Tiger Analytics

Tiger Analytics equipped a top pharmaceutical company to redefine the data retrieval process of 3K documents with a gen AI model

Tiger Analytics implemented a Gen AI model-based solution for designing an automatic information retrieval system (similar to ChatGPT) with 65% accuracy. The users could instantaneously get answers for any of their input queries - anytime and anywhere. It helped maximize the value of the client's market research documents.

The Background

Our client is a Fortune 500 Biopharmaceutical company developing and producing medicines, vaccines, biological therapies, animal health products, and more. They have a huge repository of research documents in PDF, PPT, and DOCX formats with business-critical information. The existing document retrieval system is a simple search engine portal requiring extensive manual efforts. Hence, the client wanted a secure solution to directly extract relevant content and ensure that answers contained links to the exact document.

Key Challenges

Unstructured data: Over 3000 documents/files came in unstructured .json format, with incoherent tabular data, charts and graphs.

Ineffective sentence segmentation: No proper sentence segmentation was used to distinguish between sentences and paragraphs.

Unpredictable data accuracy: There was a lack of original/raw PDFs to confirm data accuracy.

Low relevant data volume: A very low volume of data was provided to generate the answers from the model.

Our Solution

Tiger Analytics helped build a safe Gen AI-based model to address all the client's data retrieval challenges. The team leveraged Large Language Models (LLMs) with Reinforcement Learning from Human Feedback (RLHF) to create a chatbot user interface. It helped the users obtain summarized responses for their queries in a Q and A format.

Scope Identification and Prioritization was the first step, as the team defined a subset of Market Research documents and questions for POC. The second step was **Data Gathering and Exploration**, in which data was inputted, comprising information on PMR clinical trial data with key details on various drugs from different physicians across countries in the US market.

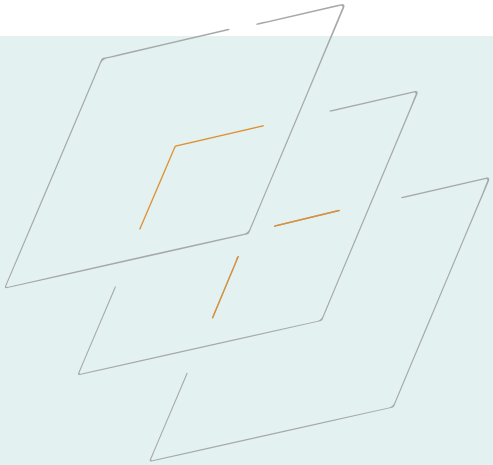
The third step was **Data Pre-Processing**, in which the team evaluated the source JSON data and resolved challenges such as incoherent charts and figure data by manually examining approximately 10 of 3000 documents. It helped identify data patterns to create a rule filter in model codes.

The fourth step was the **Extrapolation of SME Q and A pairs** to generate more data for model processing. The team fine-tuned the Language Models with a paraphraser model.

LLM Selection was the fifth step, in which the team experimented with multiple models before choosing GPT 2 as the final base LLM. The 3000 documents were exposed to this model by the Sequence-to-Sequence Modeling Method for the model to learn the context and sequence from the client's data. In step six, a **Retriever System** was built to obtain the correct context of user query to pass to the LLM model and questions to generate accurate answers. Using a BM25 retriever, the terms from the query in the document were matched to identify the top 10 documents where the answer is present.

Step seven saw the **Colossal AI Framework for RLHF and Distributed Model Training** being used to fine-tune the model on Question-Context-Answer (QCA) objectives, train reward models based on the SME-chosen answers, and update the fine-tuned model. The team also used GPT 3.5 to convert the final answers (text/ phrases) from the model engine to proper sentences.

Finally, step eight, which focused on **User Interface Development**, saw the team developing and deploying an application for users to query responses from the model.



Tech Stack

/AWS

/SageMaker (Development)

/ S3 Bucket (Storage)
/ EC2 (Deployment Environment)
/GPT2 & GPT3.5

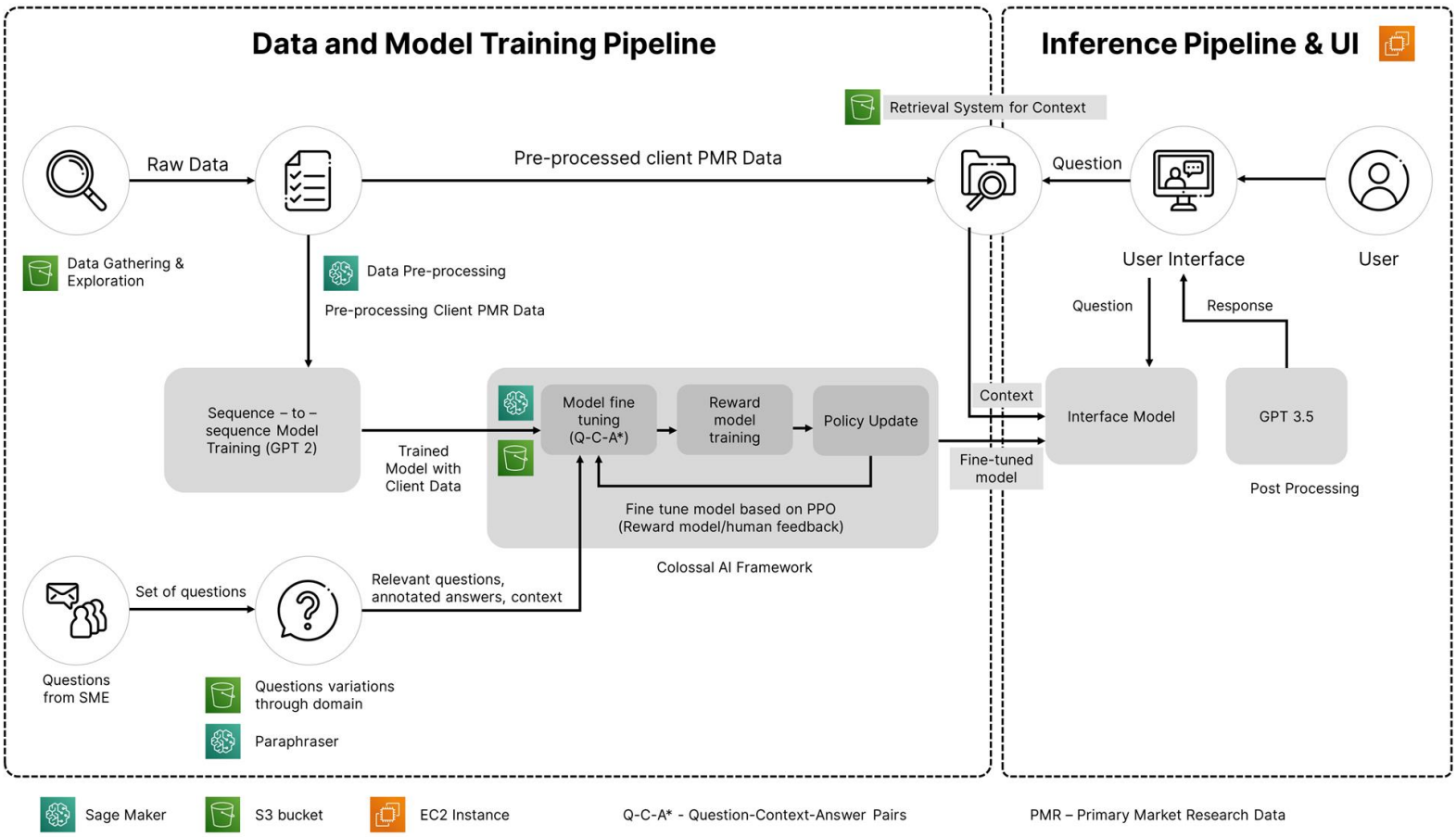
/Python Libraries

/ Haystack (Back-end Retrieval System)
/ Sentence Transformer Models

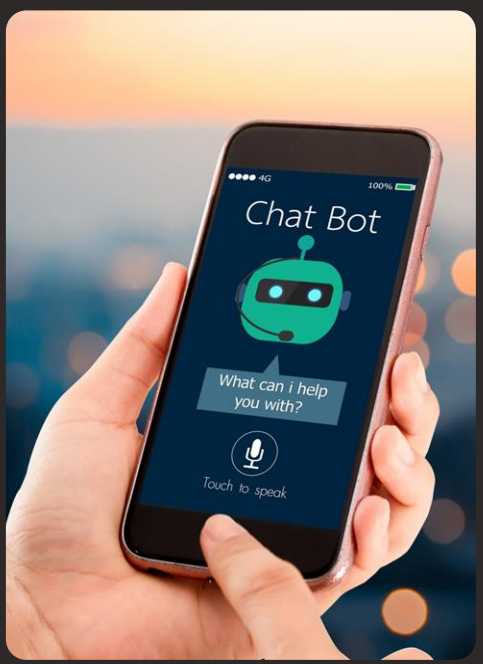
/ Fast API (Backend User Interface)

/ React JS (Frontend User Interface)

Solution Architecture:



Value Delivered



- The query response time for US markets was reduced by nine seconds by developing an AI-based Q&A chatbot.
- 65% accuracy was achieved in the final model – trained and fine-tuned through the supervised method, with human feedback and iterative retraining.
- The user experience was similar to ChatGPT, enabling users to input queries, receive summarized answers, and regenerate answers.

  The answer contained links to the internal portal, redirecting users to the exact document.

About Us

Tiger Analytics is a global leader in AI and analytics, helping Fortune 1000 companies solve their toughest challenges. We offer full-stack AI and analytics services & solutions to help businesses achieve real outcomes and value at scale. We are on a mission to push the boundaries of what AI and analytics can do to help enterprises navigate uncertainty and move forward decisively. Our purpose is to ***provide certainty to shape a better tomorrow.***

Being a recipient of multiple industry awards and recognitions, we have 4000+ technologists and consultants, working from multiple cities in 5 continents.

www.tigeranalytics.com

US | UK | Canada | India | Singapore | Australia