MICHIGAN STATE UNIVERSITY

Project Plan Presentation

AI System Testing Framework

The Capstone Experience

Team Ally

Andrew Dagher
Ethan Gomez
Vu Ho
Gabe Moraru
Michael Plante
Amit Wagh

Department of Computer Science and Engineering
Michigan State University

Spring 2025

From Students…
…to Professionals

Project Sponsor Overview

- Ally is an online bank and financial services company
- Top 25 U.S. financial holding company
- Seeking to utilize AI and ML to automate business processes

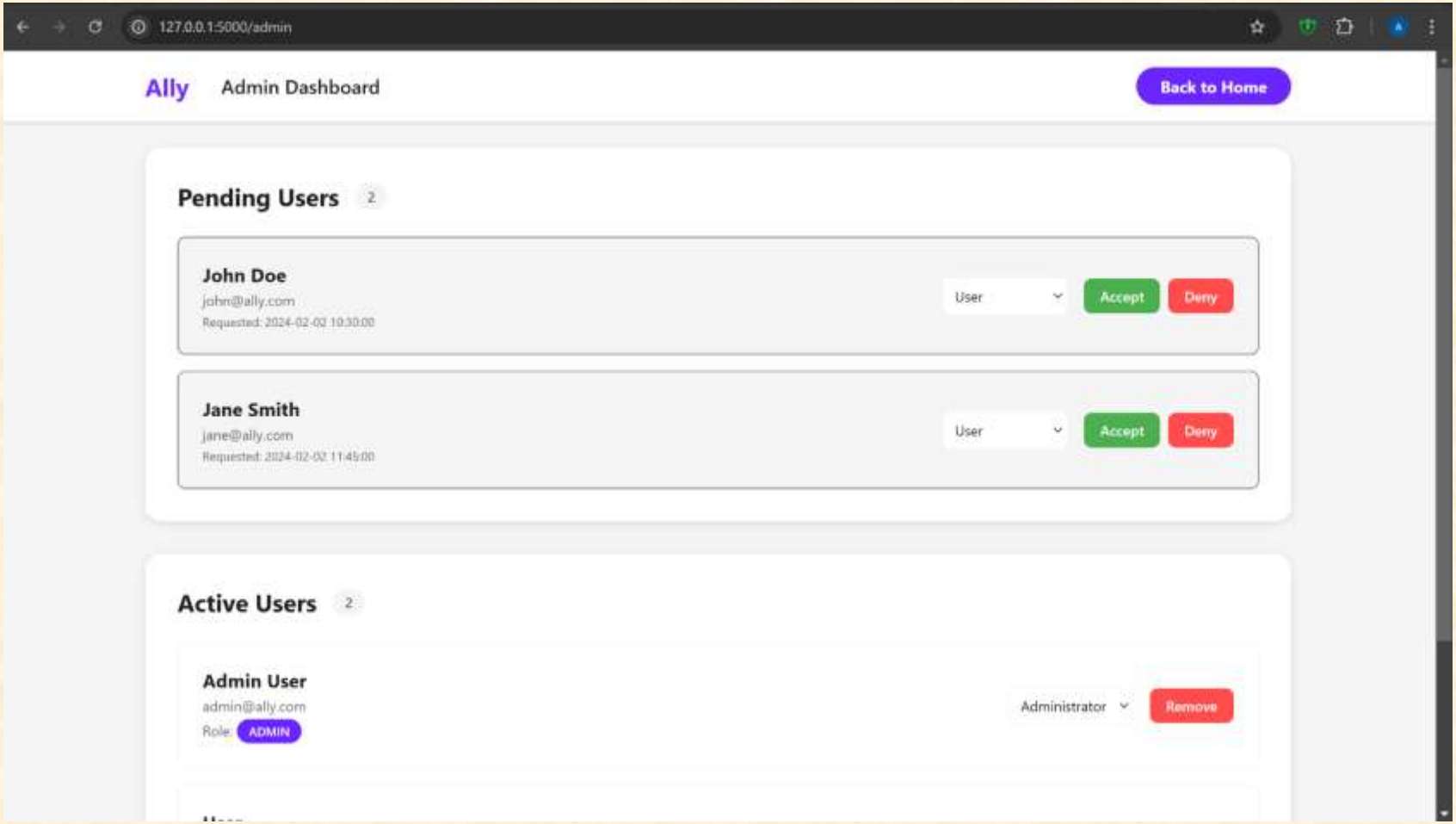
Project Functional Specifications

- How reliable is generative AI?
- Build a testing framework
- Evaluate the performance of generative AI in specific business use cases
- Provide users evaluation metrics to determine if generative AI excels in a particular area

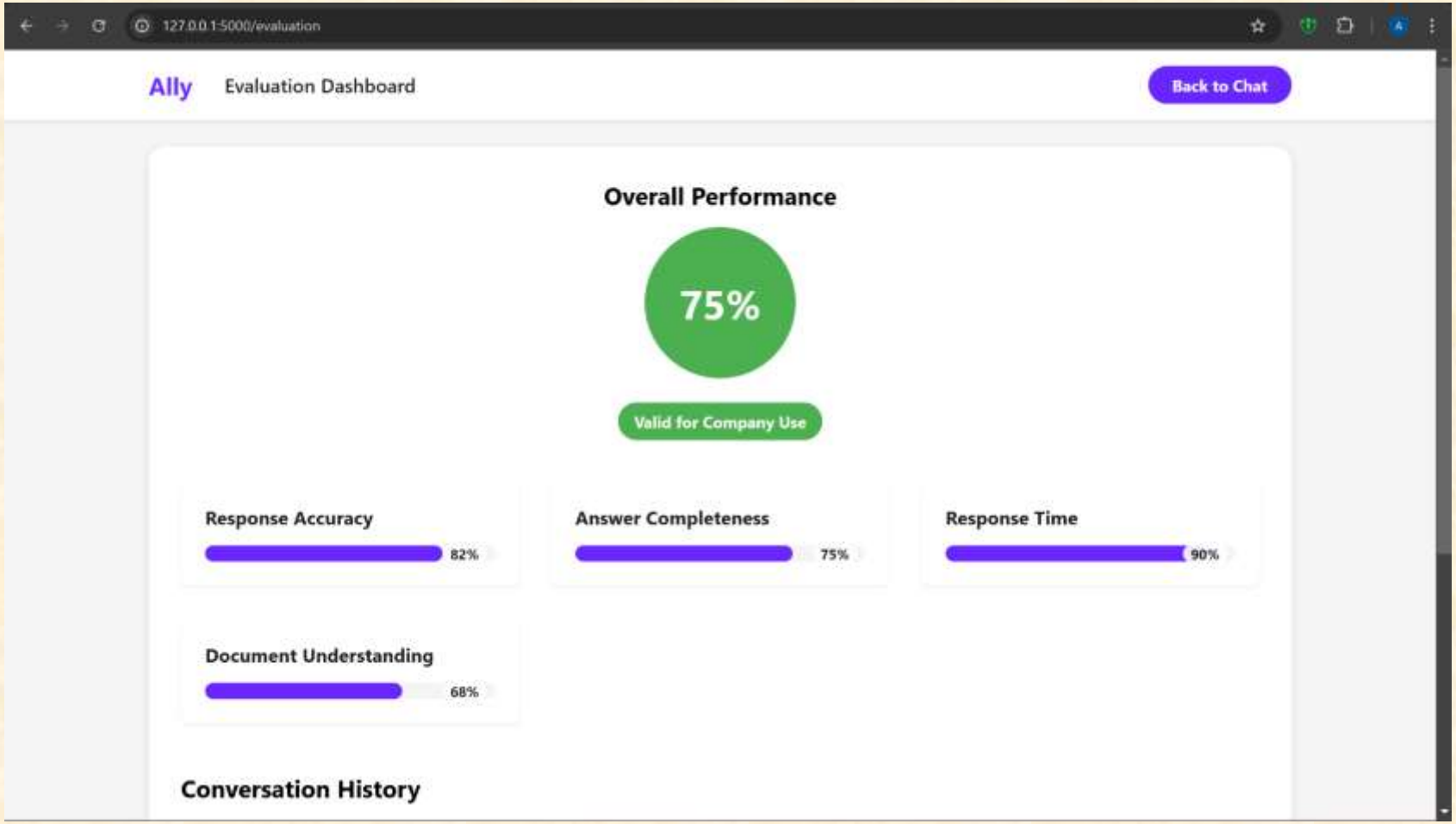
Project Design Specifications

- Homepage includes a conversation box, text input field, send button, and a file upload button
- Evaluation page includes a confidence score, and evaluation metrics (Response Accuracy, Answer Completeness, Response Time, Document Understanding)
- Conversation history displays previous responses and assessments
- Administrative page allows admins to manage accounts

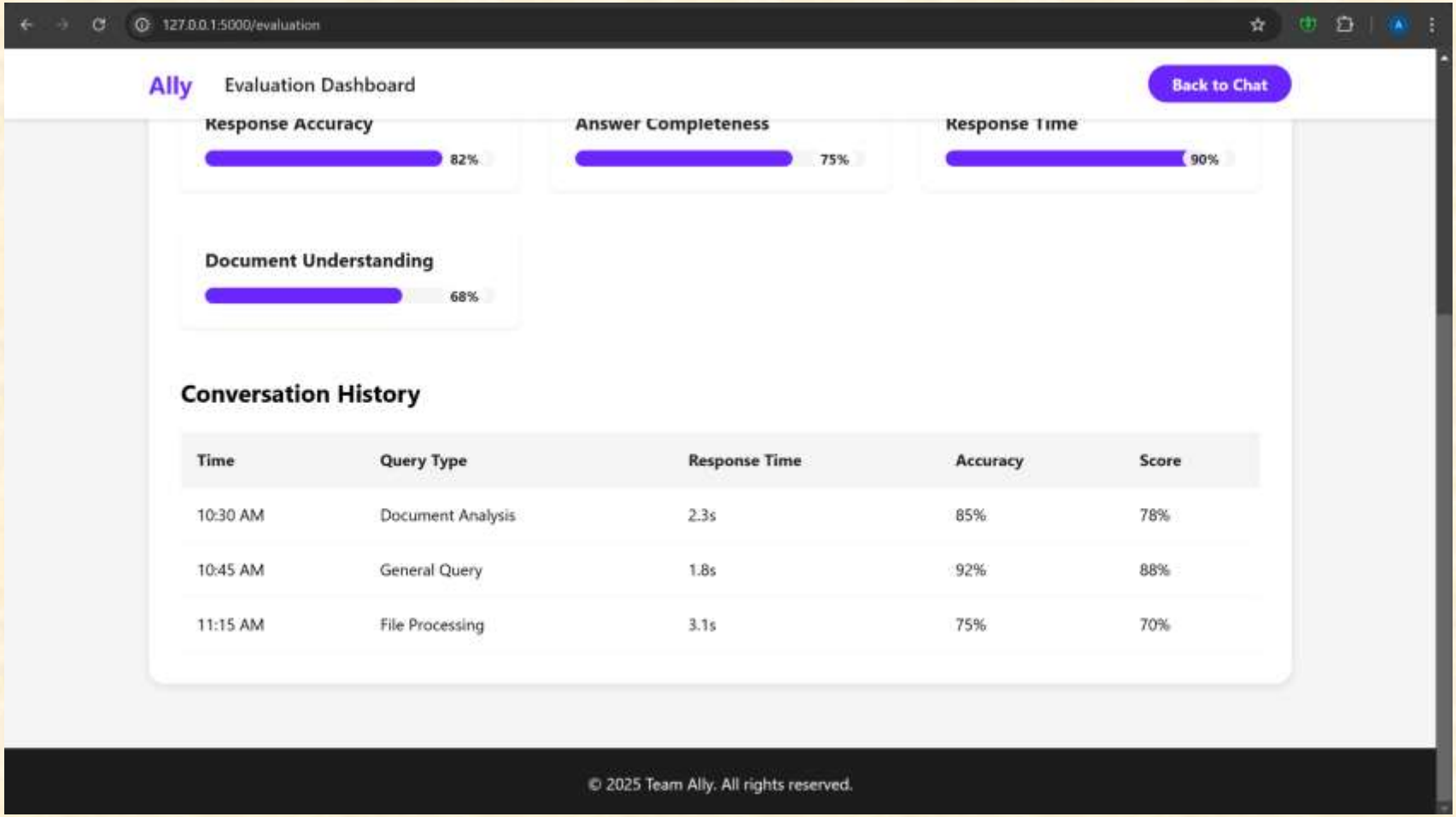
Screen Mockup: Admin Dashboard



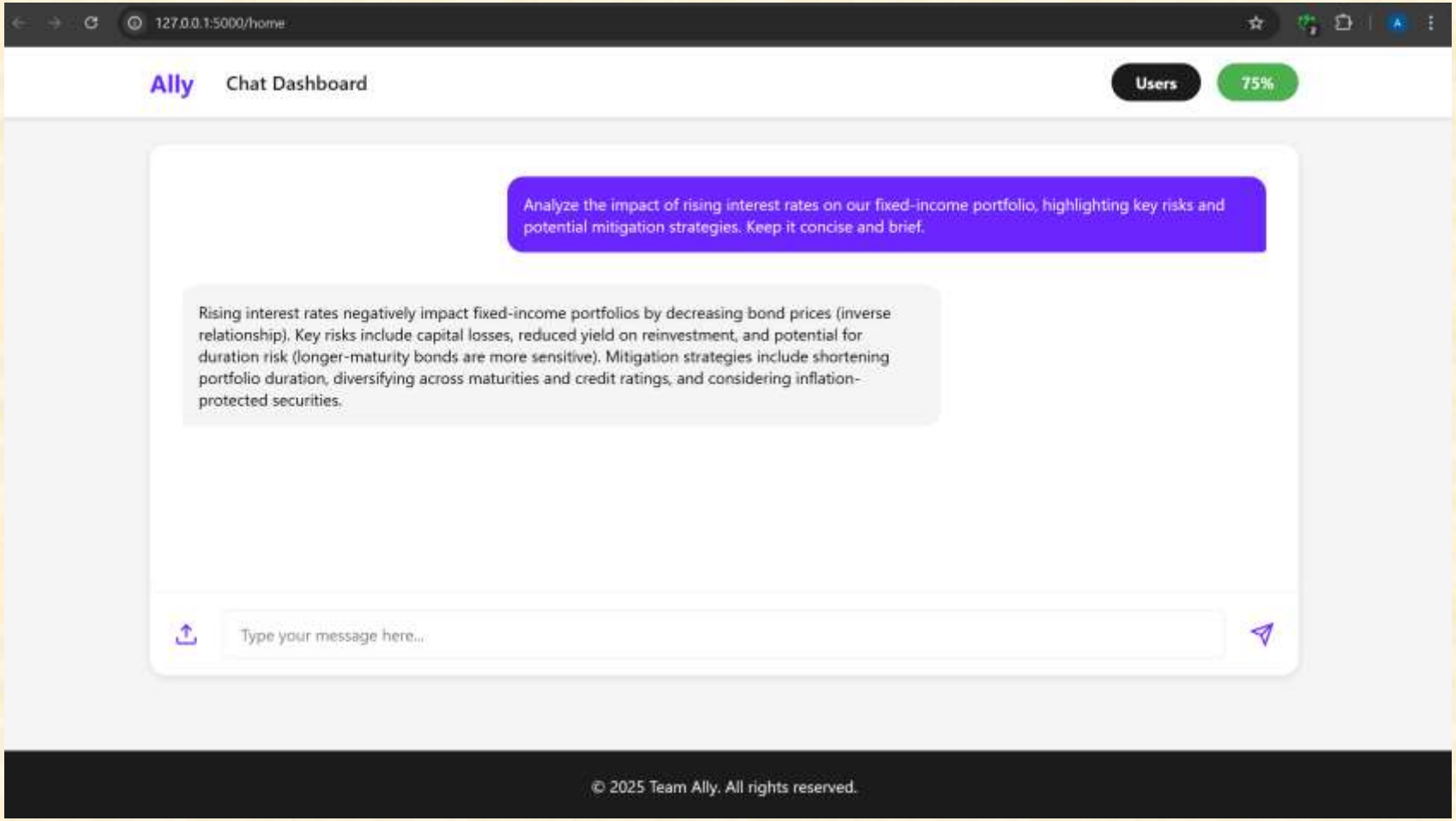
Screen Mockup: Evaluation Dashboard (Performance)



Screen Mockup: Evaluation Dashboard (Conversation History)



Screen Mockup: Chat Dashboard

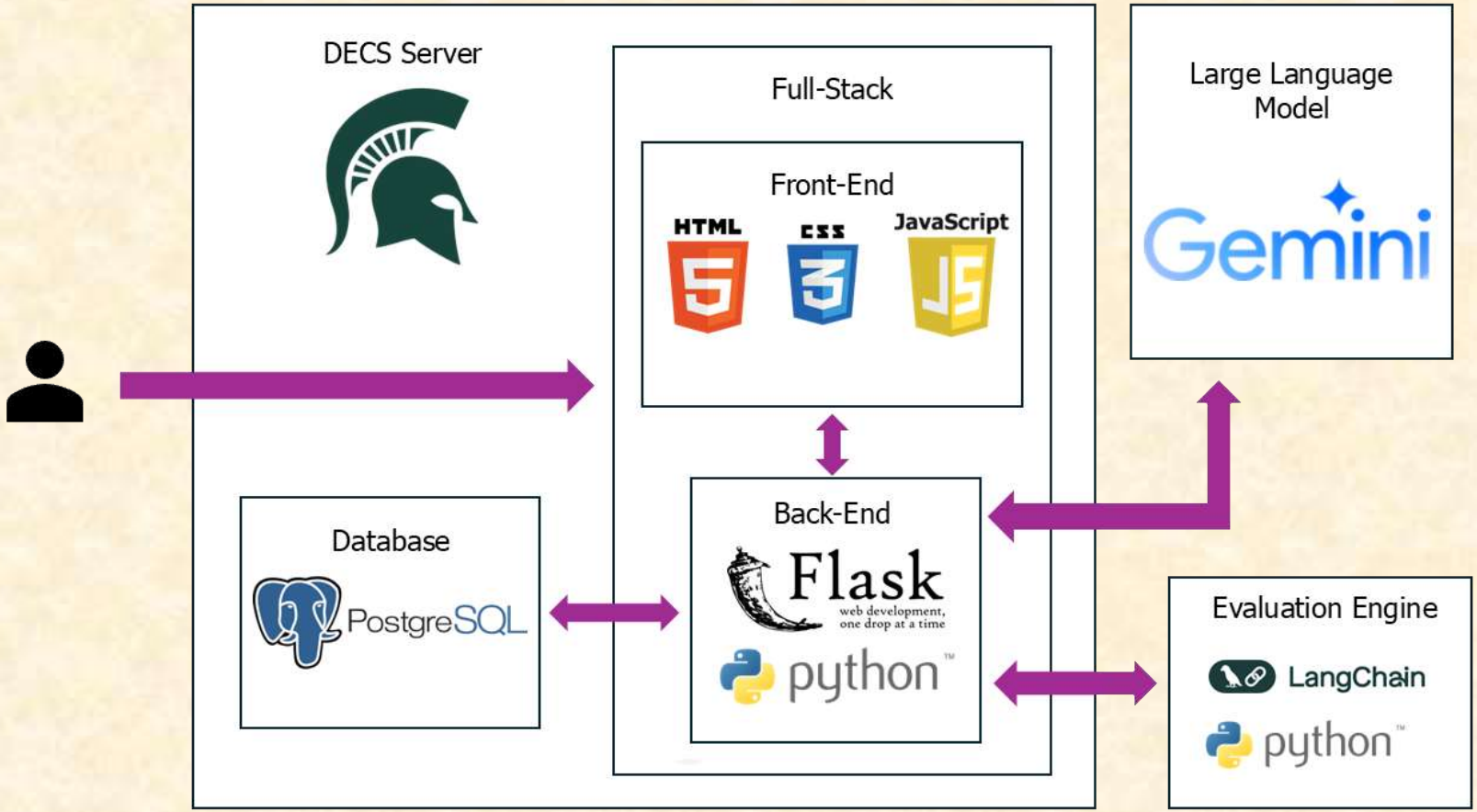


Project Technical Specifications

Development Environment: VSCode

- Hosting: MSU DECS
- Back-End: Python, Flask
- Database: PostgreSQL
- Front-End: HTML, CSS, JS
- Large Language Model: Google Gemini API
- Evaluation Engine: LangChain

Project System Architecture



Project System Components

- Hardware Platforms
  - Michigan State DECS
- Software Platforms / Technologies
  - Linux
  - Python
  - Flask
  - PostgreSQL
  - HTML, CSS, JS
  - Google Gemini API
  - LangChain

Project Risks

- Integration Issues
  - Integrating the front-end, back-end, and database may be difficult
  - Develop full-stack application with limited functionality to test for integration issues
- Noise Handling
  - The user may provide input or files that are unclear, invalid, or irrelevant. The application needs to be resilient to these inputs and not provide the user incorrect information
  - Develop a system to notify the user when their input is problematic
- Evaluation Criteria Implementation
  - A useful evaluation method must correspond to the usefulness of AI in particular use cases
  - Experiment with a variety of evaluation methods and use synthetic data to verify accuracy
- File Handling
  - Files could be provided along with user input for a use case. The files must correctly uploaded, stored, and parsed
  - Identify file types which will need to be supported and test relevant features

Questions?