ANALYTIBOT: UNLEASHING LLM, LANGCHAIN, AI-ML AND OPENAI FOR THE ANALYSIS OF RESEARCH ARTICLES.

Alekya Ramani Nalagandla – 20BPS1097
Dr.Pradeep Kumar T S
School of Computer Science and Engineering
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)
VIT CHENNAI STARTUP AND RESEARCH FOUNDATION

MOTIVATION / INTRODUCTION

Researchers are excited by the flood of new research, but struggle to keep up with all the information. Traditional methods for finding what they need are slow and can miss important discoveries. AnalytiBot is here to help! It uses powerful AI to make data analysis faster and easier, unlocking the hidden potential within this sea of knowledge.

OBJECTIVES

- **Helping Hand for Research Data:** AnalytiBot is a system built to assist researchers in making sense of unstructured research data, like PDFs, Images, CSV files and webpages.
- **Superpowered Search:** It leverages cutting-edge Natural Language Processing (NLP) and Large Language Models (LLMs) to help researchers explore and find what they need from this data.
- **Easy to Use:** AnalytiBot offers a user-friendly interface that allows researchers to interact with data in a conversational way and see information visualized clearly.
- **Digging Deeper:** Beyond exploration, it provides advanced functionalities for researchers to analyze data, including statistical analysis for data in spreadsheets (CSV format).

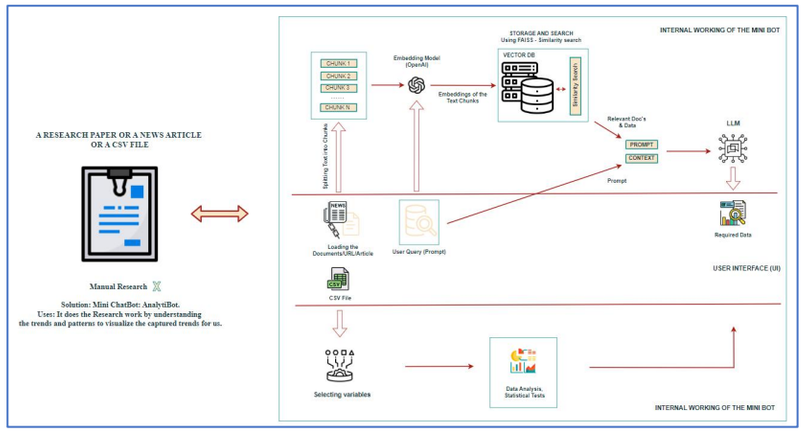
SCOPE OF THE PROJECT

- **Challenge:** Researchers drown in unstructured data (PDFs, webpags, CSV files, Images, etc.). AnalytiBot tackles this.
- **AI-powered Analysis:** Leverages NLP and LLMs to help researchers explore and analyze unstructured research data.
- **Intuitive Interface:** Conversational interaction and clear data visualization make data exploration a breeze.
- **Advanced Functionality**: Statistical analysis (e.g., linear regression) on CSV data unlocks deeper insights.

METHODOLOGY

- AnalytiBot cuts through the clutter of research data! It uses Natural Language Processing (NLP) to understand the meaning in PDFs, articles, and webpages.
- But AnalytiBot doesn't stop there. It also tackles data in spreadsheets (CSV files) and uses powerful Large Language Models (LLMs) to analyze everything. These LLMs are like super-readers, finding key information from mountains of text. To make searching super-fast, AnalytiBot uses vector embeddings - a fancy way of turning text into a code that helps find relevant information quickly.
- And the best part? You can chat with AnalytiBot in a normal way, asking questions and getting clear answers. This combo of NLP, LLMs, vector embeddings, and a user-friendly chat makes research a breeze, helping you discover hidden gems in the data sea!

ARCHITECTURE



AnalytiBot leverages NLP for unstructured data (PDFs, webpages) and utilizes LLMs & vector embeddings for efficient retrieval. It integrates statistical analysis for CSV data, offering a comprehensive research assistant.

RESULTS

Research Findings:	
Model Performance Analysis:	GPT-3.5-turbo is observed to be more accurate and faster than GPT-3.5-turbo-16k.
Impact of Processing Multiple URLs:	Questions related to content from URLs are answered faster when prompted for URL1 compared to URL2, and similarly, URL2 is faster than URL3, and so on. This suggests that there might be a correlation between the processing time and the order of URLs.
Significance of Training Data Quality and Quantity:	It is observed that the more trained the data is, the faster and more accurate the answers are. This highlights the importance of training data quality and quantity for model performance.
Effect of Question Length on Answer Accuracy:	Longer word count questions produce more accurate answers compared to smaller word count queries. This suggests that the model might require more context to generate accurate responses.
Influence of PDF Document Size on Response Speed:	Smaller PDF documents result in faster answers when prompted with questions. This indicates that document size might impact the processing time.

CONCLUSION

The research data flood is real! AnalytiBot is your AI-powered life raft. It uses smarts (NLP & LLMs) to understand research (PDFs, webpages, even CSVs!), retrieves info fast (vector embeddings!), and lets you chat for easy exploration. Dive deeper, discover more, with AnalytiBot!

CONTACT DETAILS

MAIL ID: alekya.ramani2020@vitstudent.ac.in
MOBILE NO: 6301949698

REFERENCES

- Zemčík, M. T. (2019). A brief history of chatbots. DEStech Transactions on Computer Science and Engineering, 10.
- Luo, B., Lau, R. Y., Li, C., & Si, Y. W. (2022). A critical review of state- of- the- art chatbot designs and applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12(1), e1434.