10/29 Client Meeting

Nexteer (AI Bots) Capstone Team

Agenda

Reiterate Next Steps from Midterm Presentation

Progress in the Past Week

Goals for Next Week

Deliverable

Questions

Next Steps from Midterm Presentation

1. Explore automated metadata generation based on proposed schema
2. Conduct data batching experiment
3. User query similarity search with metadata
4. Benchmark open-source Llama 3.2 performance on classification of user queries
5. Test LLM model performance with more metadata in prompts (JSON key-value pairs)
6. Document all code and files

Progress in the Past Week

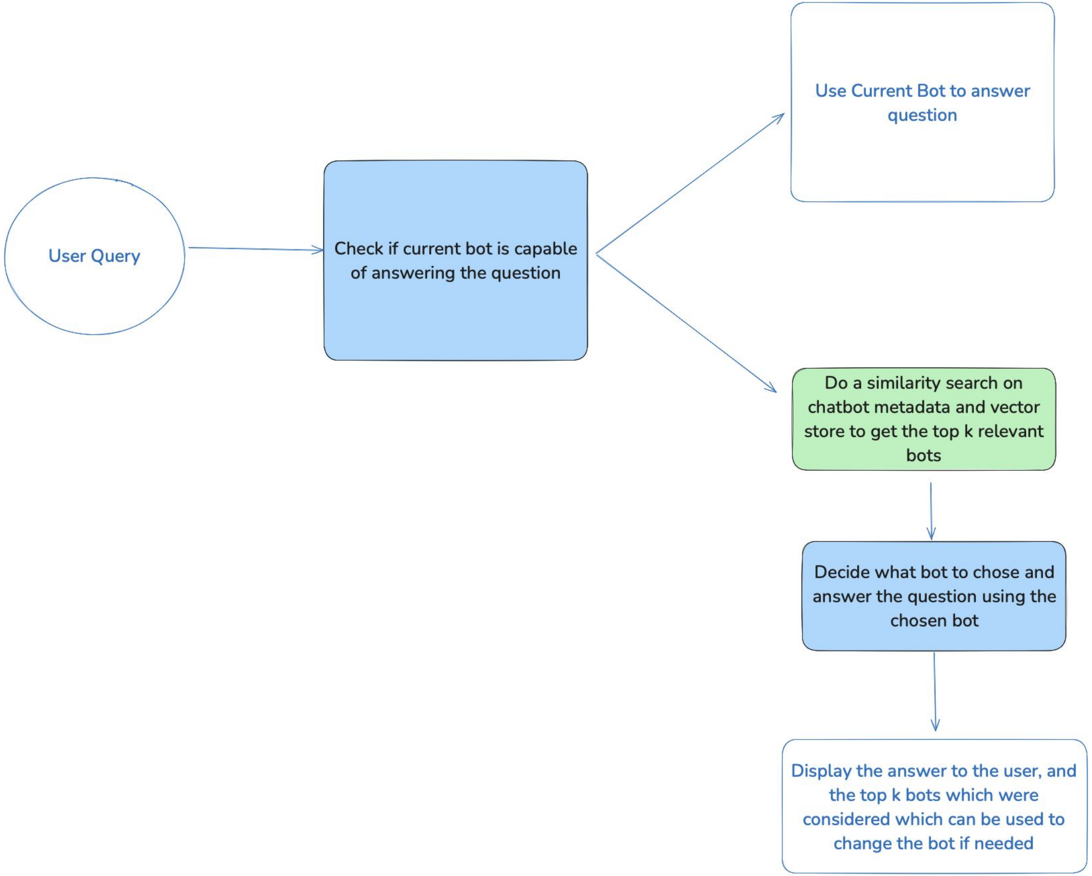
Data Team

1. 2nd draft of our POC schema: https://dbdiagram.io/d/KBChat-66f624cc3430cb846cc86bf4
2. Making a dataset from the chat history

Progress in the Past Week

1. Implemented a basic triage agent
2. Created dummy documents and metadata for three potential bots.

Triage Agent - First Implementation



User Query

Check if current bot is capable of answering the question

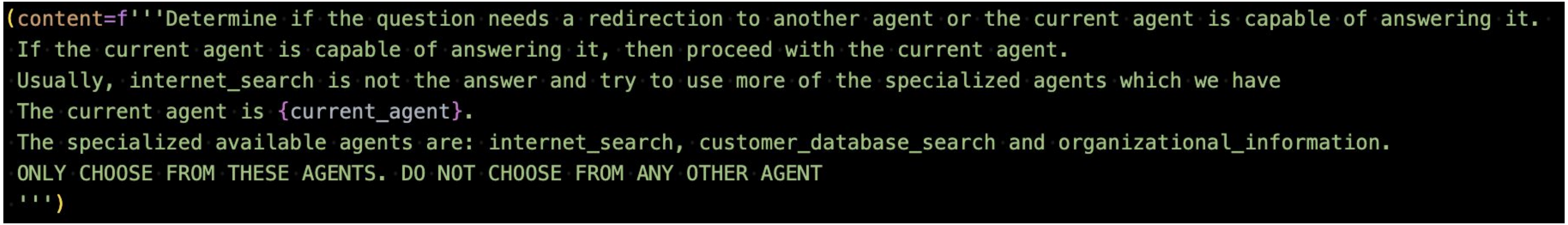
Use Current Bot to answer question

Do a similarity search on chatbot metadata and vector store to get the top k relevant bots

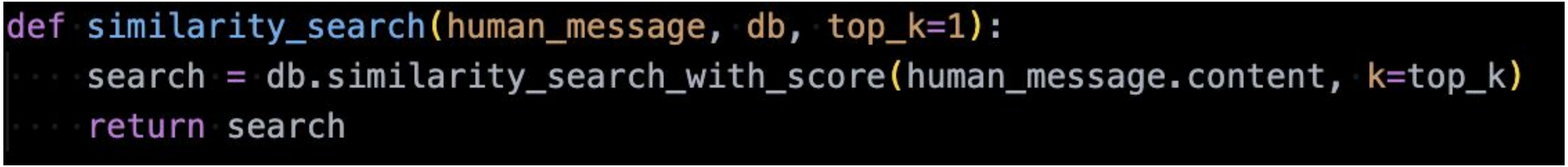
Decide what bot to chose and answer the question using the chosen bot

Display the answer to the user, and the top k bots which were considered which can be used to change the bot if needed

Triage Agent - First Implementation

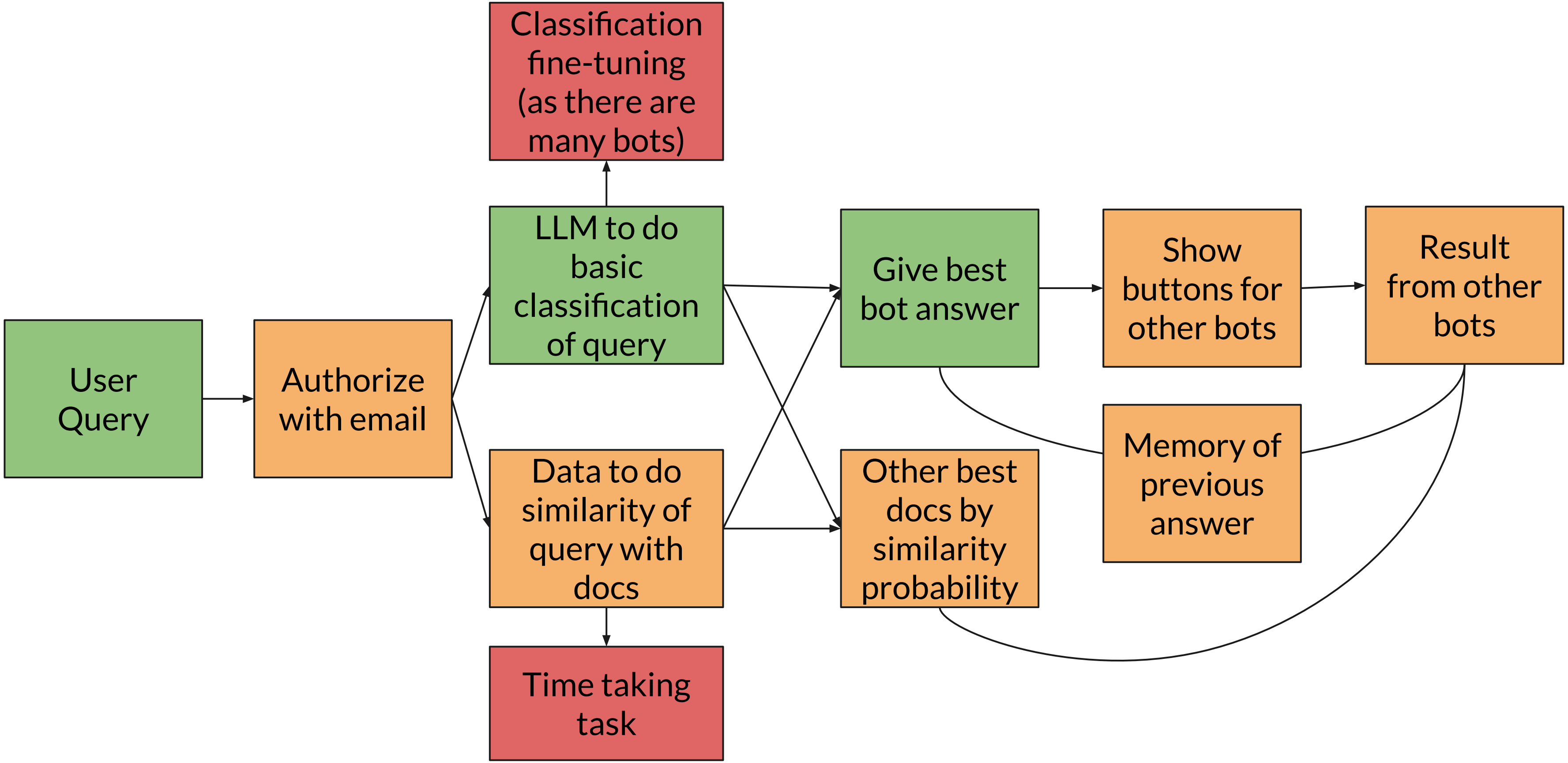


```
(content=f'''Determine if the question needs a redirection to another agent or the current agent is capable of answering it.
 If the current agent is capable of answering it, then proceed with the current agent.
 Usually, internet_search is not the answer and try to use more of the specialized agents which we have
 The current agent is {current_agent}.
 The specialized available agents are: internet_search, customer_database_search and organizational_information.
 ONLY CHOOSE FROM THESE AGENTS. DO NOT CHOOSE FROM ANY OTHER AGENT
 ''')
```



```
def similarity_search(human_message, db, top_k=1):
    search = db.similarity_search_with_score(human_message.content, k=top_k)
    return search
```

Modeling Pipeline

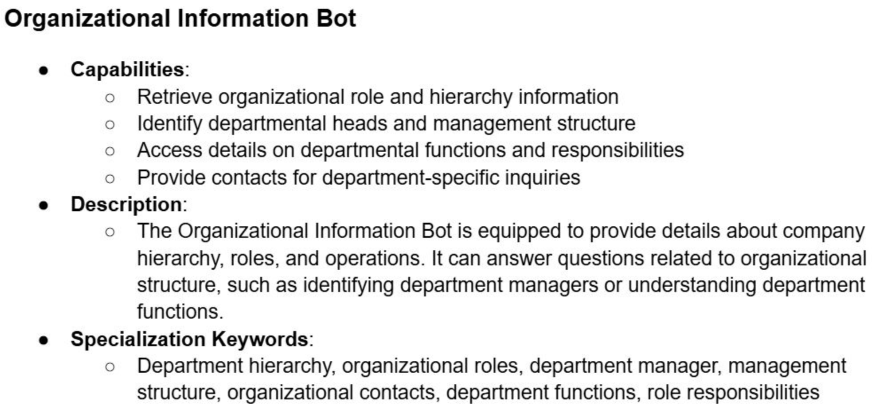

Classification fine-tuning (as there are many bots)
User Query
Authorize with email
LLM to do basic classification of query
Data to do similarity of query with docs
Time taking task
Give best bot answer
Other best docs by similarity probability
Show buttons for other bots
Result from other bots
Memory of previous answer

Auto Metadata Generation

Goal: To be able to generate metadata for each specialized chatbot based on given documents.

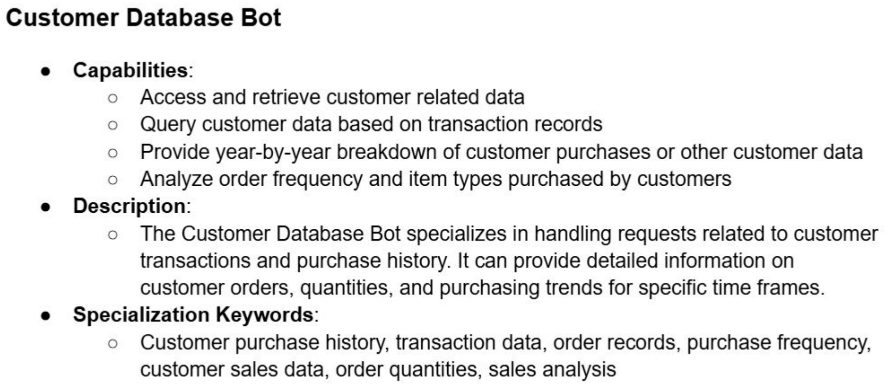
We did the following:

1. Created dummy documents that a specialized bot would use as reference.
2. Created manual metadata that we hope to auto generate from an LLM based model as follows:



Organizational Information Bot

- **Capabilities**:
  - Retrieve organizational role and hierarchy information
  - Identify departmental heads and management structure
  - Access details on departmental functions and responsibilities
  - Provide contacts for department-specific inquiries
- **Description**:
  - The Organizational Information Bot is equipped to provide details about company hierarchy, roles, and operations. It can answer questions related to organizational structure, such as identifying department managers or understanding department functions.
- **Specialization Keywords**:
  - Department hierarchy, organizational roles, department manager, management structure, organizational contacts, department functions, role responsibilities



Customer Database Bot

- **Capabilities**:
  - Access and retrieve customer related data
  - Query customer data based on transaction records
  - Provide year-by-year breakdown of customer purchases or other customer data
  - Analyze order frequency and item types purchased by customers
- **Description**:
  - The Customer Database Bot specializes in handling requests related to customer transactions and purchase history. It can provide detailed information on customer orders, quantities, and purchasing trends for specific time frames.
- **Specialization Keywords**:
  - Customer purchase history, transaction data, order records, purchase frequency, customer sales data, order quantities, sales analysis

Goals for Next Week

1. Test auto metadata generation from documents
2. Improve user query similarity search with metadata and dummy documents
3. Implement better information gathering in triage agent and use the above similarity search to reduce latency

Deliverable

POC:

1. A model that understands user query domain and skills
2. Redirection logic/demo of redirecting user to specialized chatbots

Deliverable items:

1. Code (for testing model/redirection logic)
2. Documentation/Report

Questions