

A Dynamic Query Framework for Research Accessibility using OpenAI and Langchain



by Louise Gabrielle Talip

advisee of Professor Reginald Neil C. Recario



Made with Gamma

Table of Contents

- 1 — Introduction
- 2 — Review of Related Literature
- 3 — Methodology

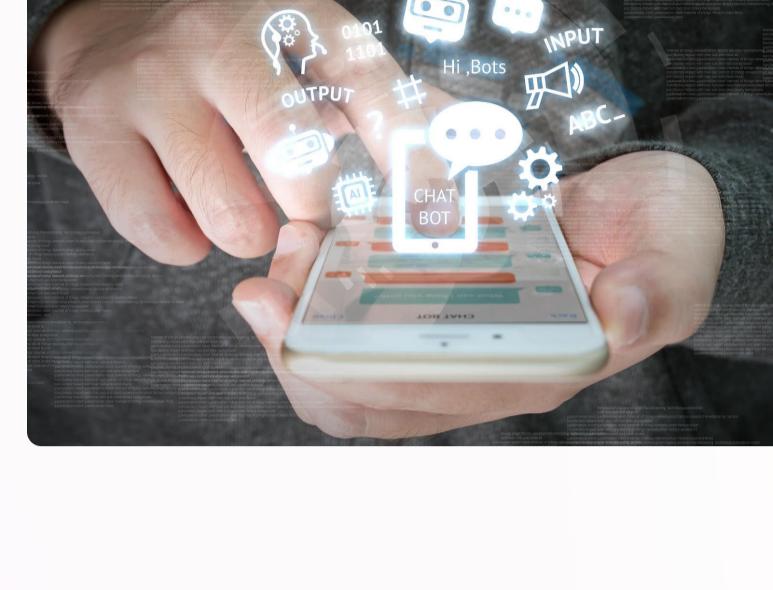
Introduction

Chatbots

Chatbots are intelligent systems that can hold conversations with humans using natural language.

From hard-coded response systems to more sophisticated Artificial Intelligence-driven advancements.

They can now be trained or prompted to understand a wider scope of topics with the aid of appropriate datasets, such as in the case of OpenAI's ChatGPT

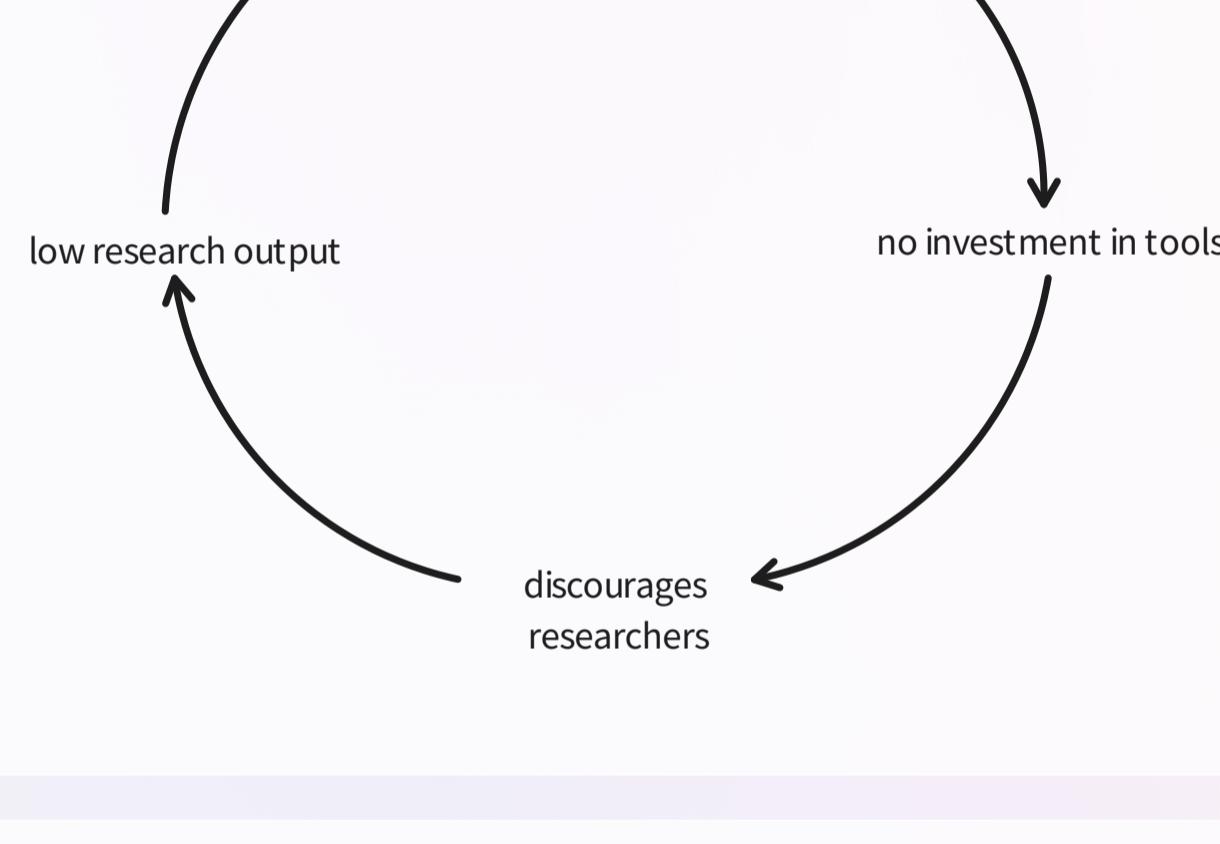


This paper explores the potential and significance of developing a chatbot-driven solution aimed at addressing inquiries and providing recommendations related to academic papers.

Statement of the Problem

There are many factors as to why the Philippines' research output has been consistently low over the years.

- Lack of research culture
- Lack of government support
- Lack of infrastructure and accessible tools
- Time-intensive task
- Steep learning curve for new researchers
- Heavy barrier to entry



Significance of the Study

The academic landscape in the Philippines is already niche, deterring many potential researchers from pursuing these paths.

Developing a tool that mitigates the barriers becomes pivotal in revitalizing researchers' interest.

Existing applications like **ChatPDF** and the **AI Research Assistant** by Elicit exhibit limitations:

- handling only one PDF at a time
- having restricted datasets
- often excluding Philippines-based papers

In contrast, a tool incorporating custom data offers accessibility to local resources in the university library.

High costs associated with proprietary models hinder accessibility, especially for entry-level scholars with limited purchasing power.

Utilizing Langchain improves scalability and optimizes efficiency by dividing tasks into smaller subtasks and targeting specific parts only containing required information.

Especially in countries like the Philippines where resources such as GPUs are limited, it is not ideal to train a custom model on the data. Langchain is crucial in enhancing accessibility even when using existing models like OpenAI.

Introduction

Objectives of the Study

The primary focus of this study is to **create a research assistant chatbot** to aid researchers in understanding scientific publications in the Institute of Computer Science of the University of the Philippines Los Banos.

1 Analyze and implement Langchain and OpenAI

to enhance the **chatbot's conversational understanding capabilities** in complex academic contexts, addressing the challenge of natural language processing

2 Utilize and employ Langchain, OpenAI, and Chainlit functionalities

to devise efficient methods for content extraction from thesis papers, focusing on

- **the retrieval of correct and simplified answers**
- **creating accessible summaries, keywords, and citations**

to help researchers comprehend academic content effectively

3 Evaluate and improve the accuracy of information retrieval

from academic papers with Chainlit's Feedback feature

4 Assess the chatbot's effectiveness

through Post-Task Evaluation survey and System Usability Scale

Review of Related Literature

Language Models and OpenAI

Model	Provider	Open-Source	Speed	Quality	Params	Fine-Tuneability
gpt-4	OpenAI	No	★★★	★★★★	-	No
gpt-3.5-turbo	OpenAI	No	★★★	★★★★	175B	No
gpt-3	OpenAI	No	★★★	★★★★	175B	No
ada, babbage, curie	OpenAI	No	★★★	★★★★	350M - 7B	Yes
claude	Anthropic	No	★★★	★★★★	52B	No
claude-instant	Anthropic	No	★★★	★★★★	52B	No
command-xlarge	Cohere	No	★★★	★★★★	50B	Yes
command-medium	Cohere	No	★★★	★★★★	6B	Yes
BERT	Google	Yes	★★★	★★★★	345M	Yes
T5	Google	Yes	★★★	★★★★	11B	Yes
PaLM	Google	Yes	★★★	★★★★	540B	Yes
LLaMA	Meta AI	Yes	★★★	★★★★	65B	Yes
CTRL	Salesforce	Yes	★★★	★★★★	1.6B	Yes
Dolly 2.0	Databricks	Yes	★★★	★★★★	12B	Yes

Large Language Models (LLMs)

- are a groundbreaking development in the field of AI
- a class of language models that have demonstrated exceptional performance in natural language processing tasks both in **understanding** and **generation**

* Parameters for gpt-4: **1.76 trillion**

from MindsDB

OpenAI Models

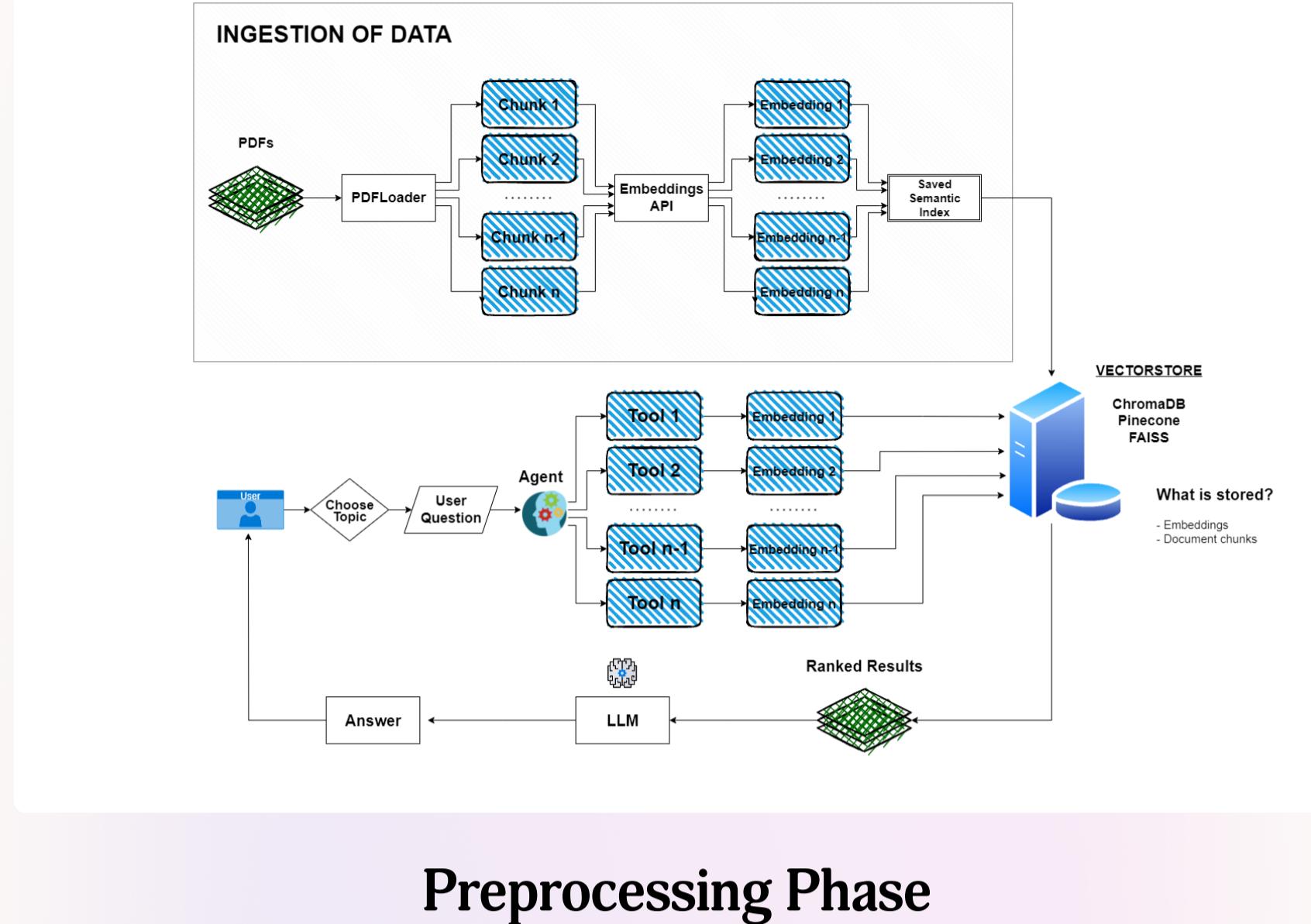
Model Name	Tokens	Training Dataset	Pricing (Input)	Pricing (Output)
gpt-4-1106-preview	128,000	Up to Apr 2023	\$0.01 / 1K	\$0.03 / 1K
gpt-3.5-turbo-1106	16,385	Up to Sep 2021	\$0.0010 / 1K	\$0.0020 / 1K
gpt-3.5-turbo	4,096	Up to Sep 2021	\$0.002 / 1K	\$0.002 / 1K
gpt-3.5-turbo-16k	16,385	Up to Sep 2021	\$0.003 / 1K	\$0.004 / 1K
gpt-3.5-turbo-instruct	4,096	Up to Sep 2021	\$0.0015 / 1K	\$0.0020 / 1K

from OpenAI

Methodology

based on Topsakal and Akinci (2023) and Sreeram and Sai (2023)

Theoretical Framework



Preprocessing Phase

1

2

3

Data Collection

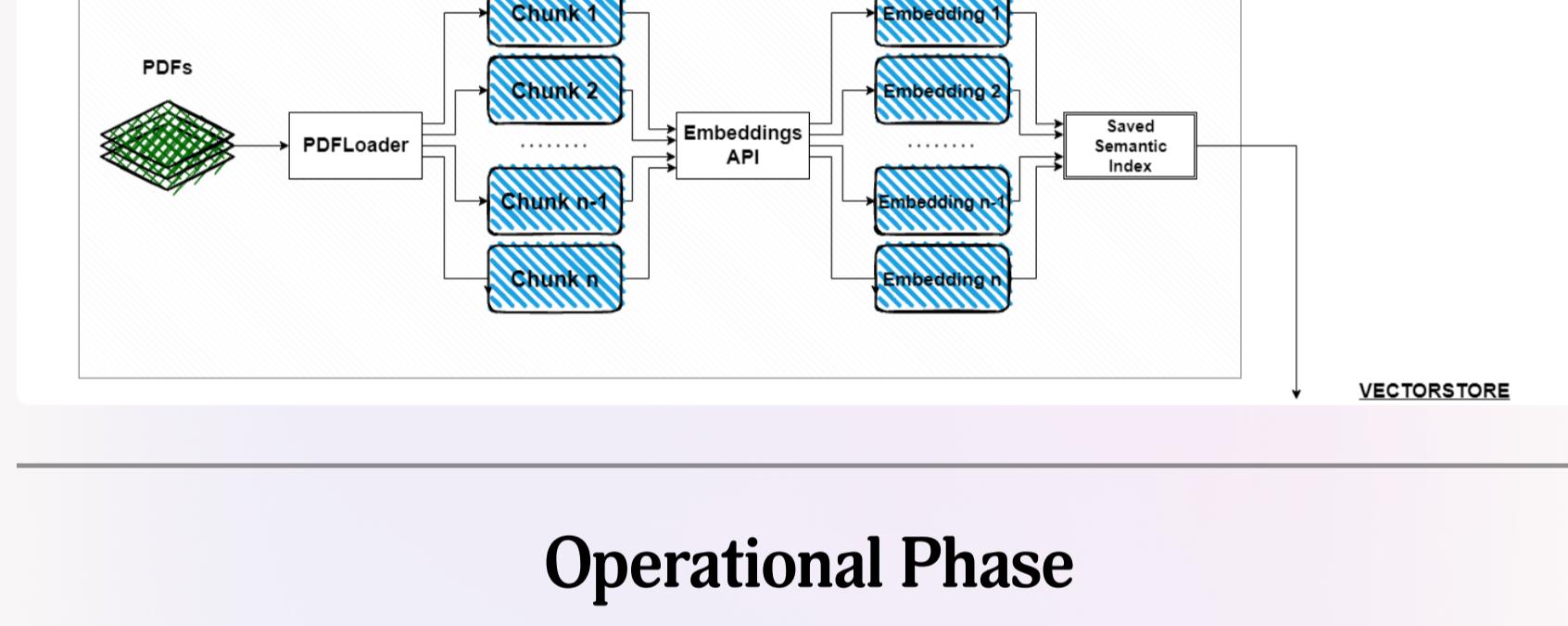
of Special Problems and Thesis manuscripts from the Institute of Computer Science (archives from January 2018 to December 2023)

PDF Chunking and Text Splitting

of lengthy documents into smaller, more digestible sections, enabling efficient handling and analysis

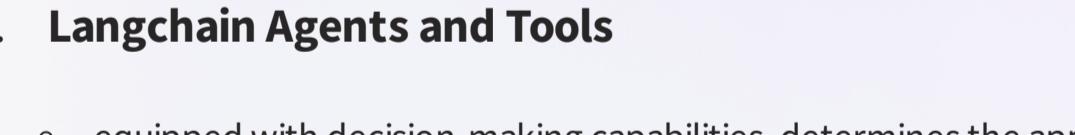
Database Storing

of the extracted text segments that are converted into structured formats and is stored within a vectorstore



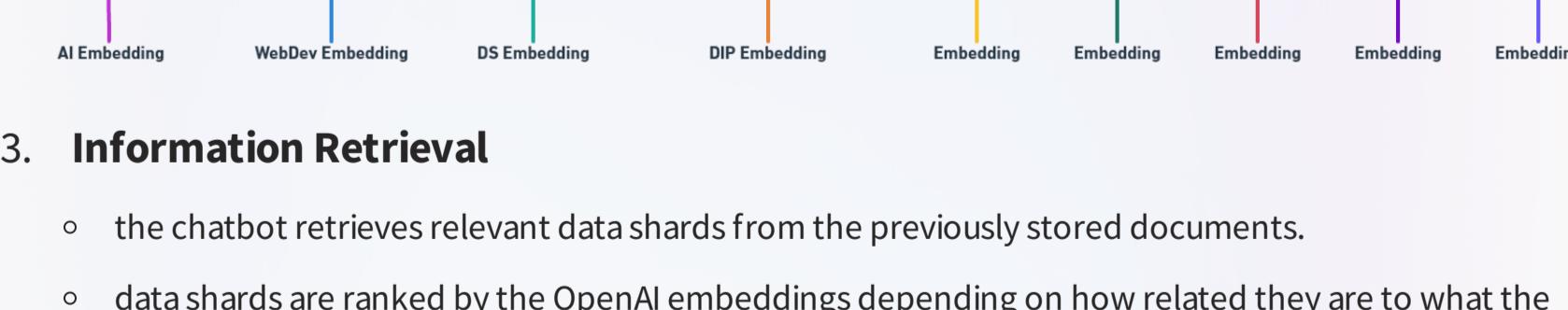
Operational Phase

1. User Selection of Topic and Query



2. Langchain Agents and Tools

- equipped with decision-making capabilities, determines the appropriate tool to employ and dataset to access for acquiring the proper response

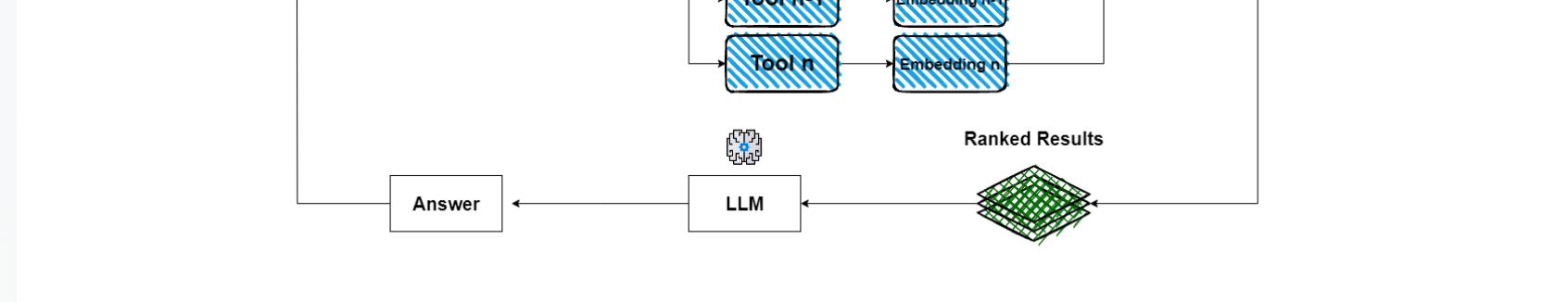


3. Information Retrieval

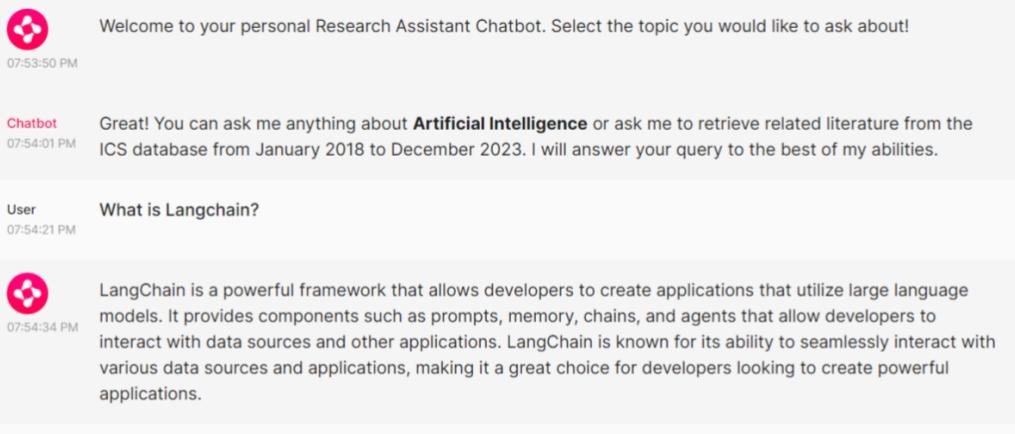
- the chatbot retrieves relevant data shards from the previously stored documents.
- data shards are ranked by the OpenAI embeddings depending on how related they are to what the user is querying for

4. Formulation of Response

- the information retrieved from the embeddings is then passed to the LLM to process this data and formulate a comprehensive, properly formatted response tailored to address the user's query



Types of Queries



Welcome to your personal Research Assistant Chatbot. Select the topic you would like to ask about!

07:53:50 PM

Chatbot Great! You can ask me anything about **Artificial Intelligence** or ask me to retrieve related literature from the ICS database from January 2018 to December 2023. I will answer your query to the best of my abilities.

07:54:01 PM

User What is Langchain?

07:54:21 PM

Chatbot LangChain is a powerful framework that allows developers to create applications that utilize large language models. It provides components such as prompts, memory, chains, and agents that allow developers to interact with data sources and other applications. LangChain is known for its ability to seamlessly interact with various data sources and applications, making it a great choice for developers looking to create powerful applications.

07:54:34 PM

Direct Answer Queries

These questions seek specific, concise information. The chatbot engages in a conversational manner, delivering the requested information in a user-friendly format. Ideally, with references to the paper.

Types of Queries

List-based Queries

These are inquiries requesting lists of papers. The chatbot responds with a list of research paper titles and the relevant information related to the queried topic

Welcome to your personal Research Assistant Chatbot. Select the topic you would like to ask about!

08:05:47 PM

08:06:08 PM Great! You can ask me anything about **Artificial Intelligence** or ask me to retrieve related literature from the ICS database from January 2018 to December 2023. I will answer your query to the best of my abilities.

User 08:06:26 PM Can you give me papers related to research productivity in the Philippines?

08:06:26 PM Sure! I found these research papers related to your query.

1. Crisostomo, L. C., Guido, R. M., & Villanueva, M. A. (2022). **Business Technology and innovation research among Southeast Asian countries: Examining Philippine performance in Technoinnovation Research.** 2022 IEEE 14th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM).
<https://doi.org/10.1109/hnicem57413.2022.10109275>
2. Guido, R. M. D., & Orleans, A. V. (2020). **Philippine Research Productivity in Education Research: A Comparative Performance in Southeast Asia.** Asia Pacific Journal of Multidisciplinary Research, 8(4), 76–90.
3. Guido, R. M., & Mangali, G. R. (2022). **Research productivity as performance dynamics of Pisa among Southeast Asian countries.** 2022 IEEE 14th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM).
<https://doi.org/10.1109/hnicem57413.2022.10109453>

[Regenerate Answer](#) [Retrieve More Papers](#)

Fine-Tuning and Evaluation

Prompt Engineering

- a technique that involves the use of task-specific hints, known as prompts, to adapt large pre-trained models to new tasks.
- it enables the ability to perform predictions based solely on prompts without updating model parameters
- Even altering a single word within a prompt can significantly impact the AI's output.



"Give me the list of restaurants"

"Give me the complete list"

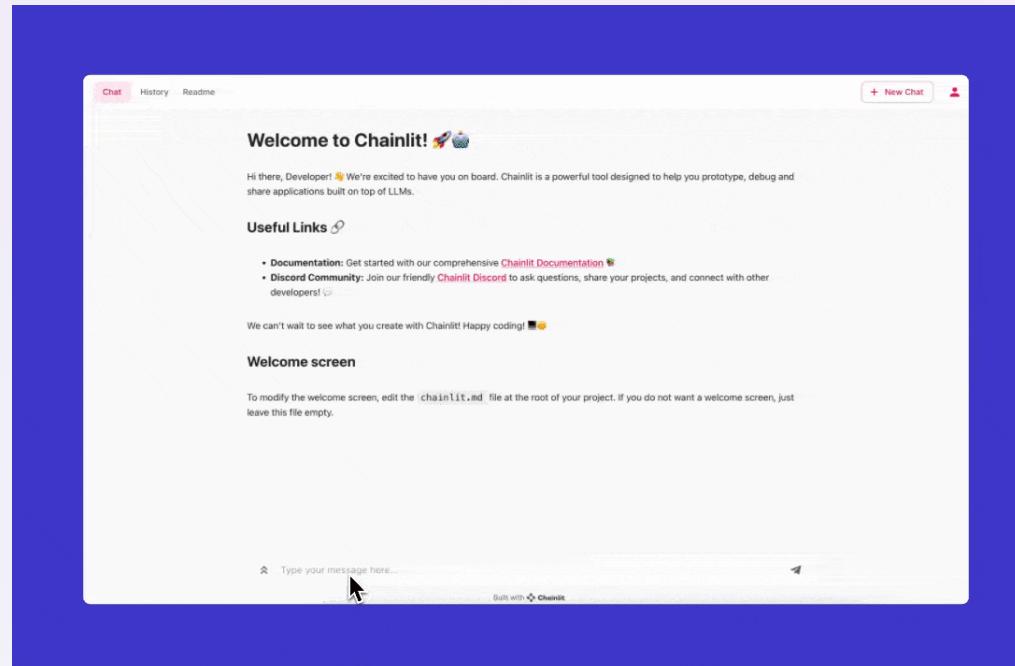
"Give me the whole list"

"List out all the ..."

"Be thorough when giving the total,
unabridged list"

Chainlit's Human Feedback Feature

- Allows users to provide direct feedback on the interaction, which can be used to improve the performance and accuracy of the system.



from: Chainlit documentation

- **Feedback scores** enable objective measurement and comparison of different agent versions. By collecting this, can create a dataset that can be used to improve the system's accuracy.
- **Feedback** allows for direct model training and fine-tuning

Surveys

- Post-Task Evaluation survey
 - Content, Accuracy, Format, Ease of use, and Timeliness
- System Usability Scale (SUS)
 - Strongly Agree to Strongly disagree
 - Examples:
 - I think that I would like to use this system frequently
 - I found the system unnecessarily complex
 - I thought the system was easy to use.



Thank you!

Louise Gabrielle Talip

lltalip@up.edu.ph

2018-11154