Title:: Deploy AI-enabled chatbots on retail platforms to assist customers in product selection, answer common inquiries, and elevate the shopping experience

Problem Statement:
Inefficient customer support and lack of personalized assistance hinder the retail experience, leading to decreased customer satisfaction and potential loss of sales. Integrating AI-enabled chatbots on retail platforms is imperative to address these issues by providing real-time product guidance, addressing common queries, and enhancing overall shopping engagement.

Description: This project ai. ms to develop retail platforms to assist customers in product selection, answer common inquiries by artificial intelligence (AI) technologies. The system will be capable of processing PDF documents,
extracting text chunks using LangChain, generating embeddings using LangChain's Sentence Transformers, and leveraging Azure OpenAI's RAG (Retrieval-Augmented Generation) models for efficient document retrieval and summarization.

data sources:

in every retail Product company provide manuals which are in pdf documents ,so client provides users with detailed information on a product's installation, operation, maintenance, and troubleshooting. some amount pdf
pdf files extract text and metadata from PDFs, merge multiple PDF files, split PDFs into separate pages, using this Python library like PYPDF, PDF Miner .data extracting to pdf to chunks

nlp preprocessing:

embeddings: embeddings involves transforming each chunk of text into a numerical representation using Word2 vec, glove or pre train BERT, GPT trbuo3.5 ,embedding like tokenization,Aggregate Chunk Embeddings

vector store: vector store using embedding to save the number a vector database service like (pinecone) data save numpy arry,Database Storage(sql, my sql)

user: user asked different prompts(Instructional, text, value prompts,list prompts Error,Confirmation,FeedbackPrompts)
a prompts changes into question embeddings

Semantic search: using Azure OpenAI to enhance search capabilities.
technique used to improve the relevance of search results by understanding the meaning (semantics) of the search query and the content being searched.Natural Language Processing (NLP),Semantic Matching,Ranking and Relevance

or
operated on keyword-based algorithms, focusing on matching words in user queries with those in documents, understanding of the query's context or the user's intent. The advent of semantic search marked a paradigm shift towards understanding the meaning behind words, revolutionizing how information is retrieved.

RAG: straightforward approach to Retrieval-Augmented Generation (RAG),hree-step process comprises indexing, retrieval, and generation.The indexing phase involves preparing your data for retrieval, encompassing information you wish your Large Language Model (LLM) , libraries combination of RAG ,

like LlamaIndex and LangChain , Open AI and costom gpt

LLM: The indexing phase involves preparing your data for retrieval, encompassing information you wish your Large Language Model (LLM).The retrieval process is triggered when a user queries the LLM. Instead of immediately forwarding the user's question to the LLM, you retain the query and enhance it with additional information gleaned from the text chunks in your index.he identified text chunks are inserted into a prompt that incorporates the user's original query

Evaluate model performance:

BLEU score assesses the quality of machine-generated translations compared to human-generated references,
perplexity evaluates the performance of language models in predicting the next word in a sequence of text. Both metrics are essential for assessing the effectiveness and quality of natural language processing models

Responsibilities:
* Utilize LangChain for preprocessing PDF documents, extracting them into text chunks.
* Use LangChain's Sentence Transformers to generate embeddings for text chunks.
* Develop and fine-tune RAG models using Azure OpenAI for efficient document retrieval and summarization.
* Integrate Azure services for advanced NLP capabilities, including RAG model deployment and management.
* Implement Chroma DB to store embeddings, metadata, and document representations.
* Optimize database queries in Chroma DB for efficient retrieval and analysis.
* Evaluate model performance and optimize based on accuracy and relevance metrics.