

Personal AI Assistant Using LangGraph and LangChain

Vishal Vardhan Adepu — 334003567

November 3, 2024

Abstract

This report describes the design and implementation of a personal AI assistant built using LangGraph and LangChain. The assistant provides functionalities such as sending emails, reading and answering questions from PDF files, scheduling meetings, and searching the Internet. Privacy concerns are mitigated through a local LLM model to ensure that private information is handled securely. This report details the features, implementation, challenges, and future work of the project.

1 Problem Statement

The goal of this project is to create a personal AI assistant that can perform the following tasks:

- Write and send emails on behalf of the user (1pt)
- Read multiple PDF files and answer questions (1pt)
- Schedule meetings for the user (2pt)
- Search the Internet (2pt)
- Ask clarifying questions when uncertain (2pt)

Key Requirement: The assistant must not leak private information and should use a local LLM for sensitive data. Public LLM APIs may be used for non-private data.

2 Implementation

The assistant is implemented as a command-line application using the `click` library in Python. The following components and tools were used:

2.1 Workflow

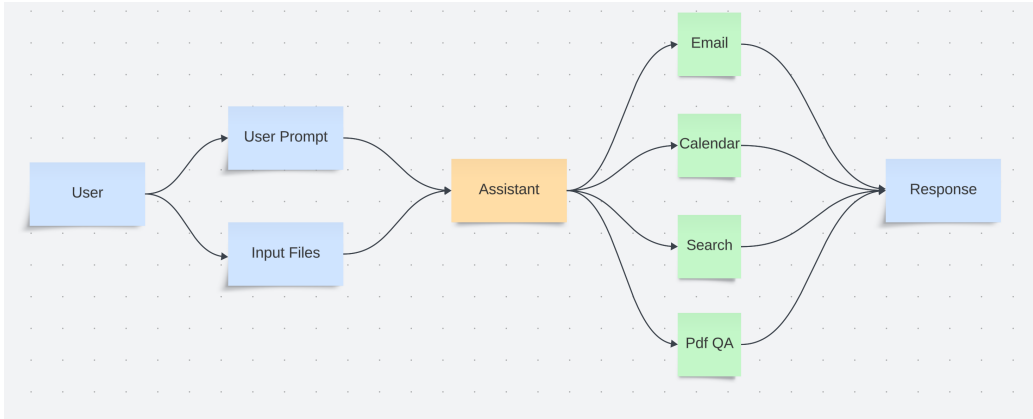


Figure 1: Workflow diagram of the personal AI assistant

2.2 Features and Tools

- **Gmail Tool:** Implemented using the `composio` library for composing and sending emails.
- **Search Tool:** Uses the DuckDuckGo API from LangChain to perform Internet searches.
- **PDF QA Tool:** Built using LangChain and backed by a Chroma vector database for efficient question answering. The assistant can also answer questions about previously uploaded PDFs, as the data is persisted in the local file system using ChromaDB.
- **Calendar Tool:** Also implemented using the `composio` library for scheduling meetings.

- **Privacy Checks:** Utilized the Mistral-NeMo 12B model running locally via Ollama. The local LLM first checks for any private information in the user input. If private data is detected, the assistant flags the information and requests user consent before proceeding. If consent is denied, the input is erased from memory.

2.3 Privacy Mechanism

The assistant uses a local LLM model to identify and handle private information securely:

1. User input is analyzed for private information using the local model.
2. If private data is detected, the assistant informs the user and asks for consent to share it with a public LLM.
3. If consent is given, the information is shared; otherwise, the input is erased and the user is prompted again.

3 Clarifying Questions

The assistant asks for additional details when prompts are vague. For example, in the Gmail feature demo, the assistant requests clarification if the input lacks necessary information. This behavior is consistent across all tools.

4 Screenshots

4.1 Gmail Feature

```
(programming-llms) C:\Users\vva\Documents\MS\colleges\Texas A&M University\post-admit-docs\in-college\semester-3\csce-689-programming-llms\hw\3-hw>python script.py
Welcome to the CLI Chat!
Type 'exit' or '\q' to end and '\c' to clear the chat.

User (end with an empty line):

send an email

Assistant>
Okay, let's draft an email. Please provide the following details:

1. Recipient's email address
2. Subject line
3. Body content
4. Any file attachments (file paths)
5. If the body is HTML formatted (True/False)

I'll put together a draft email for your review before sending. Please provide the necessary information.

User (end with an empty line):
subject: Hello from vishal vardhan adepu
body: Howdy, How are you?. Best regards, Vishal Vardhan Adepu, contact: 9475816842
- recipient: vishalvardhan418@gmail.com

!!PRIVATE INFO FOUND!!
Private Information:
* Email address: 'vishalvardhan418@gmail.com'
* Phone numbers: '9475816842'

Do you want to proceed sending this info to a public LLM? (y/n): : y
Assistant>
Here is the draft email for your review:

To: vishalvardhan418@gmail.com
Subject: Hello from vishal vardhan adepu

Howdy, How are you?

Best regards,
Vishal Vardhan Adepu
Contact: 9475816842

Please confirm if you would like me to send this email as is, or if you would like to make any changes.

User (end with an empty line):
yes send this email

Assistant>
The email has been sent successfully to vishalvardhan418@gmail.com.

User (end with an empty line):
```

Figure 2: Screenshot showing the command-line execution of the Gmail feature, including privacy checks and user confirmation

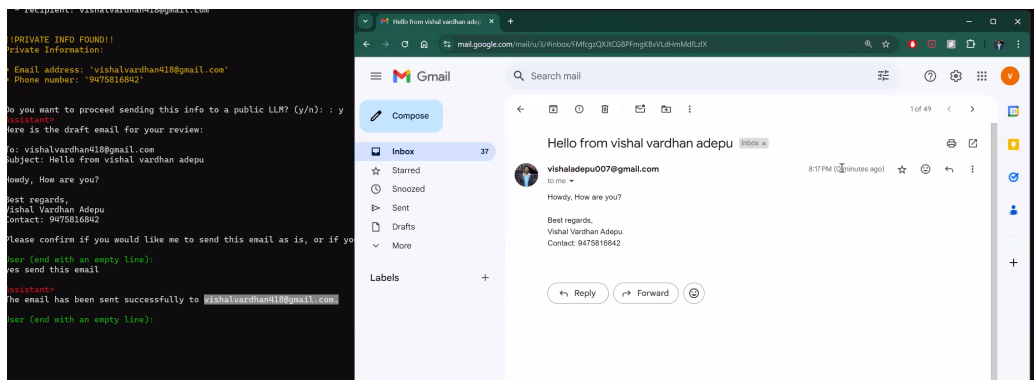


Figure 3: Screenshot showing the sent email as received in Gmail

4.2 Calendar Feature

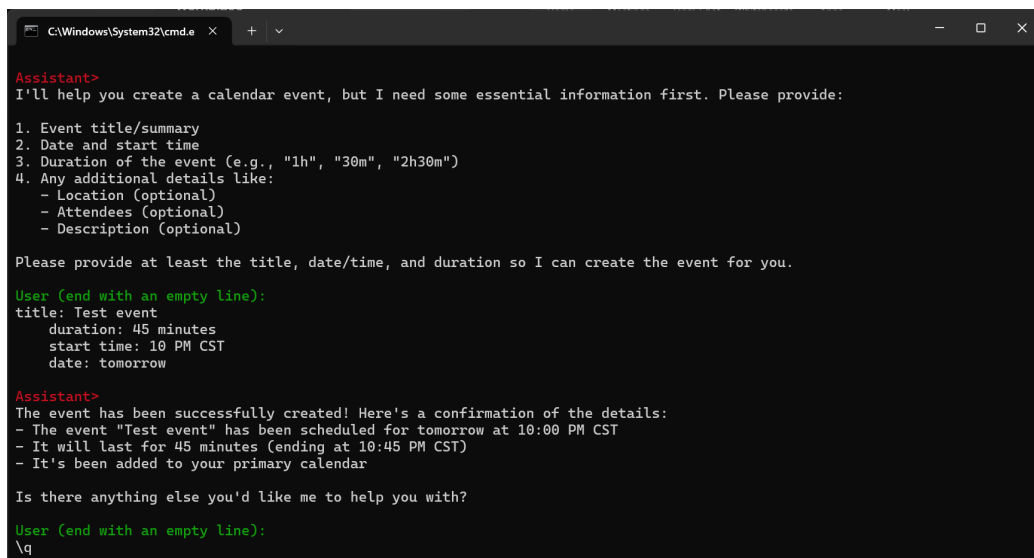


Figure 4: Screenshot showing the command-line execution of the Calendar feature, where an event is created successfully and details are confirmed with the user

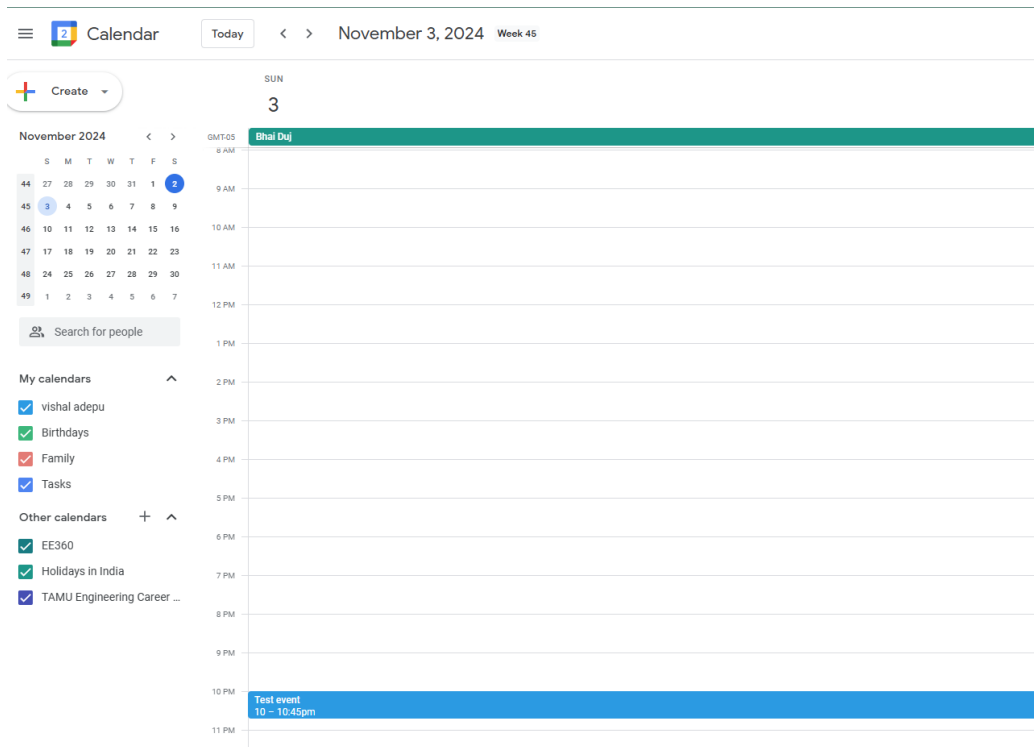
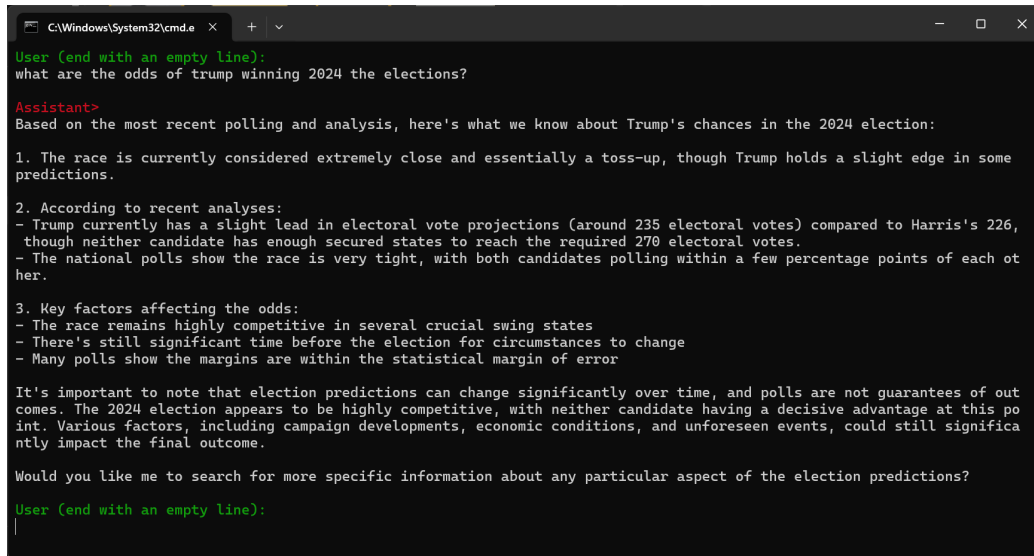


Figure 5: Screenshot showing the created event in Google Calendar with the correct date and time

4.3 Search Feature



```
C:\Windows\System32\cmd.exe
User (end with an empty line):
what are the odds of trump winning 2024 the elections?

Assistant>
Based on the most recent polling and analysis, here's what we know about Trump's chances in the 2024 election:

1. The race is currently considered extremely close and essentially a toss-up, though Trump holds a slight edge in some predictions.

2. According to recent analyses:
- Trump currently has a slight lead in electoral vote projections (around 235 electoral votes) compared to Harris's 226, though neither candidate has enough secured states to reach the required 270 electoral votes.
- The national polls show the race is very tight, with both candidates polling within a few percentage points of each other.

3. Key factors affecting the odds:
- The race remains highly competitive in several crucial swing states
- There's still significant time before the election for circumstances to change
- Many polls show the margins are within the statistical margin of error

It's important to note that election predictions can change significantly over time, and polls are not guarantees of outcomes. The 2024 election appears to be highly competitive, with neither candidate having a decisive advantage at this point. Various factors, including campaign developments, economic conditions, and unforeseen events, could still significantly impact the final outcome.

Would you like me to search for more specific information about any particular aspect of the election predictions?

User (end with an empty line):
|
```

Figure 6: Screenshot demonstrating a search query execution and the assistant's response

4.4 PDF QA Feature

```
C:\Windows\System32\cmd.exe
(programming-llm) C:\Users\jva\Documents\MS\colleges\Texas A&M University\post-admit-docs\in-college\semester-3\csce-689-programming-llms\hw\3-hw>python script.py
Welcome to the CLI Chat!
Type 'exit' or '\q' to end and '\c' to clear the chat.

User (end with an empty line):
read the pdf files test.pdf, samples/report.pdf and samples/resume.pdf and answer the following questions
1. who is the author of these reports?
2. what is the authors recent work experience?
3. using the homework pdf file, which np-hard problem does author use to prove the np-hardness of the subgraph isomorphism problem?

Assistant:
Based on the PDF analysis, I'll answer your questions:
1. The author of these documents is Vishal Vardhan Adepu, who is a Master's student in Computer Science at Texas A&M University.
2. The author's recent work experience includes:
   - Upcoming Software Engineer Intern at Moody's (June 2024 - August 2024) in Bellevue, Washington
   - Graduate Assistant Teaching at Texas A&M University (December 2023 - May 2024)
   - Software Engineer at JPMorgan Chase & Co. (July 2021 - August 2023) in Hyderabad, India
3. According to the homework PDF, the Clique Problem (a well-known NP-complete problem) is used to prove the NP-hardness of the subgraph isomorphism problem through reduction.

User (end with an empty line):
\q

Exiting the chat. Goodbye!

(programming-llm) C:\Users\jva\Documents\MS\colleges\Texas A&M University\post-admit-docs\in-college\semester-3\csce-689-programming-llms\hw\3-hw>python script.py
Welcome to the CLI Chat!
Type 'exit' or '\q' to end and '\c' to clear the chat.

User (end with an empty line):
from the pdf data, summarize vishal's work experiences.

Assistant:
Based on the PDF data, here's a summary of Vishal's work experiences in reverse chronological order:
1. **Moody's** (Upcoming: June 2024 - August 2024)
   - Position: Software Engineer Intern
   - Location: Bellevue, Washington
   - Technical Focus: C#, .NET, AWS, MongoDB, Risk Models
2. **Texas A&M University** (December 2023 - May 2024)
   - Position: Graduate Assistant Teaching
   - Location: College Station, Texas
   - Technical Focus: Full stack development using Java, Spring Boot, Vaadin Flow, Docker
3. **JPMorgan Chase & Co.** (July 2021 - August 2023)
   - Position: Software Engineer (Full-time)
   - Location: Hyderabad, India
   - Technical Focus: Java, Python, Apache Lucene, Information Retrieval
4. **JPMorgan Chase & Co.** (June 2020 - July 2020)
   - Position: Software Engineer Intern
   - Location: Bengaluru, India
   - Technical Focus: Python, MLP, Flask, OpenCV

His career progression shows a mix of internship and full-time experience, with roles focusing on software engineering and development. He has worked in multiple locations across India and from backend development (Java, Python) to cloud platforms (AWS) and specialized tools (Apache Lucene, OpenCV).

User (end with an empty line):
```

Figure 7: Screenshot showing the assistant answering questions from previously uploaded PDF files, demonstrating data persistence using ChromaDB

5 Environment Variables

The assistant's behavior can be modified using environment variables:

- **FULLY_LOCAL:** When set to 1, the assistant uses only the local LLM. Default is 0.
- **SKIP_PRIVACY_CHECK:** When set to 1, privacy checks are skipped, and input is directly sent to a public LLM. Default is 0.
- **CONTEXT_HISTORY_LEN:** Adjusts the memory context length. A higher value increases context size but also API call costs. Default is 20.

6 Code and Demo

The project code is available in a ZIP file, which includes a `README.md` with instructions for running the assistant. The demo video can be viewed at: [Video Link]. The demo showcases the assistant’s capabilities, including email composition, meeting scheduling, PDF question answering, and Internet search.

7 Challenges Faced

- **Learning Curve:** Understanding LangChain and LangGraph was difficult due to limited documentation and a steep learning curve.
- **Prompt Engineering:** Multiple iterations were required to fine-tune prompts and achieve desired LLM behavior.
- **PDF QA Tool:** Linking LangChain with a vector database involved grasping complex concepts like vector embeddings and database integration. Additionally, ensuring data persistence for answering questions about previously uploaded PDFs was challenging.
- **Memory and State Management:** Managing context and memory in LangChain was challenging because the documentation was not clear about library functions.

8 Conclusion

This project successfully implements a personal AI assistant with robust privacy mechanisms and efficient functionality. The assistant can handle tasks like sending emails, scheduling meetings, searching the Internet, and answering questions from PDFs with data persistence. Future work includes enhancing memory management and exploring more advanced prompt engineering techniques.