URBAN
SYSTEMS

Final Report
LLM Powered AI Chatbot

Team Members:

Sadia Khan
Jauhar Fathima
Om Patel
Sanket Patel

Submission Date: April 15th, 2025

1. Introduction

Project Overview

The project focuses on developing an advanced LLM-powered AI chatbot to enhance information retrieval within organizations. The chatbot is designed to address employees' challenges when searching for information across multiple systems and folders, which slows collaboration and productivity. The system aims to provide an intuitive and interactive AI-driven solution for efficient information access.

Problem Statement

Employees struggle with locating information due to:

- Fragmented storage across multiple folders and systems.
- Static websites lack interactive and efficient search capabilities.
- Knowledge silos that reduce collaboration efficiency.

To solve this, we propose an AI chatbot and an adding information page to:

- Provide LLM (GPT-4o mini) powered responses for accurate information retrieval
- Seamlessly integrates with internal databases via APIs
- Allow non-technical users to add and edit content easily
- Enhance collaboration and streamline workflow

2. Tech Stack and Justification

Component	Technology Used	Justification
Frontend	React.js	● Provides a fast, scalable, and interactive UI ● All the members have experience with this.
Backend	FastAPI (Python)	● Requested by industry partner ● Helps developers build applications quickly and efficiently
Database	MongoDB	● A flexible NoSQL database suited for chatbot interactions
LLM	OpenAI GPT API	● State-of-the-art language model for accurate responses
Information Retrieval	ChromaDB (Vector Database)	● Open-source & easy to set up ● Works well with LLMs like OpenAI GPT, LangChain, and FastAPI ● Stores text embeddings for semantic search
User Authentication & Security	MultiFactor Authentication (MFA) – Email OTP	● Instantly sends OTP to email for secure, passwordless login ● Simple, smooth experience with stronger protection against unauthorized access

3. Project Timeline

Task	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
UI/UX design, database setup	[redacted]	[redacted]						
LLM integration, API dev			[redacted]	[redacted]	[redacted]			
Testing & optimization						[redacted]	[redacted]	
Deployment & finalization								[redacted]

4. Deliverables & Task Distribution

Task	Importance	Status	Team Member(s)
UI/UX Design	-	Completed	Sadia Khan, Jauhar Fathima
Database Setup	-	Completed	Om Patel
LLM Integration	Must Have	Completed	Sanket Patel
API Development	Must Have	Completed	Sadia Khan
Frontend Development	Must Have	Completed	Sadia Khan, Jauhar Fathima
Information Retrieval (RAG)	Must Have	Completed	Om Patel, Sanket Patel
User Authentication	Must Have	Completed	Om Patel, Sanket Patel
Chat History	Nice to Have	Completed	Om Patel, Sanket Patel
Customizable Themes (Light vs Dark Mode)	Nice to Have	Completed	Team Collaboration
RAG Enhancement	Must Have	Completed	Team Collaboration

UI/UX Improvements	-	Completed	Team Collaboration
Testing and Debugging	-	Completed	Team Collaboration
Deployment with Docker	Nice to Have	Completed	Team Collaboration

5. Final Pages

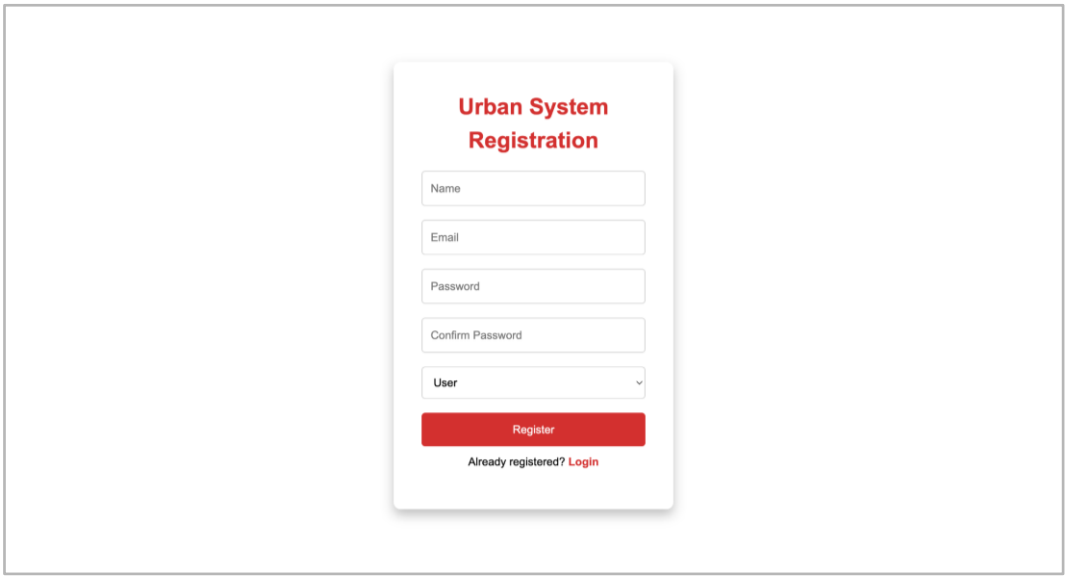

Figure 4.1: Registration Page – New User

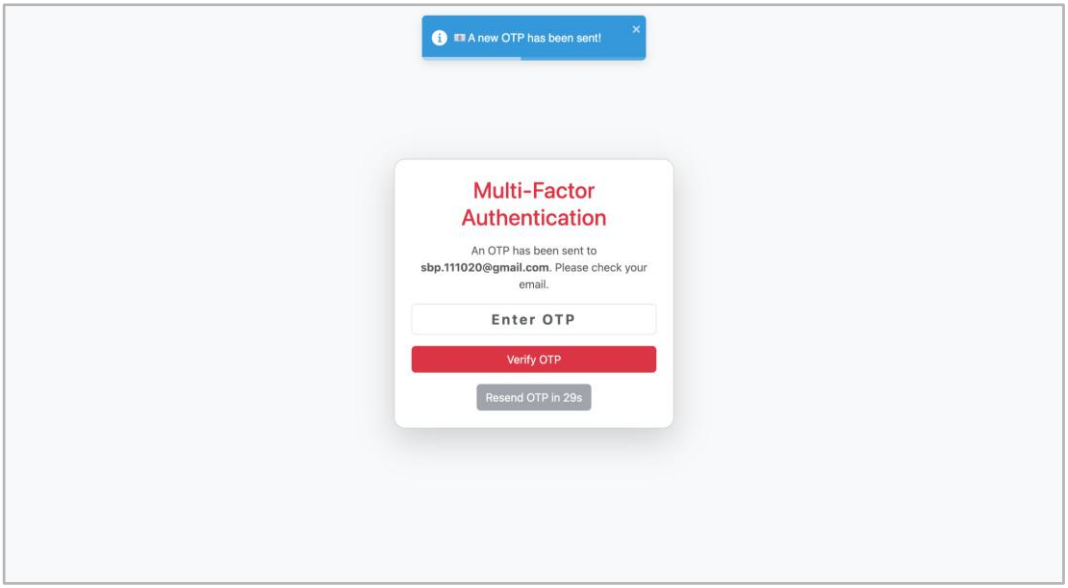

Figure 4.2: Multi-Factor Authentication Page – Enter OTP

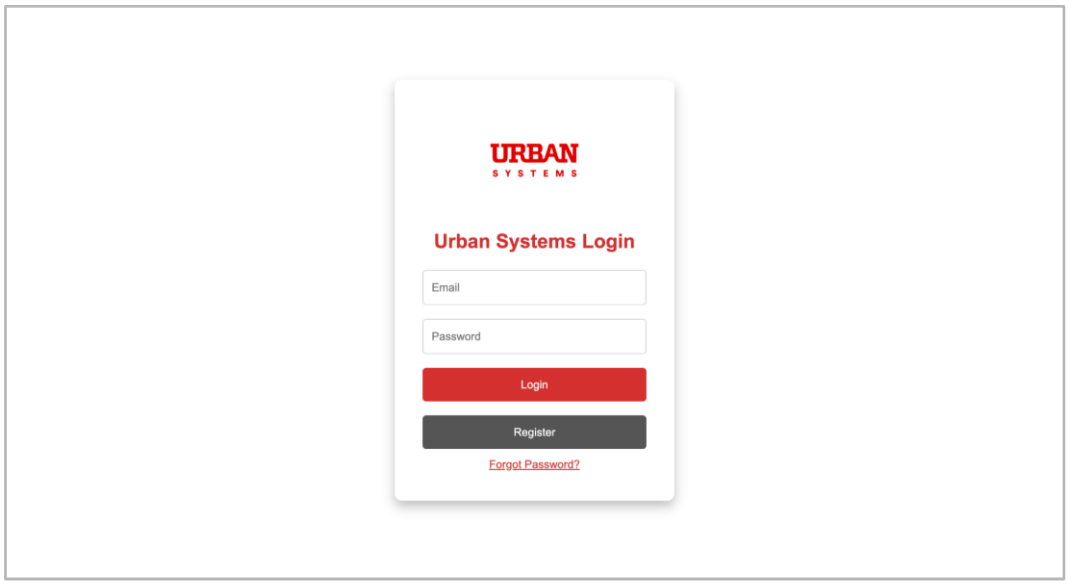

Figure 4.3: Login Page – Existing User

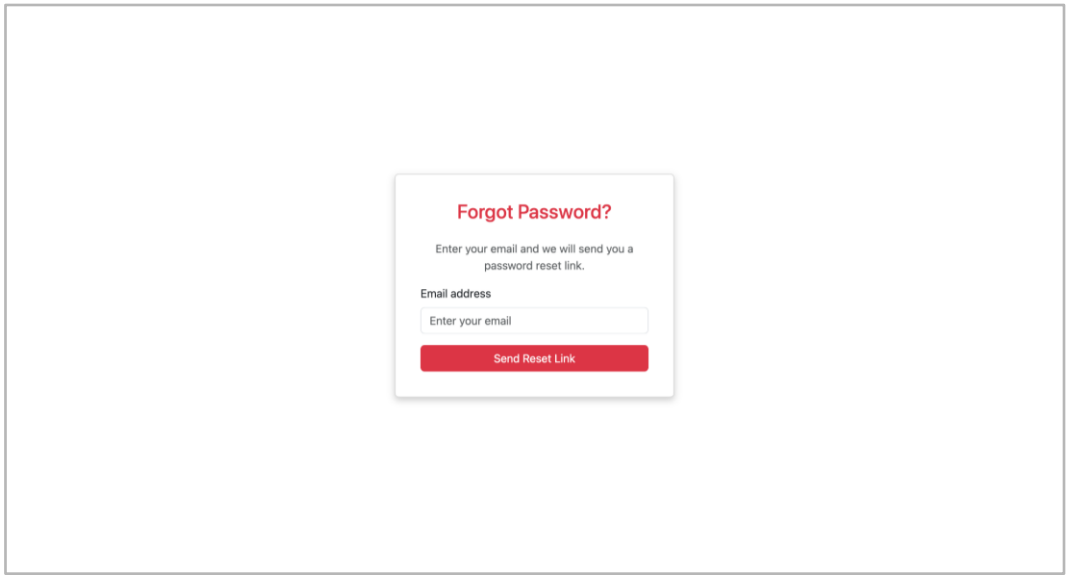

Figure 4.4: Forgot Password Page

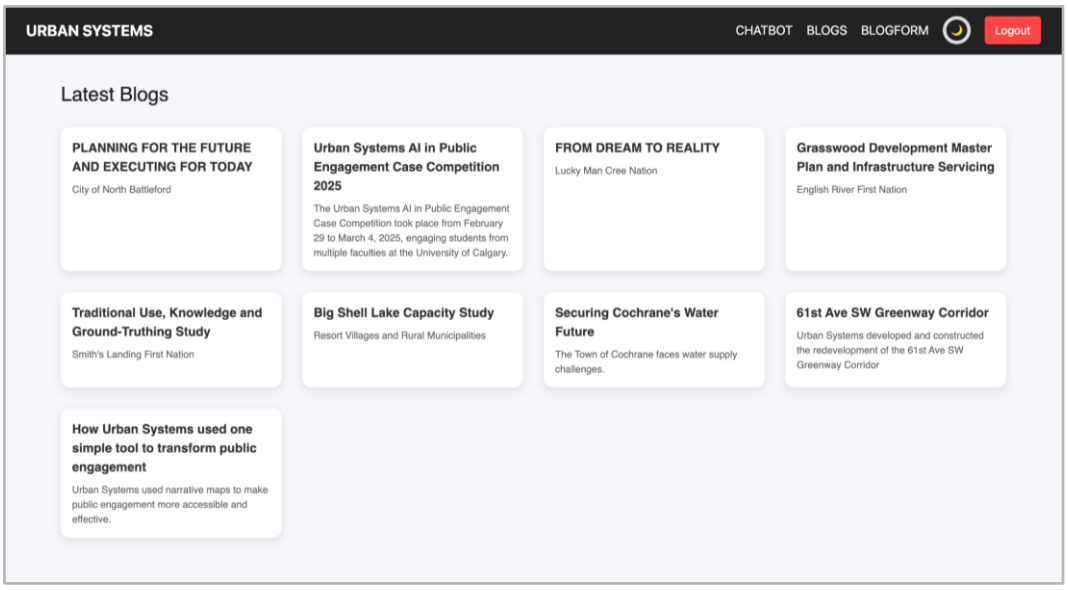

Figure 4.5: Latest Blog Page – Light Mode

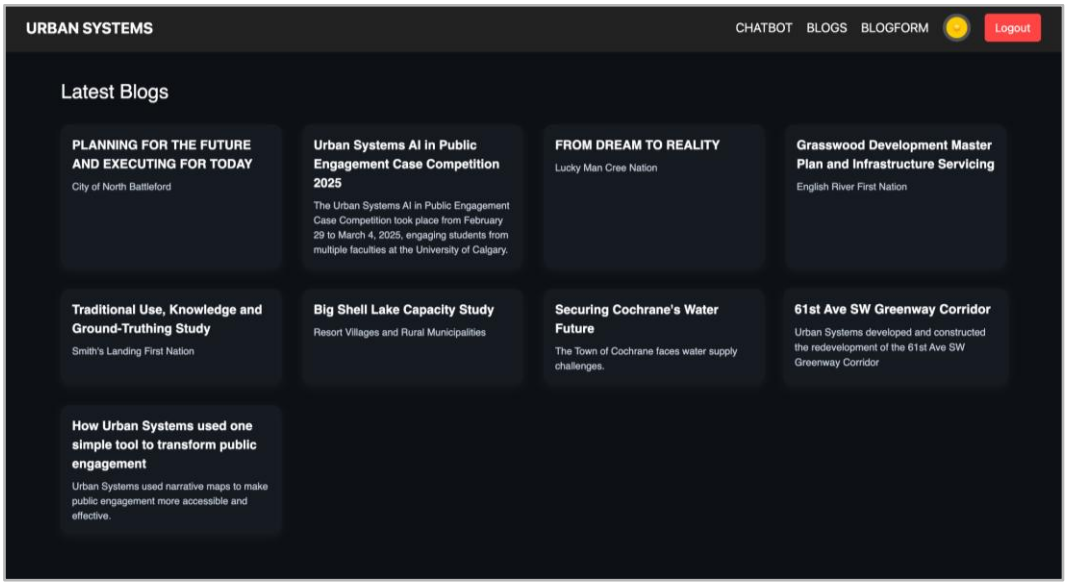

Figure 4.6: Latest Blog Page – Dark Mode

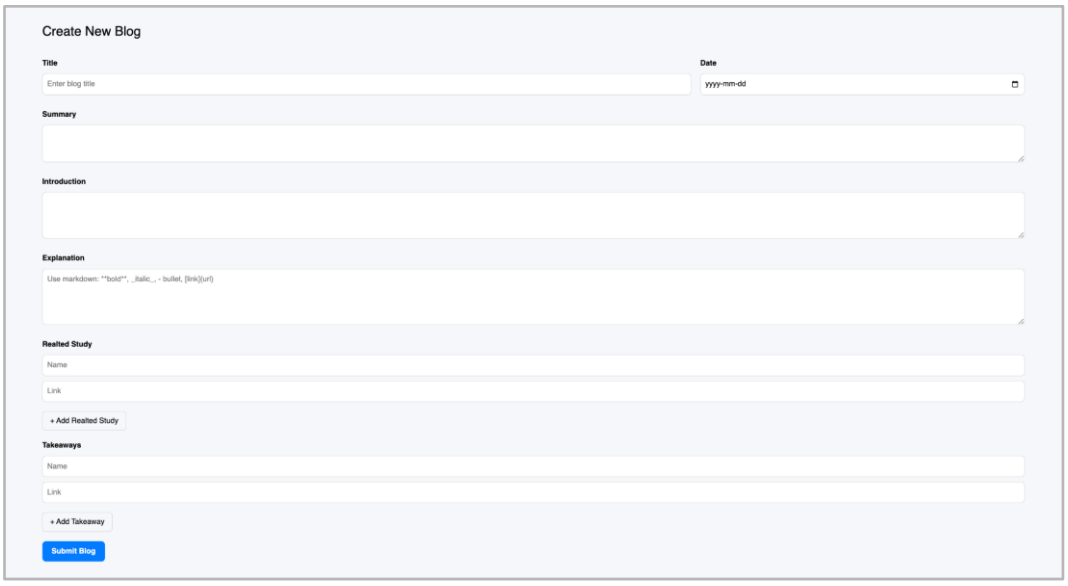

Figure 4.7: Create New Blogpost Page

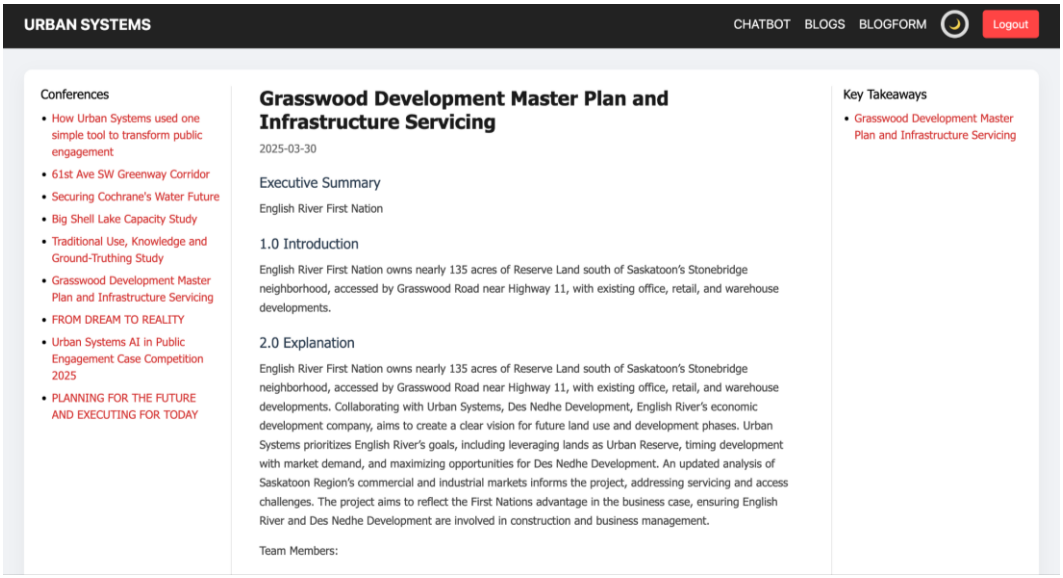

Figure 4.8: Blog Post Page – Light Mode

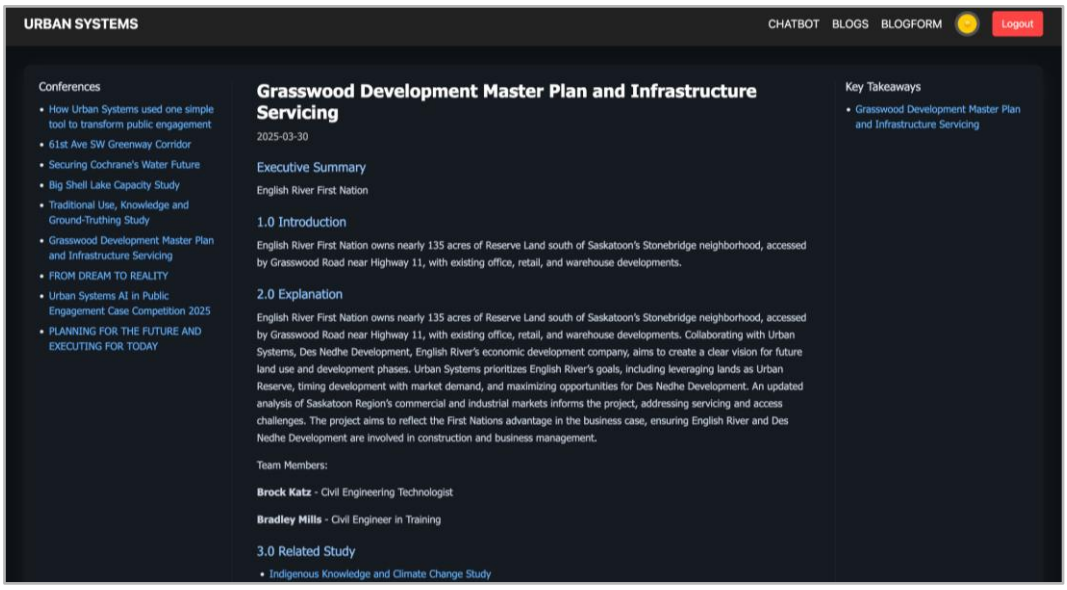

Figure 4.9: Blog Post Page – Dark Mode

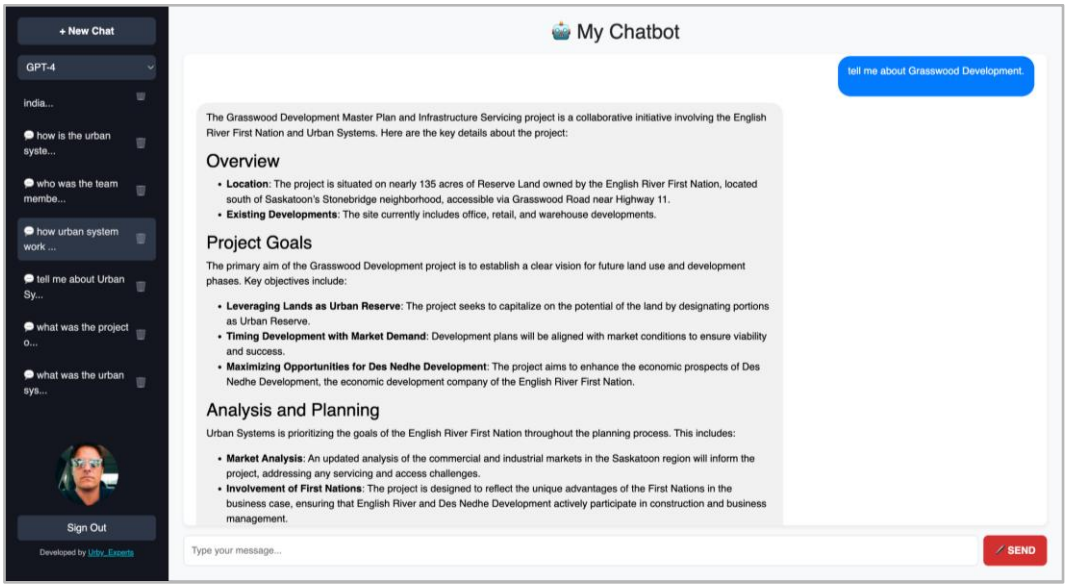

Figure 4.10: Chatbot Page – Light Mode

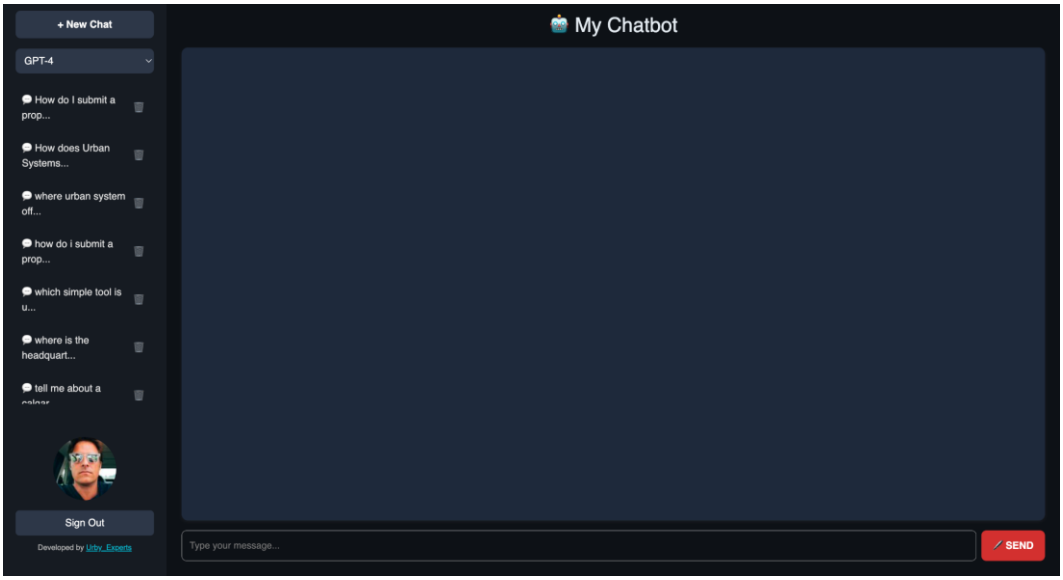

Figure 4.11: Chatbot Page – Dark Mode

6. System Architecture Diagram

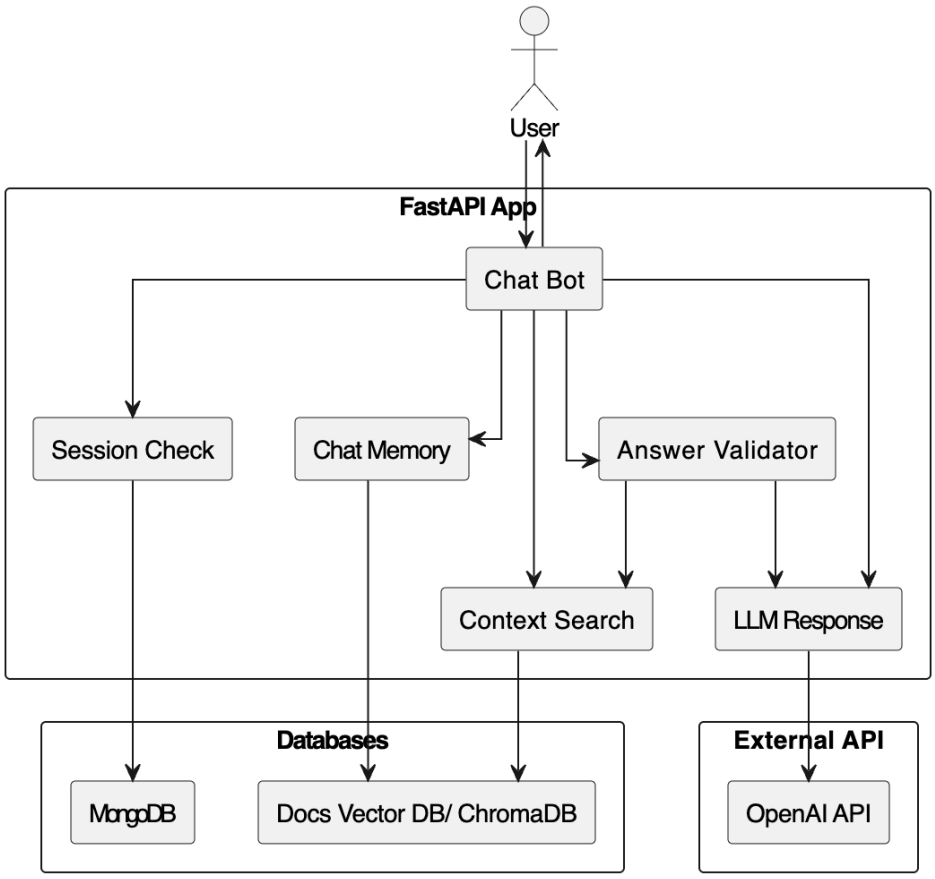

Figure 5.1: Overall Architecture Model

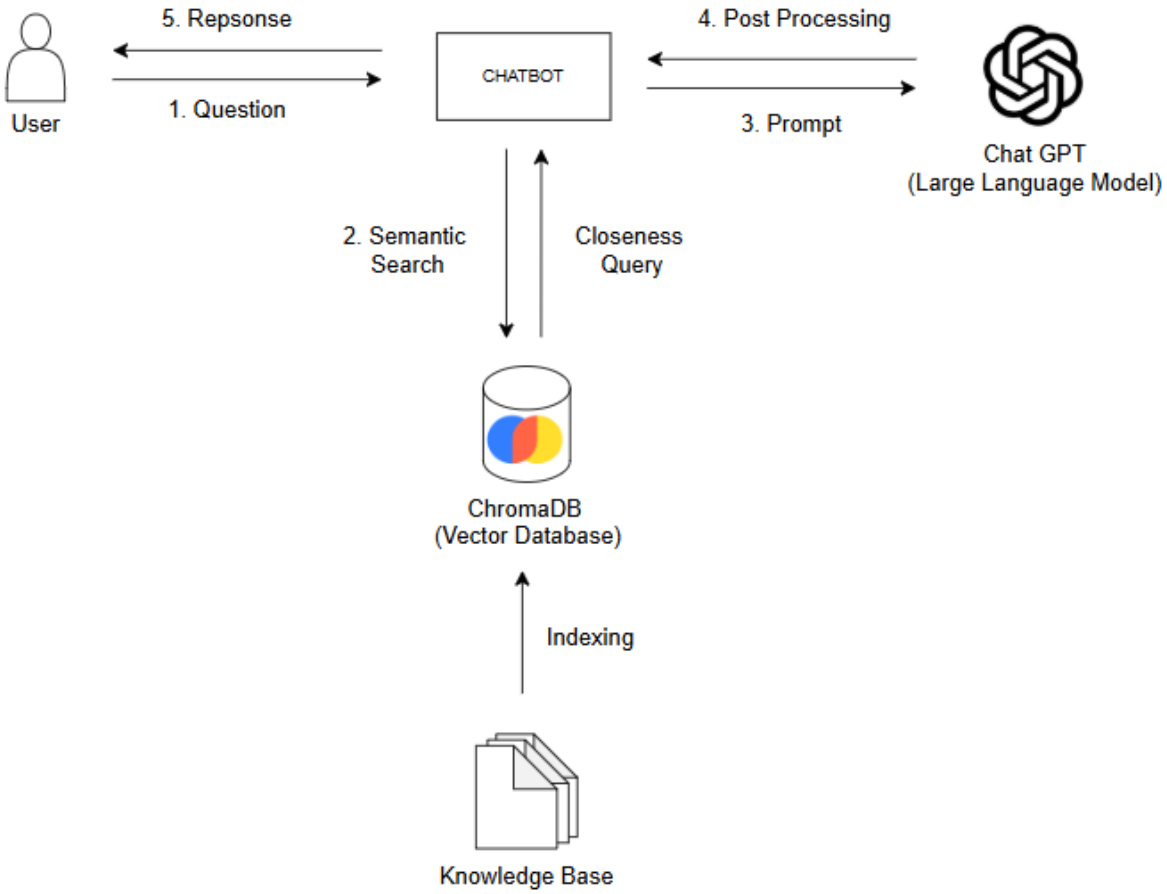

Figure 5.2: Retrieval-Augmented Generation (RAG) Architecture Model

7. System Interaction Diagram

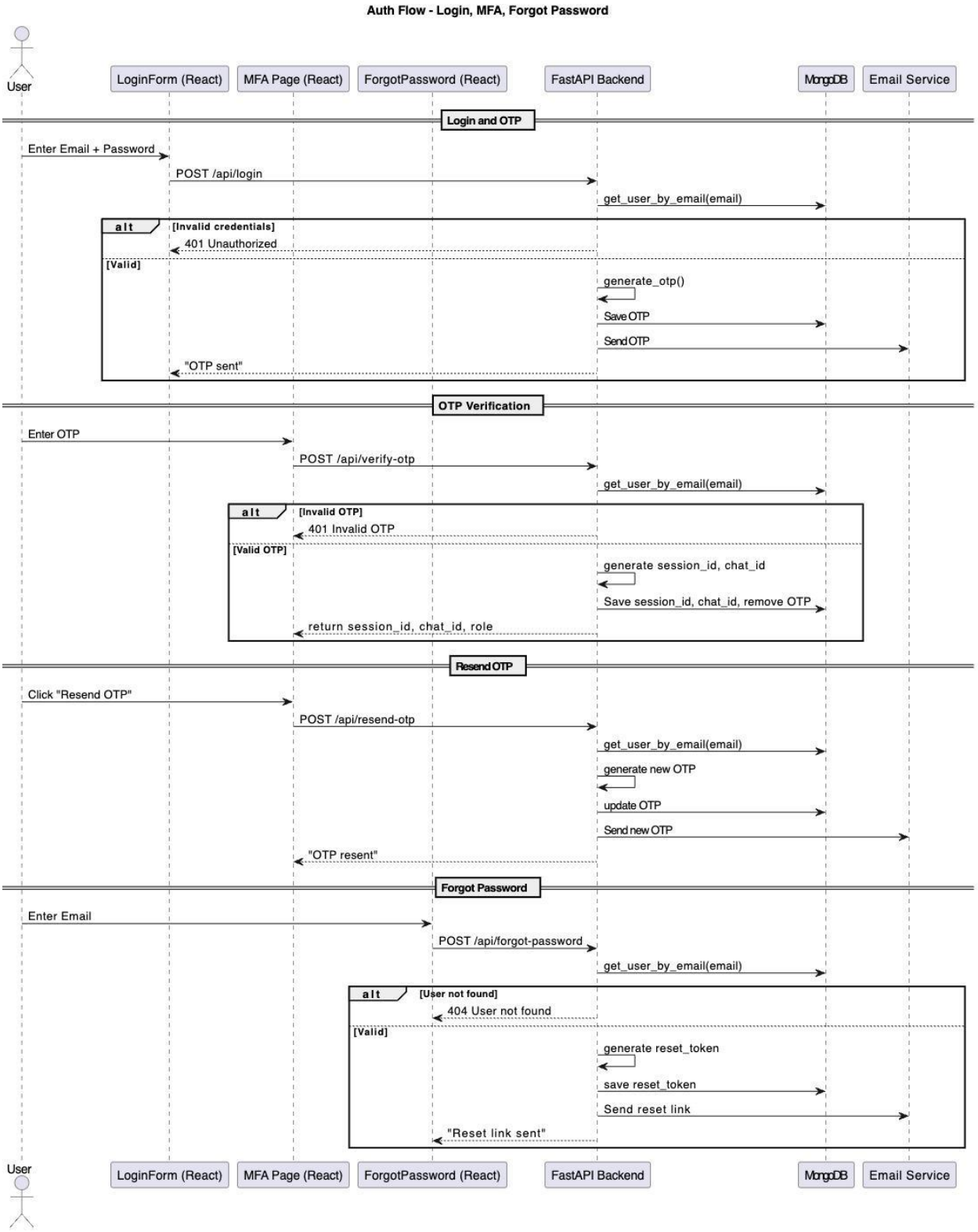

Figure 6.1: Login system interaction diagram

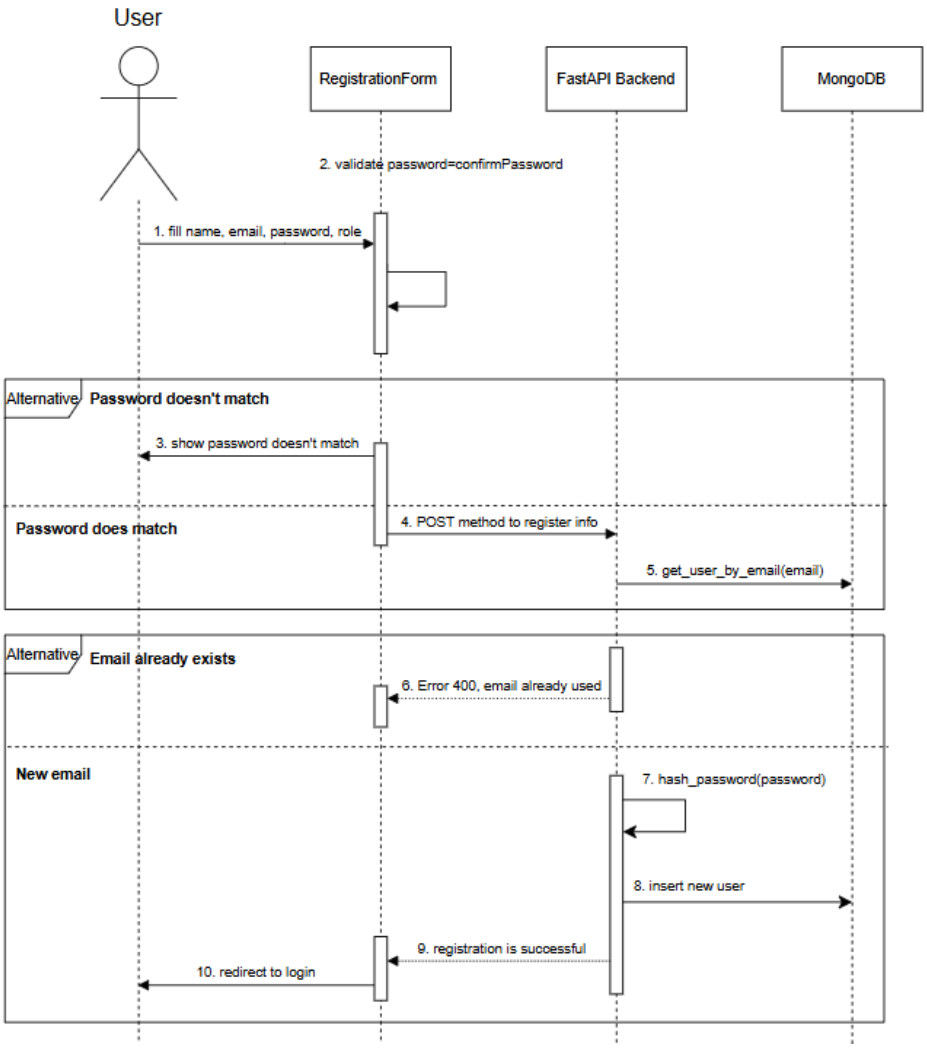

Figure 6.2: Registration system interaction diagram

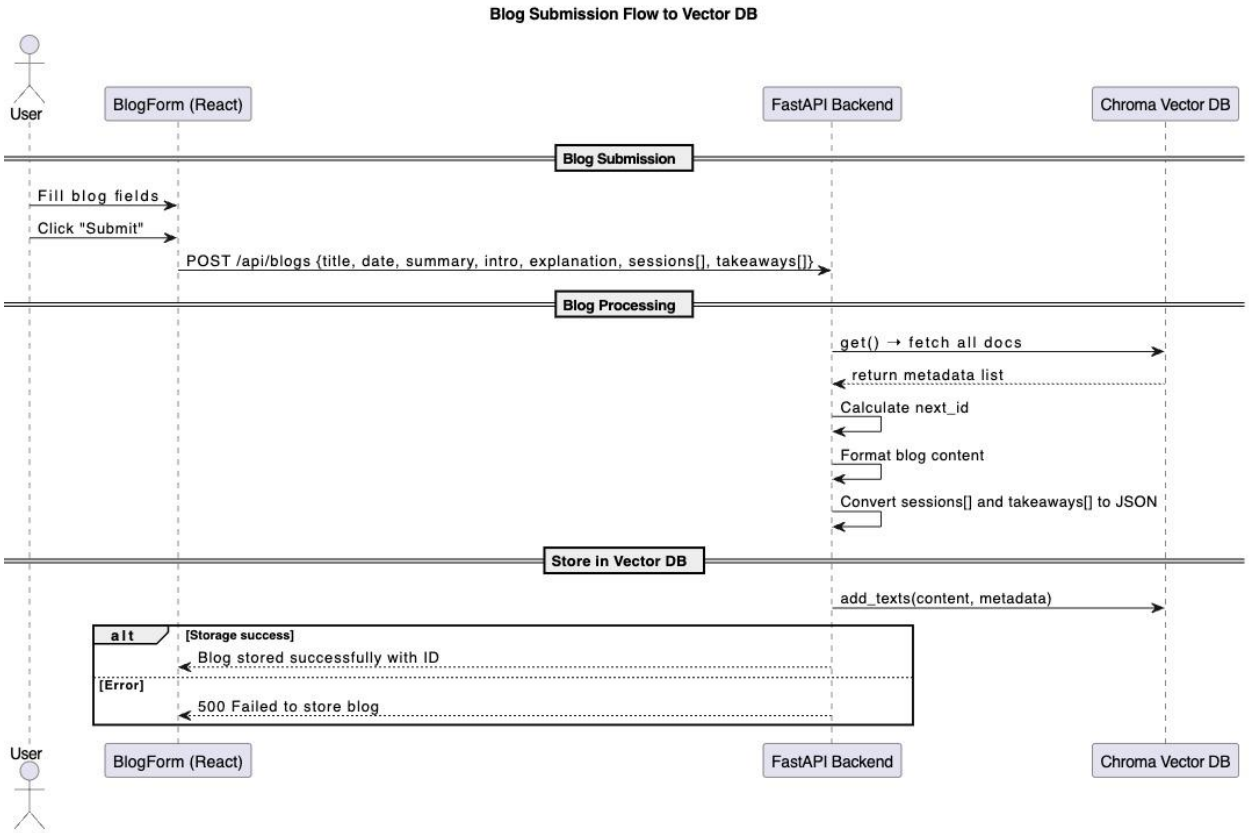

Figure 6.3: Create a new blog post system interaction diagram

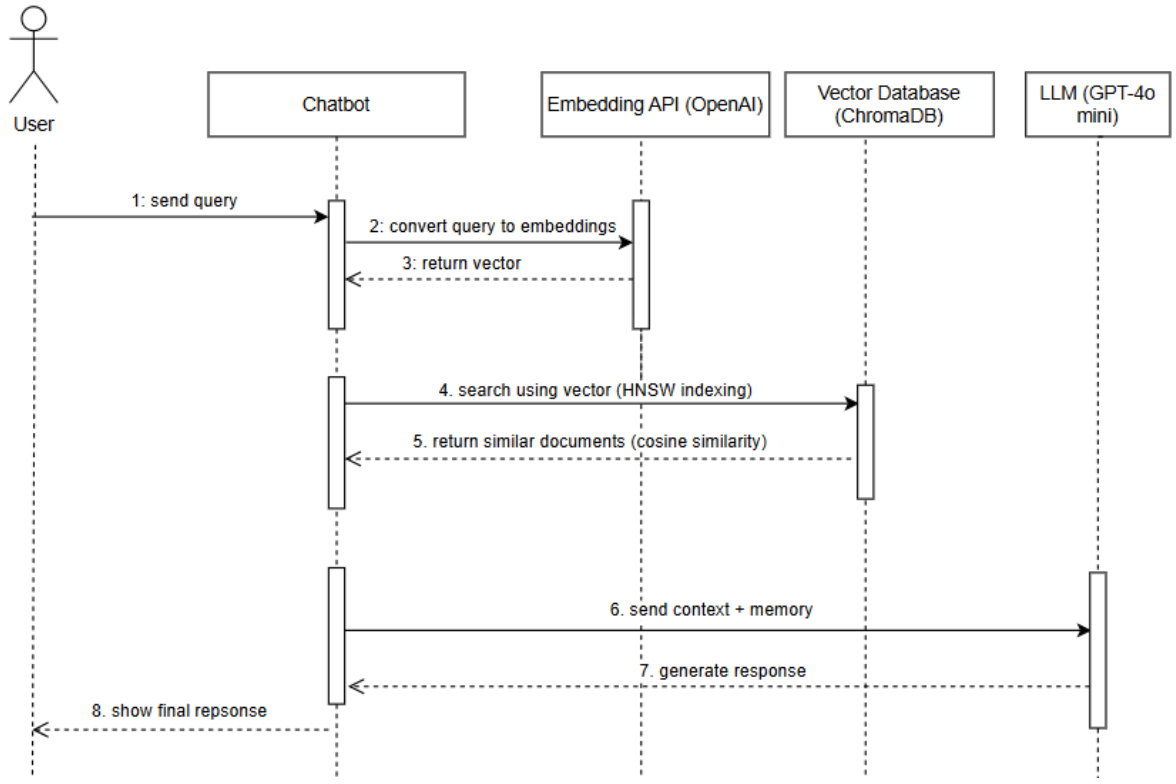

Figure 6.4: Query with chatbot system interaction diagram

8. Problems Encountered and Solutions

1. **Login and Registration Validation:**
    - Challenge: Ensuring that users provide a valid email during registration
    - Solution: Implemented email verification by sending a code to users during registration

2. **Forgot Password Functionality:**
    - Challenge: Users lacked an option to reset their passwords
    - Solution: Implemented a password reset feature using OTP verification

3. **Handling Large API Requests Efficiently:**
    - Challenge: LLM API calls could cause delays under heavy load
    - Solution: Used caching mechanisms and optimized request handling to improve response times

4. **User Role Management:**
    - Challenge: Defining different access for users and admins
    - Solution: Implemented role-based access control (RBAC) to ensure proper authorization

5. **Vector Search Accuracy:**
    - Challenge: Some queries returned irrelevant results or lacked context.
    - Solution: Implemented proper embedding and indexing to improve semantic search. Adjusted generation settings, such as temperature, to ensure responses were more accurate and contextually relevant.

9. Suggestions for Future Work

Feature	Recommendation
Deployment	Deployment with Docker
Multilingual Chat Support	Integrate OpenAI multilingual models or add a translation layer before prompt generation
Voice Command Input	Use Web Speech API for browser voice recognition or Whisper API for transcription
Analytics Dashboard	Track queries, user activity, most-used documents; visualize using Chart.js or Tableau
Document Upload & Extraction	Added file upload and extracted content using LangChain PDF loaders and summarizers

10. References

1. FastAPI. (n.d.). *FastAPI Documentation*. Retrieved April 15, 2025, from https://fastapi.tiangolo.com/

2. Motor. (n.d.). *Asynchronous Python Driver for MongoDB*. Retrieved April 15, 2025, from https://motor.readthedocs.io/

3. OpenAI. (n.d.). *OpenAI GPT API*. Retrieved April 15, 2025, from https://platform.openai.com/docs/guides/gpt

4. LangChain. (n.d.). *LangChain Chroma Vectorstore Integration*. Retrieved April 15, 2025, from https://docs.langchain.com/docs/integrations/vectorstores/chroma

5. ChromaDB. (n.d.). *Chroma Vector Database – GitHub*. Retrieved April 15, 2025, from https://github.com/chroma-core/chroma

6. bcrypt. (n.d.). *bcrypt – Python Package Index (PyPI)*. Retrieved April 15, 2025, from https://pypi.org/project/bcrypt/

7. Python Software Foundation. (n.d.). *Email and SMTP Handling in Python*. Retrieved April 15, 2025, from https://docs.python.org/3/library/email.examples.html

8. Python dotenv. (n.d.). *python-dotenv – Environment Variable Loader*. Retrieved April 15, 2025, from https://pypi.org/project/python-dotenv/

9. Python Software Foundation. (n.d.). *UUID and Secrets Modules – Official Docs*. Retrieved April 15, 2025, from https://docs.python.org/3/library/uuid.html and https://docs.python.org/3/library/secrets.html

10. Meta (React). (n.d.). *React – JavaScript Library for Building Interfaces*. Retrieved April 15, 2025, from https://react.dev/