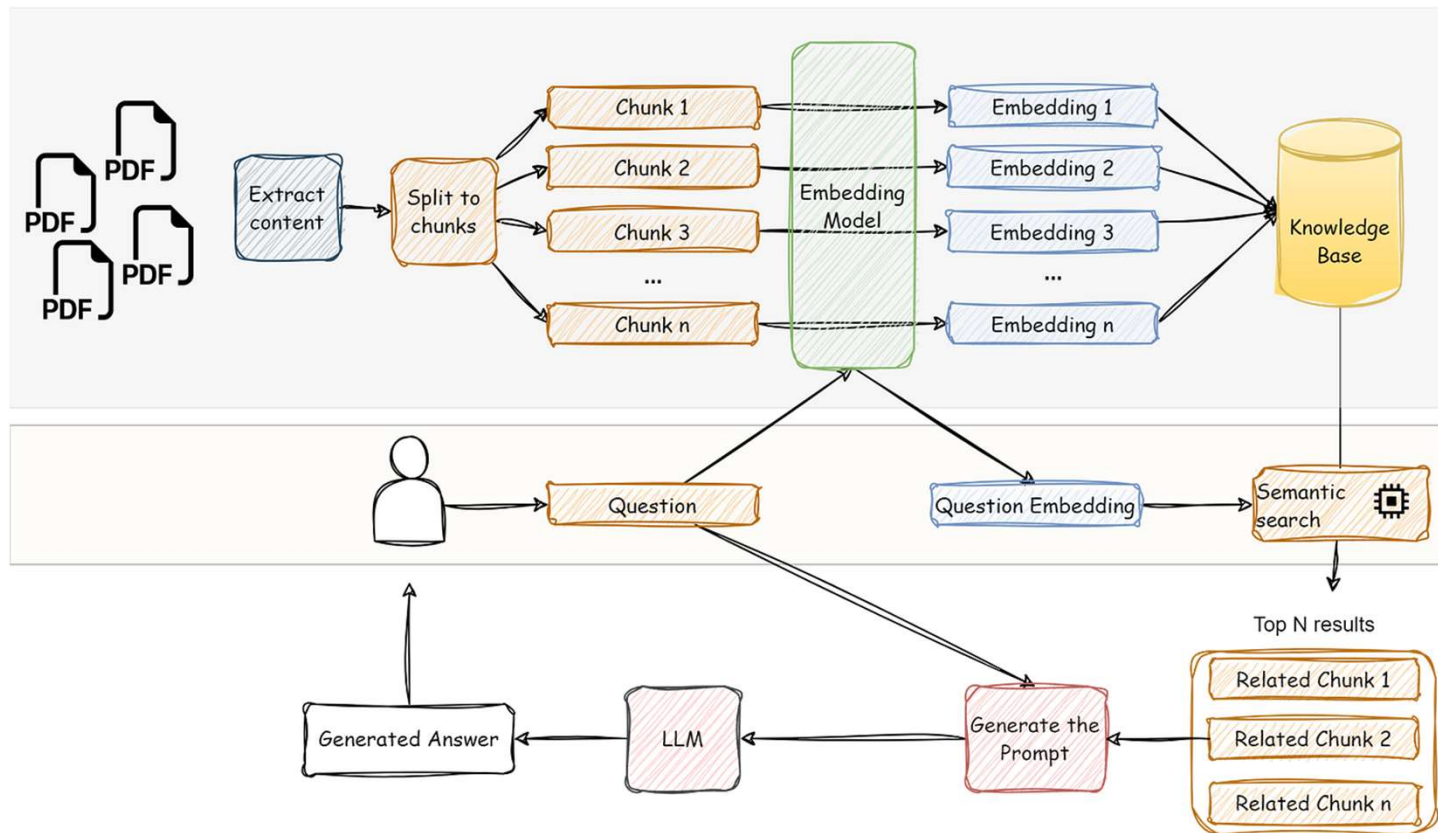
Adding Persistence	Med Risk		0%	10/20/2023	7
--------------------	----------	--	----	------------	---

Add User Login	Med Risk		0%	10/14/2023	7
----------------	----------	--	----	------------	---

Explore Free LLM/Embedding Models	Med Risk		0%	10/21/2023	7
-----------------------------------	----------	--	----	------------	---

Deploy Application	Med Risk		0%	10/28/2023	7
--------------------	----------	--	----	------------	---

# HIGH LEVEL ARCHITECTURE

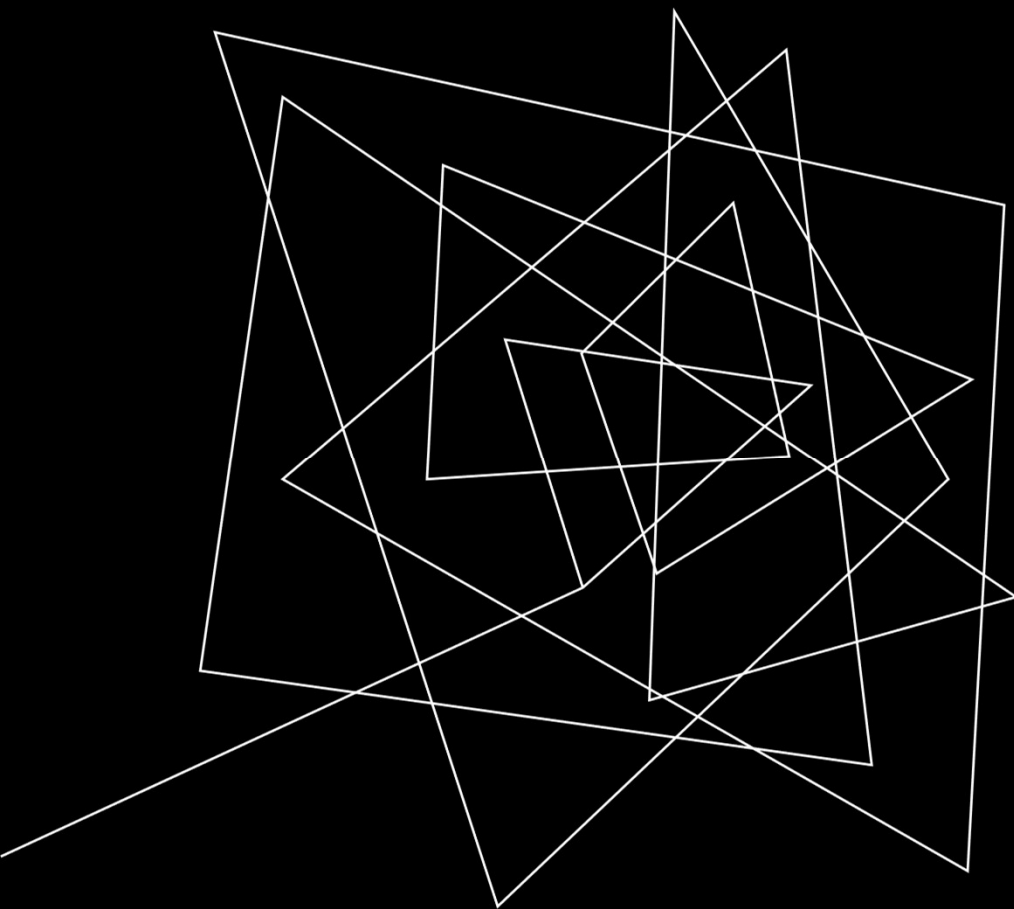


[Source](#)

# PYTHON IMPLEMENTATION

Using the following libraries

1. LangChain – framework for apps using language models
2. Streamlit - UI
3. Python-dotenv – virtual environment
4. Pypdf2 – pdf reader
5. Faiss-cpu – vector store



DEMO TIME



# ANY QUESTIONS