FAITH

FAITH : **F**ramework for **AI** **I**ntegrity and **T**esting **H**allucinations

https://faith-cyfzeshhbjcdefhb.eastus2-01.azurewebsites.net/

TEAM MEMBERS

- **TEAM MEMBERS**
- **Aruthra S - (Female, Asian)**
  - **https://github.com/AruthraS** (GitHub)
  - **https://www.linkedin.com/in/aruthra-s-66b97b256/** (LinkedIn)

- **R S Kierthana - (Female, Asian)**
  - **https://github.com/KierthanaRS** (GitHub)
  - **https://www.linkedin.com/in/kierthana-rajesh-8b8b42256/** (LinkedIn)

- **Arun Pranav A T - (Male, Asian)**
  - **https://github.com/arunpranav-at** (GitHub)
  - **https://www.linkedin.com/in/arunpranavat/** (LinkedIn)

PROBLEM STATEMENT

- **PROBLEM STATEMENT**

- **Hallucination Detection and Context Validation**

- Build a framework that identifies hallucinations in AI-generated outputs and validates them against trusted data sources

- Our systems need to detect when AI generates content that is unsupported by any source, such as fabricated statistics or made-up historical facts. How could output automatically cross-reference with a reliable knowledge base or API? How do you provide confidence scores and explanations for detected hallucinations? For this challenge, build a framework that identifies hallucinations in AI-generated outputs and validates them against trusted data sources or a pre-defined knowledge base. Your project should provide false positive and false negative rates for hallucination detection, have the ability to integrate and validate with multiple external knowledge sources, and provide user-friendly visualization of hallucination metrics.

SOLUTION

F - **F**ramework for

A - **A**I

I - **I**ntegrity and
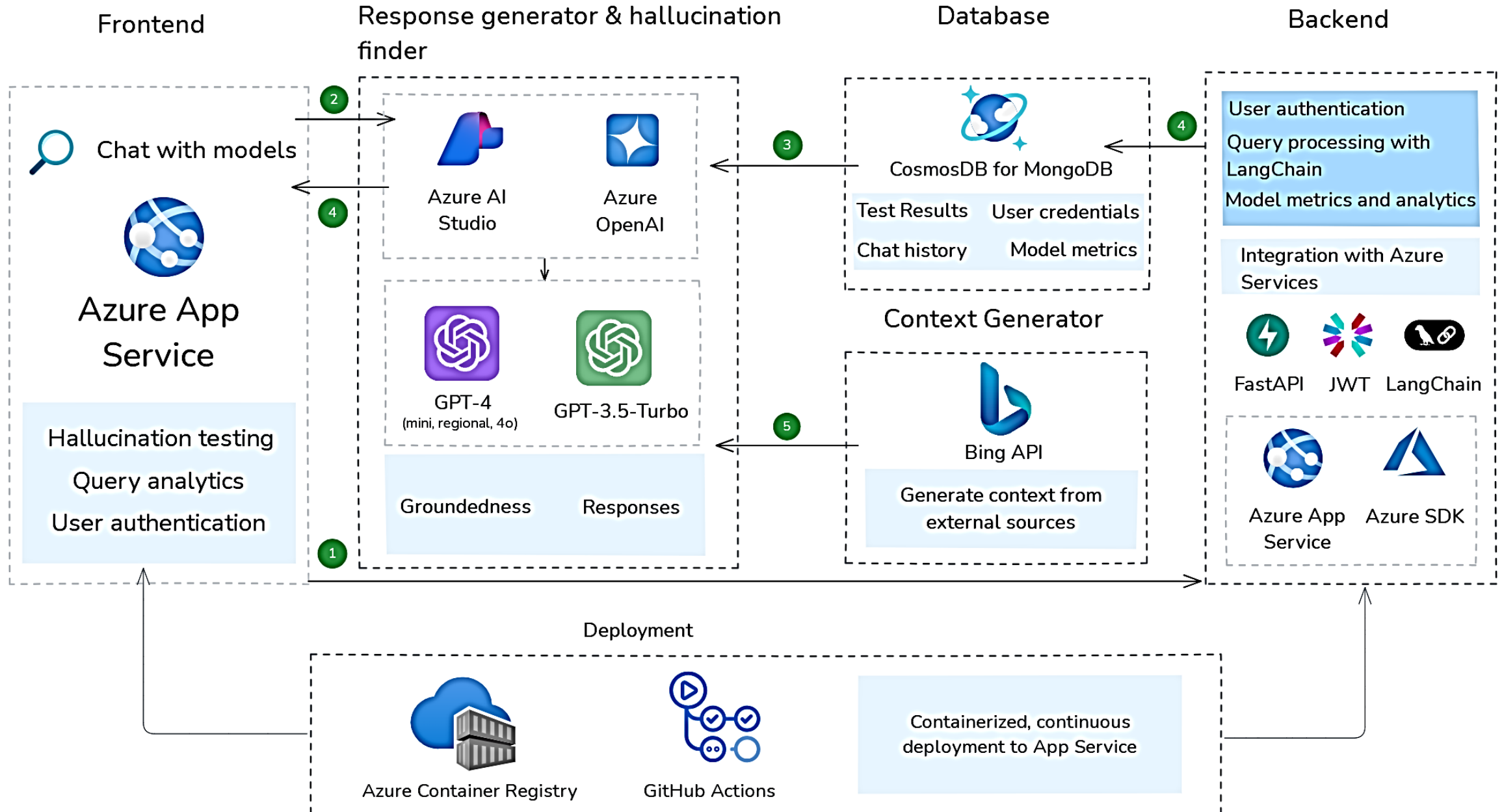
T - **T**esting

H - **H**allucinations

FAITH is an Azure AI based web application which is used to find hallucinations and ensure integrity among various AI models and LLMs along with confident scores, complete reasoning, detailed analytics and visualizations by comparing with external knowledge sources.

Multiple model support(gpt-4, gpt-4o, gpt-4o-mini, gpt-35-turbo-16k, etc.,), True and False negatives and positives, user friendly interface, detailed analytics, ready to use hosted platform are some of the unique selling points of our solution.

It is completely built on using Azure technologies like Azure App Service, Azure OpenAI, Bing Search API, Azure AI Evaluators, Azure Cosmos DB, Azure AI Foundry, Azure Container Registry, etc.,

ARCHITECTURE DIAGRAM

Frontend

Response generator & hallucination finder

Database

Backend

Chat with models

Azure App Service

Hallucination testing

Query analytics

User authentication

Azure AI Studio

Azure OpenAI

GPT-4
(mini, regional, 4o)

GPT-3.5-Turbo

Groundedness   Responses

CosmosDB for MongoDB

Test Results   User credentials

Chat history   Model metrics

Context Generator

Bing API

Generate context from external sources

User authentication

Query processing with LangChain

Model metrics and analytics

Integration with Azure Services

FastAPI   JWT   LangChain

Azure App Service   Azure SDK

Deployment

Azure Container Registry

GitHub Actions

Containerized, continuous deployment to App Service

AZURE TECHNOLOGIES USED

- Azure App Service
- Azure Cosmos DB
- Azure OpenAI
- Azure AI Foundry
- Azure AI Services
- Azure Bing Search API
- Azure Container Registry
- Azure AI Evaluators
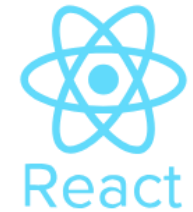- Azure Cognitive Services
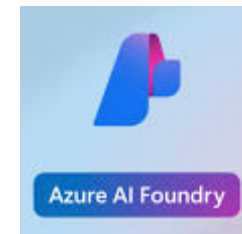
TECH STACK USED

- Next.js
- Tailwind CSS
- Chart.js
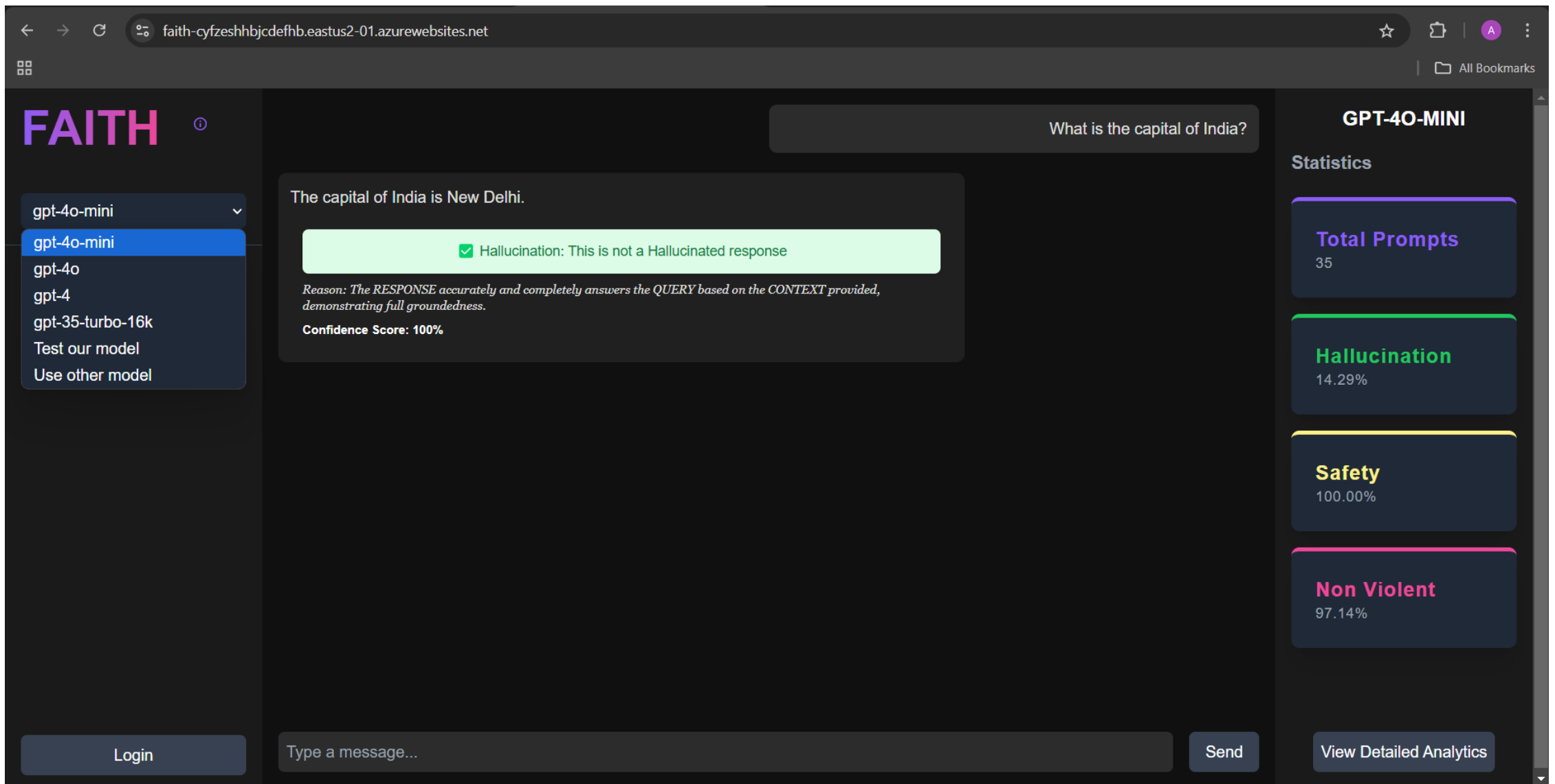- Python
- Fast API
- MongoDB
- GitHub
- Docker
- LangChain
- VSCode
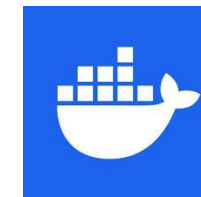
FEATURES

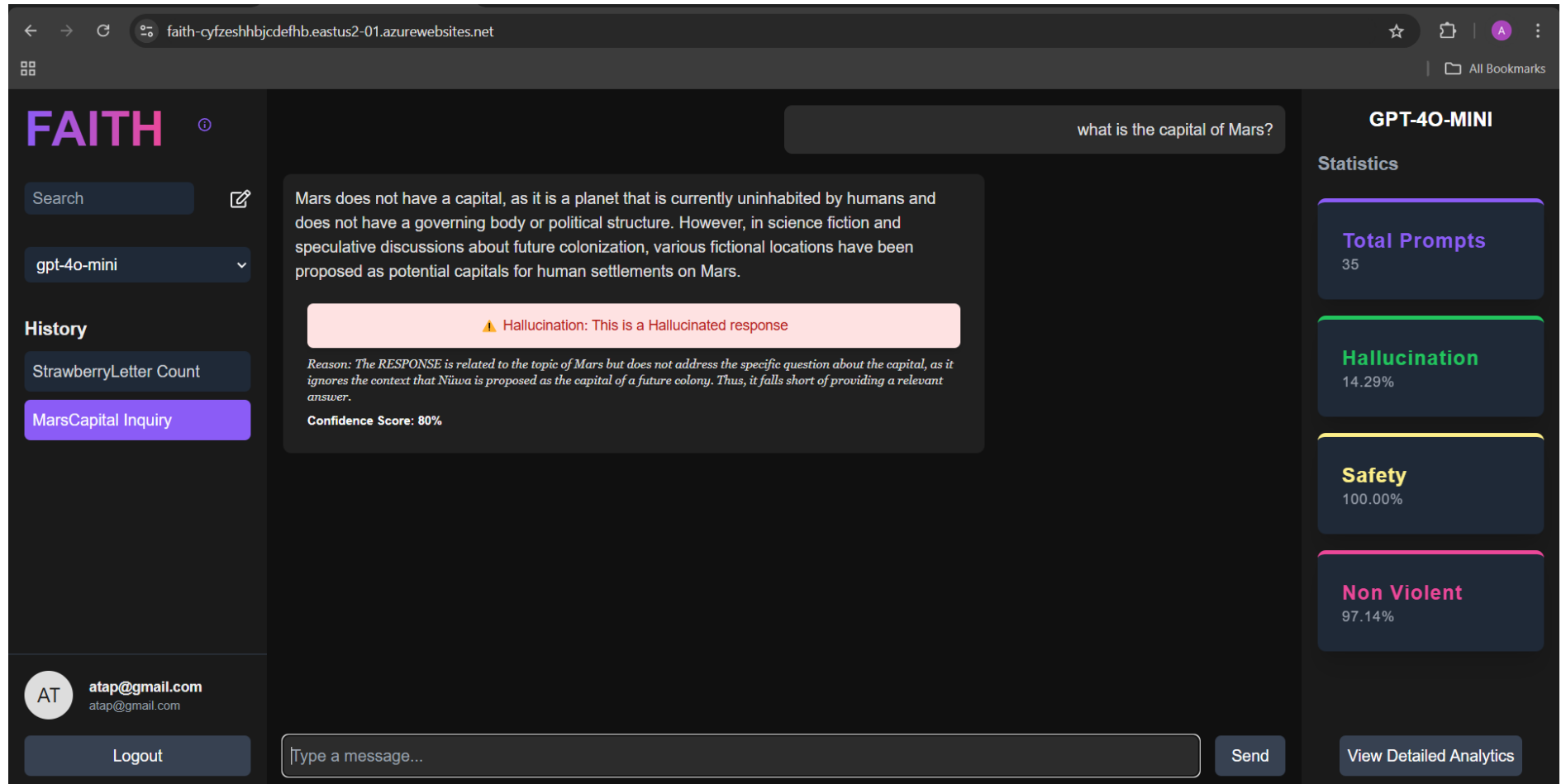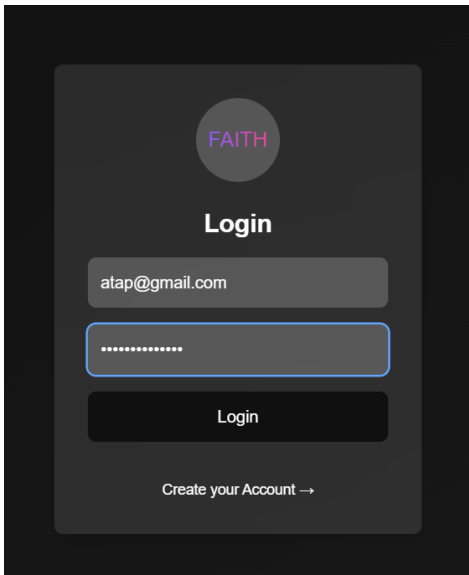Main Chat Interface to check prominent four LLM /AI models (gpt-4, gpt-4o, gpt-4o-mini, gpt-35-turbo-16k) and any other custom LLM /AI model too.

User authentication and chat history maintenance for each LLM model along with corresponding statistics, finding hallucination with reason and confident scores for every response generated.

Detailed analytics for every type of LLM/AI model including hallucination percentage and several other AI evaluator metrics like violence, bias, sexual, self harm, etc.,
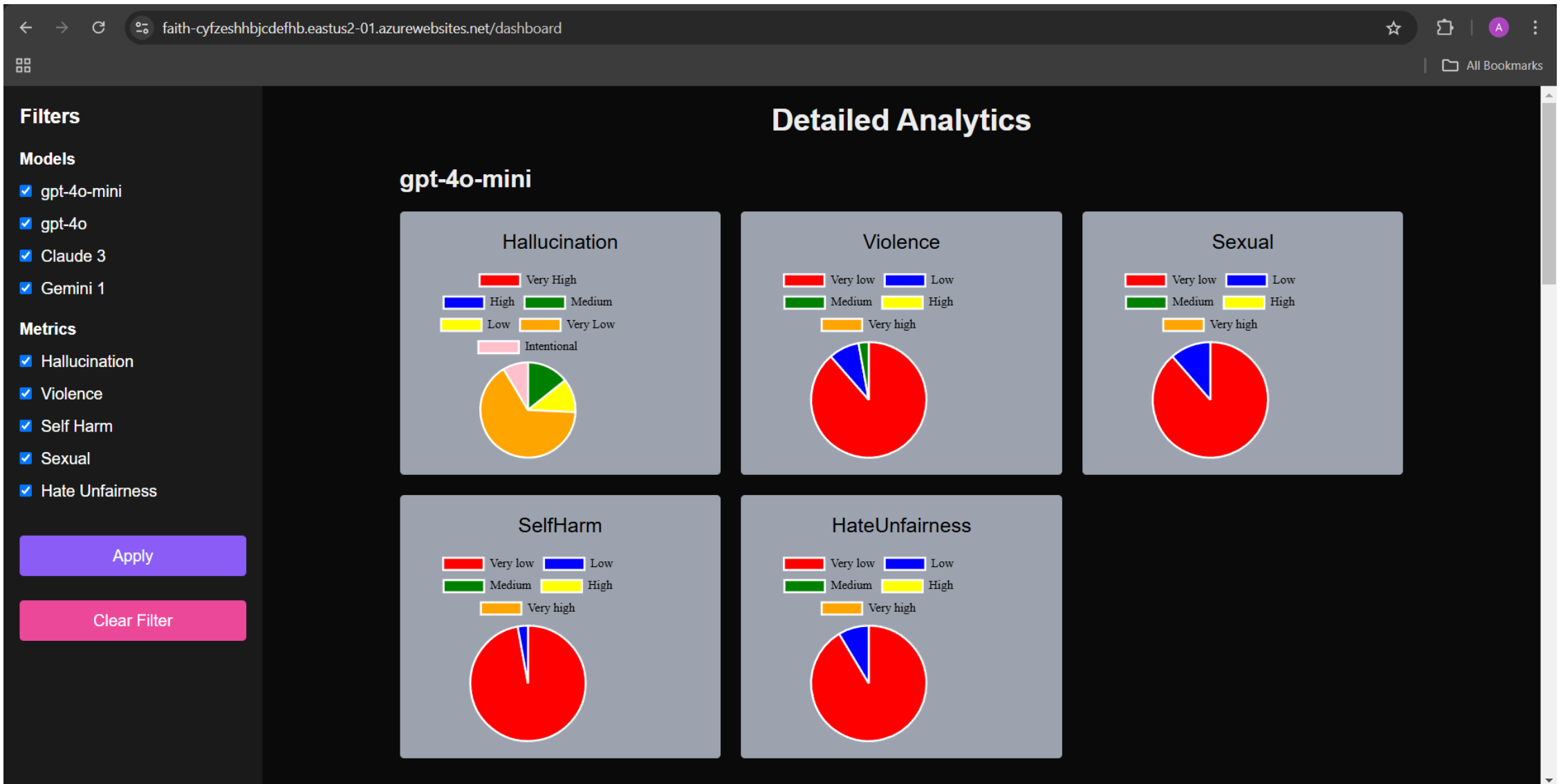
Test our platform my comparing your hallucination results with our predictions and have a look at our performance including true positive, true negative, false positive and false negative counts in form of confusion matrix.

Main Chat Interface to check prominent four LLM /AI models (gpt-4, gpt-4o, gpt-4o-mini, gpt-35-turbo-16k) and any other custom LLM /AI model too.

User authentication and chat history maintenance for each LLM model along with corresponding statistics, finding hallucination with reason and confident scores for every response generated.

Detailed analytics for every type of LLM/AI model including hallucination percentage and several other AI evaluator metrics like violence, bias, sexual, self harm, etc.,

Test our platform my comparing your hallucination results with our predictions and have a look at our performance including true positive, true negative, false positive and false negative counts in form of confusion matrix.

Thank You

FAITH : **F**ramework for **AI** **I**ntegrity and **T**esting **H**allucinations

https://faith-cyfzeshhbjcdefhb.eastus2-01.azurewebsites.net/