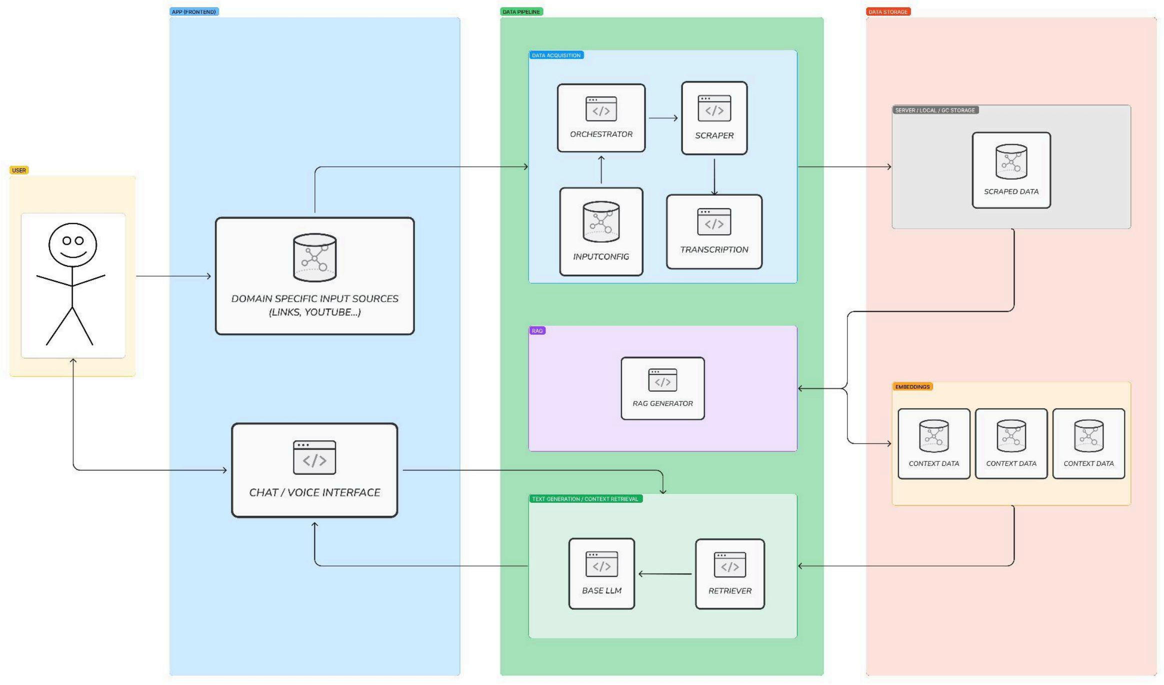
Design Documentation

Introduction

The “Ailixir” project is a framework that utilises Large Language Models (LLMs) to develop customised AI agents tailored to various domains (e.g., medical, business, biology). These agents can communicate via text and voice, adapt to user input and refine interactions over time. They are based on handpicked content sources (e.g. youtube videos, scientific papers), from which information is retrieved to customise the LLM.

Code Components



App (Frontend)

The user provides a set of input sources and/or a set of user-specific parameters to configure the data pipeline. The actual information from these input sources (e.g. youtube video, scientific paper, ...) will then be retrieved in the data pipeline components to later feed the customised AI agent. When the agent has been generated, a user can have a conversation with the agent through a chat or voice interface. The app frontend has not been implemented this far into the project.

Data Acquisition Pipeline

Input Config

The *Config* component provides functionality to define sources to scrape information from (also called scraping targets). It currently supports:

- Youtube channel (retrieve information from all videos on the channel) or individual video links
- Podcasts from the Peter Attia Podcast Archive
- Articles from arXiv and PubMed
- Recipes from allrecipes.com
- Nutrition related blog posts from NutritionFacts.org

Furthermore, the component is used to set up a local database to store the information scraped from the input sources.

Orchestrator

The *Orchestrator* component manages the information retrieval process by scheduling scraping jobs and making sure that no source is scraped twice (storing an id for each scraped source). This component will be run after the input config is set up in order to perform the data acquisition.

Scrapers

For each type of input source (as detailed in the Input Config section), there exists a scraper. Scrapers are responsible for retrieving information - transcript and metadata - from a provided source. Metadata includes for instance publication date, authors, etc.. Once a source was scraped, the data is saved in the local database and an id is stored for the specific source, in order to avoid scraping from it again.

Custom Agent (LLMs)

RAG Generator

The RAG component takes the scraped information stored in the local database and creates embeddings based on them. This context data is stored in a vector database which will be used by the LLM. More detailed information on RAG can for example be found here.

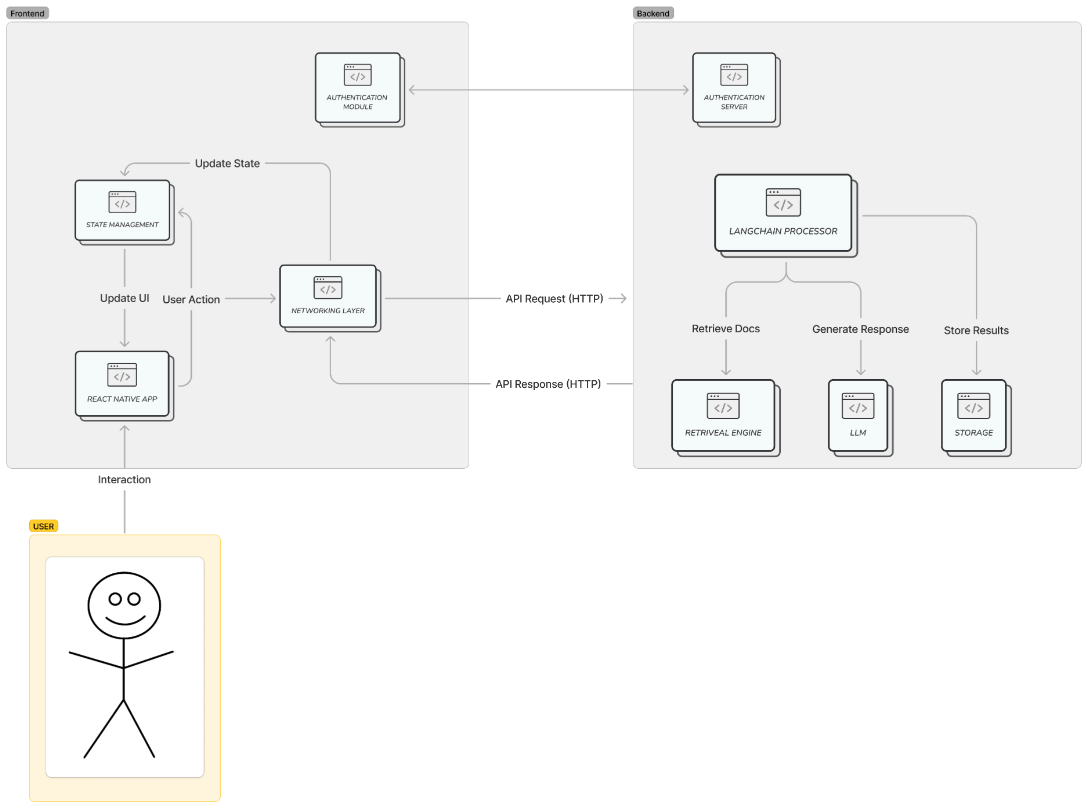
Text Generation/Context Retrieval

We use the LangChain framework to generate answers from a user prompt, the generated context data and the LLM. This component is currently still being worked on.

Data Storage

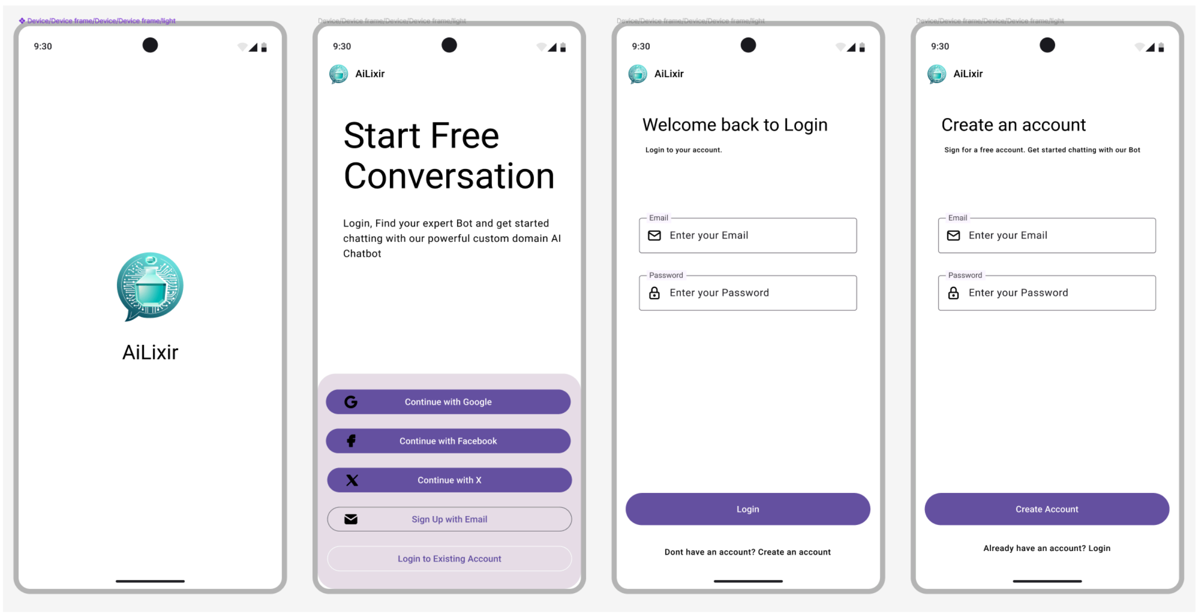
We use the computing resources of the FAU HPC to scrape data and store it. This data will be fed to LLMs later in the development.
We are adapting a Google Cloud solution to work with our embeddings database for the RAG Generator which is currently being implemented.

App System Architecture



1. The Authentication Module interacts with the Authentication Server to verify user credentials and manage sessions.
2. Users interact with the React Native app, triggering actions like querying the RAG system.
3. The Frontend (Networking Layer) sends requests to the Backend via HTTP.
4. Data Processing:
    a. The LangChain Processor first interacts with the Retrieval Engine to fetch relevant documents from the Document Store.
    b. A new prompt will be created with the fetched documents and passed to the LLM.
    c. (The response and possibly the retrieved documents are stored in Storage as needed.)
5. (Processed data and model outputs are stored in the Storage.)
6. Generated responses are sent back to the frontend via HTTP.

Initial App Wireframes



9:30

AiLixir

9:30

AiLixir

Start Free Conversation

Login, Find your expert Bot and get started chatting with our powerful custom domain AI Chatbot

Continue with Google

Continue with Facebook

Continue with X

Sign Up with Email

Login to Existing Account

9:30

AiLixir

Welcome back to Login

Login to your account.

Email

Enter your Email

Password

Enter your Password

Login

Dont have an account? Create an account

9:30

AiLixir

Create an account

Sign for a free account. Get started chatting with our Bot

Email

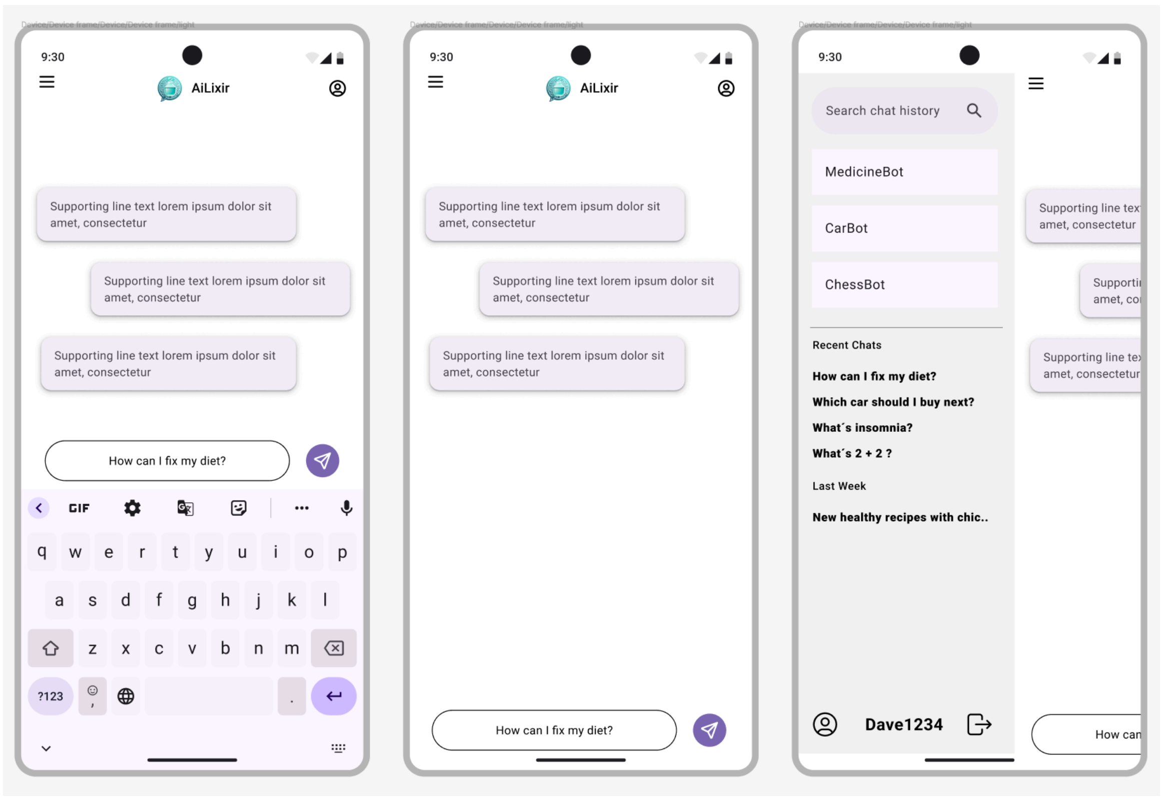
Enter your Email

Password

Enter your Password

Create Account

Already have an account? Login



9:30

AiLixir

Supporting line text lorem ipsum dolor sit amet, consectetur

Supporting line text lorem ipsum dolor sit amet, consectetur

Supporting line text lorem ipsum dolor sit amet, consectetur

How can I fix my diet?

9:30

AiLixir

Supporting line text lorem ipsum dolor sit amet, consectetur

Supporting line text lorem ipsum dolor sit amet, consectetur

Supporting line text lorem ipsum dolor sit amet, consectetur

How can I fix my diet?

9:30

Search chat history

MedicineBot

CarBot

ChessBot

Recent Chats

How can I fix my diet?

Which car should I buy next?

What´s insomnia?

What´s 2 + 2 ?

Last Week

New healthy recipes with chic..

Dave1234