

HelpMateAi.Books 

Project Submission

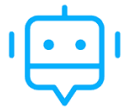
Student Name: Tarriq Ferrose Khan

Batch : DSC 68

Email : tarriqferrosekhan@gmail.com

upGrad

upGrad & IIITB | Data Science Program - May
2024



- Problem Statement
- Solution Approach
- Why LLAMAINDEX

Problem Statement

Based on Option 2: Build Your Own Project (BYOP), a Book product recommendation system is being built to recommend Books from a catalogue (CSV file). This dataset has a bunch of books of various Genre targeting research Students. The User here is a Academic Research Student who intends to ask for Book recommendation based on Key attributes like : Genre , Published Year (recency), volume (Pages) and price.

Solution Approach

A Chat based **Prototype** is provided using LLAMAINDEX , indexing and searching the Catalogue CSV File basis the User Prompt provided to the Chat Engine. The Solution is composed of various layers , detailed in the upcoming section, with necessary validations like Moderation Check before the actual prompt is based to the core AI Engine.

Why LLAMAINDEX

- Simple and most suitable for a Prototype
- This situation we don't have versatile data sources and the data is of limited size, where Llmindex would pay off well compared to Lang Chain.
- The Search and Retrieval is streamlined in LLAMAINDEX

Lessons Learnt while using LLAMAINDEX

- Handle Hallucination through temperature parameters.
- Restrict LLAMAINDEX to search from only within the data set provided and discourage responses from outside world.

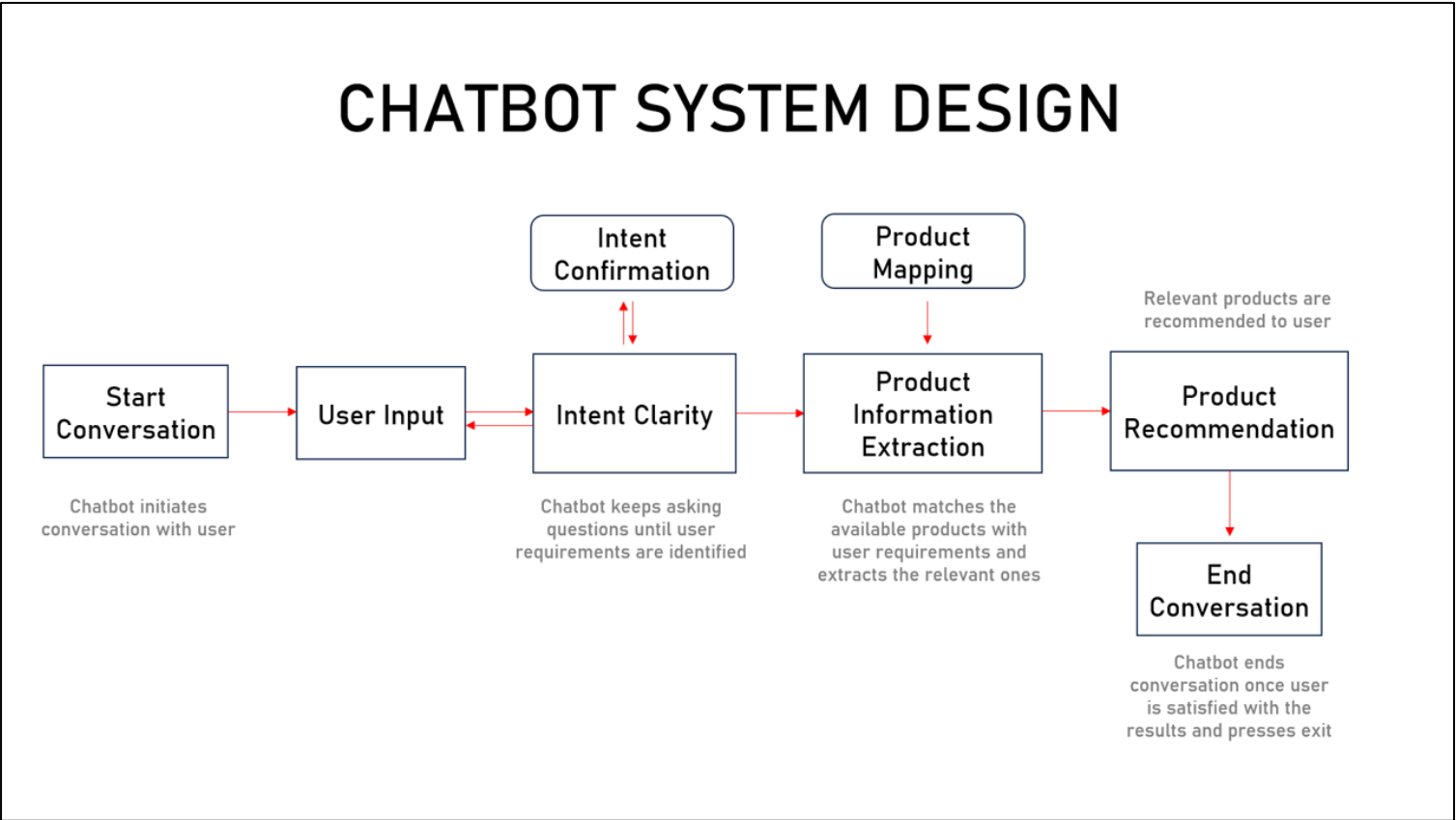
Solution Approach

A Chat based **Prototype** is provided using LLAMAINDEX , indexing and searching the Catalogue CSV File basis the User Prompt provided to the Chat Engine. The Solution is composed of various layers , detailed in the upcoming section, with necessary validations like Moderation Check before the actual prompt is based to the core AI Engine.

Why LLAMAINDEX

- Simple and most suitable for a Prototype with a specific Catalogue data (CSV)
- Lang chain suits better for versatile data sources and larger scale.
- The Search and Retrieval is streamlined in LLAMAINDEX
- Performance for Small Data set is almost same of Lang Chain , still went ahead with LLAMAINDEX for simplicity

- OVERALL SYSTEM DESIGN Flowchart
- OVERALL SYSTEM DESIGN Description
- Stagewise and Layers Implementation – A Detailed Documentation



The Overall System Design is Categorised into 3 Stages:

- **Stage 1 – Initialization , Intent Clarity and Confirmation**
 - Initializes the conversation with the user
 - Sets the context as to what to expect from the system
 - Recommended way to interact , showing the necessary attributes to be presented in the prompt
 - What not to do – like not to prompt with unwarranted messages like “ I want to kill” etc.,
 - Reads and loads the Books_gpt.csv into DataFrame for further processing
- **Stage 2 - Product Mapping and Information Extraction**
 - Intent clarity - Once The Prompt is received its checked for moderation Flags and processed further only if its Not Flagged as inappropriate by the OpenAi.Completions API
 - Intent Confirmation - The prompt is sent to an Intent confirmation layer:
 - This layer creates a Vecto store index from the Data Source CSV File
 - Sets the Context , Prevents Hallucination (temperature set to Zero) and queries the User Prompt via LLAMAINDEX’s index.query API
 - Responds a “No” , if out of context is question is asked otherwise the query response is sent for Scoring.
 - In the Scoring process basis the number of expected attributes presented in the user Query a Scoring is provided
 - Product mapping returns the actual product information from the data store Basis the scored result
- **Stage 3 - Product Recommendation**
 - The Final result is displayed in a Human Readable format
 - The Conversation is ended and asking the user for the next prompt

Stagewise and Layers Implementation – A Detailed Documentation

Stages	Layers	Methods	Descriptions
Stage - 1	Entry Point	HelpMateAiBookBot():	Entry point for the ProtoType
Stage - 1	Get Data Source	getDataSource()	Reads the CSV File Does Necessary Formatting
Stage - 1	Start Conversation & User Input	initialize_conversation()	Transparency :Tells the user they are interacting with LLM Explanation : Tells the user what they can expect from this HelpMate Bot and how their prompt should be provided for better response. Disclaimer & Warning – To avoid any unwarranted inputs like Killing etc.,
Stage - 1	Intent Clarity	moderation_check()	Checks if the User Input is Unwarranted if yes then returns the flag <ul style="list-style-type: none">- uses openai.moderations.create API- checks if the User input has any unwarranted content- if yes , it appropriately returns the Flag.
Stage - 1	Intent Confirmation	llamaindex_intent_confirmation_layer	Once moderation Check is through, the user prompt is passed to this layer which : <ol style="list-style-type: none">1. Reads the data Source and creates Documents from it.2. Creates a VectorStoreIndex for querying basis the prompt.3. Sets the Context to avoid Hallucination (temperature Zero)4. Templating – Using Binary Prompt to return No if the Query is out of context from the data set being catalogued.5. Executes the query basis the User Prompt6. Returns the Query Response and RetrieverNodes for scoring down the line.

Stagewise and Layers Implementation – A Detailed Documentation (Continued)

Stages	Layers	Methods	Descriptions
Stage -2	Product Mapping	compute_match_score	<ol style="list-style-type: none">1. Extracts the attributes from the Query2. Matches with the set of important columns in the data ('Genre', 'Publication Year', 'Page Count', 'Price (INR)')3. Calculates and returns a Score ,on the Range of 1,2 and 3, with 1 being the lowest.
Stage -2	Product Mapping (continued)	mapProductsByScoring	<ol style="list-style-type: none">1. Using RetrievedNodes from the LLAMAINDEX Search Query and the Query from user calculates the Score using the compute_match_score method explained above.2. If the Score is above the defined Threshold it is added to a dictionary.3. The scored result is returned as a Data Frame for Product information extraction.
Stage -2	Product information extraction	information_extraction	<ol style="list-style-type: none">1. The Scored Result Dataframe is JOINED with the original dataframe to return the Extracted product Information
Stage -3	Product Recommendation	showProductRecommendation	<ol style="list-style-type: none">1. The Extracted Producted information is displayed to user as Prompt Response in Human readbale and not as JSON or table format.
Stage -3	End Conversation	HelpMateAiBookBot	<ol style="list-style-type: none">1. If user types Quit Conversation is ended.

Demo Screenshots

MESSAGE PROMPTING USER TO ENTER THE QUERY IN THE TEXTBOX



HelpMateAiBookBot()

... Welcome to BookBot! Ask me for book recommendations.

You are Interacting with an AI system that will be help you to fetch books from specific catalogue and return top n matches

This bot scores your result basis these attributes :Genre,Published Year, Pages Count,Price :

For Faster and better results please ensure to include these attributes in your prompt.
Recommended way of interacting with this bot are as follows:

=====

can you recommend a Physics book'
can you recommend a book for research'
can you recommend a book of price say 1000 INR'
can you recommend a book on Physica of price around 1500 INR'
can you recommend a book on Physica of price around 1500 INR and of 450 pages'
can you recommend a book on Physica of price around 1500 INR and of 450 pages' published within 5 years

=====

Any query out of scope of the Books Catalogue will not be served
Any unwarranted questions that endangers the society, unethical etc,.will not be answered by the B

Please enter your query , or type 'exit' to quit:

User Enters the Prompt – “I need a Book in Data Science”



HelpMateAiBookBot()

... Welcome to BookBot! Ask me for book recommendations.

You are Interacting with an AI system that will be help you to fetch books from specific catalogue and return top n matches

This bot scores your result basis these attributes :Genre,Published Year, Pages Count,Price :

For Faster and better results please ensure to include these attributes in your prompt.
Recommended way of interacting with this bot are as follows:

=====

can you recommend a Physics book'
can you recommend a book for research'
can you recommend a book of price say 1000 INR'
can you recommend a book on Physica of price around 1500 INR'
can you recommend a book on Physica of price around 1500 INR and of 450 pages'
can you recommend a book on Physica of price around 1500 INR and of 450 pages' published within 5 years

=====

Any query out of scope of the Books Catalogue will not be served
Any unwarranted questions that endangers the society, unethical etc,.will not be answered by the B

Please enter your query , or type 'exit' to quit:

I need a Book in Data Scier

Chat Bot responded with all Data Science Books

Here's your current recommendation basis your prompt:

Title :Data Science in Practice

Genre:Data Science

Published in the Year:2019

Priced at :2100

of Pages :500

=====

Title :Introduction to Data Science

Genre:Data Science

Published in the Year:2022

Priced at :2000

of Pages :550

=====

Title :Data Science for Business

Genre:Data Science

Published in the Year:2020

Priced at :2300

of Pages :650

=====

Title :R for Data Science

Genre:Data Science

Published in the Year:2021

Priced at :1700

of Pages :400

=====

Title :Machine Learning Yearning

Genre:Data Science

Published in the Year:2021

Priced at :1100

of Pages :150

=====

Missing attribute verification

Please enter your query , or type 'exit' to quit:I need a Book on Research

Processing you Query=' I need a Book on Research '...

Sorry I cant process this, it may be out of scope

Please reword your Prompt it may be missing the recommended attributes:'Genre', 'Publication Year', 'Page Count', 'Price (INR)'

Please enter your query , or type 'exit' to quit:

Added Genre and the Bot Shared the recommendation

Please enter your query , or type 'exit' to quit:I need a Book on Research

Processing you Query=' I need a Book on Research '...

Sorry I cant process this, it may be out of scope

Please reword your Prompt it may be missing the recommended attributes:'Genre', 'Publication Year', 'Page Count', 'Price (INR)'

Please enter your query , or type 'exit' to quit:I need a Book on Research of Physics

Processing you Query=' I need a Book on Research of Physics '...

Sorry I cant process this, it may be out of scope

Here's your current recommendation basis your prompt:

Title :Advanced Quantum Mechanics

Genre:Physics

Published in the Year:2020

Priced at :1800

of Pages :450

=====
Title :Theoretical Physics: From Classical to Modern

Genre:Physics

Published in the Year:2019

Priced at :2500

of Pages :620

=====
Title :Fundamentals of Optics

Genre:Physics

Published in the Year:2019

Priced at :2000

of Pages :580

Intent Confirmation Layer responded a “No” when out of context question asked and appropriate response returned

=====

Query again as recommended for more

Please enter your query , or type 'exit' to quit:Who is the winner of IPL

Processing you Query=' Who is the winner of IPL '...

Sorry I cant process this, it may be out of scope

Please reword your Prompt it may be missing the recommended attributes:'Genre', 'Publication Year', 'Page Count', 'Price (INR)'

Please enter your query , or type 'exit' to quit:

Moderation Checked and Flagged “No” the appropriate response is returned and the Bot Quit the process

Please enter your query , or type 'exit' to quit:Who is the winner of IPL

Processing you Query=' Who is the winner of IPL '...

Sorry I cant process this, it may be out of scope

Please reword your Prompt it may be missing the recommended attributes:'Genre', 'Publication Year', 'Page Count', 'Price (INR)'

Please enter your query , or type 'exit' to quit:I want to kill them

Processing you Query=' I want to kill them '...

'Sorry, I cannot process this as this query seems inappropriate from Books Context (Message flagged)\n. Quitting.'



[Google Drive Folder HelpMateAI.Books](#)

Thank you

Tarriq Ferrose Khan

email: tarriqferrosekhan.com