

LLMao: Retrieval Augmented Generation using Large Language Models for Toxicology

Omkar Chavan, Ashley Fenton, Kenny Lam, Ali Mahmoud, Jhanvi Rana

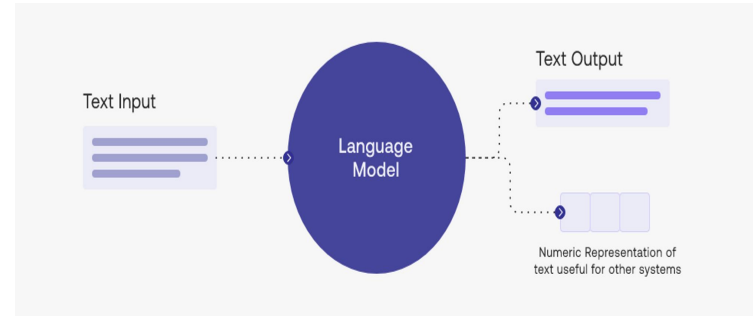
BE BOUNDLESS

W

Introduction

What are LLMs?

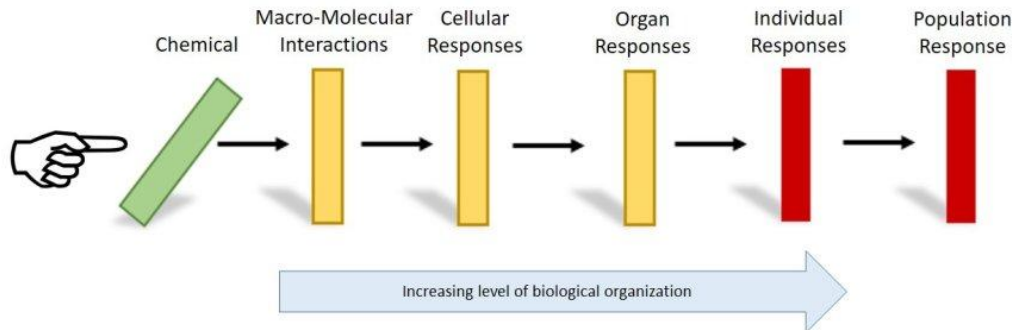
- A large language model is a deep learning algorithm that summarizes, translates and generates text to convey ideas and concepts
- Notable examples include OpenAI's GPT-3 and GPT-4



Challenge

Incompetency of LLMs

- LLMs often give inaccurate and made-up responses to scientific questions
- This problem is experienced more heavily in toxicology where ChatGPT does not have access to the **Adverse Outcome Pathway (AOP)** database



Objective

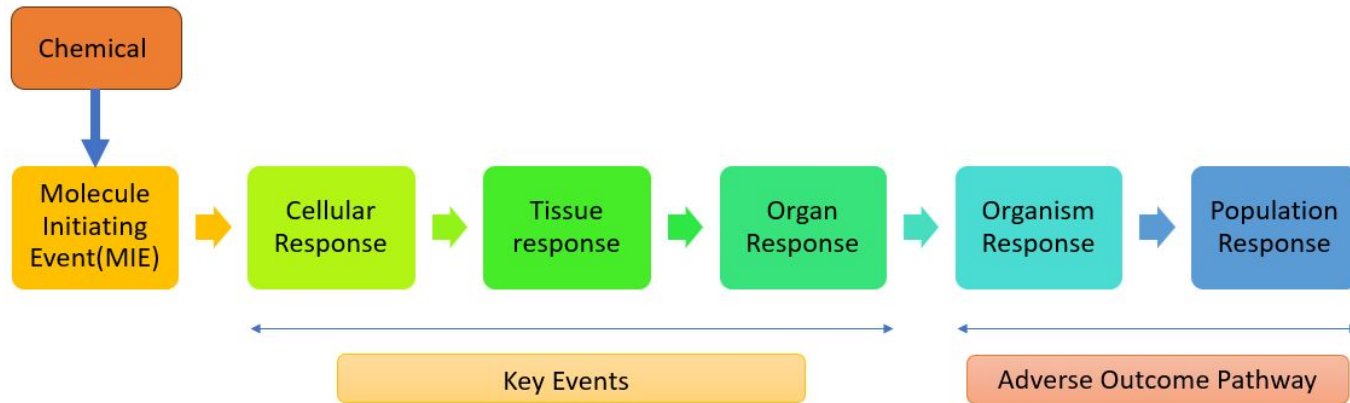
Bridge the gap
between academia
and general public

- Build LLM that utilizes AOP database from EPA
- Combine that database with **cloud computing platform** like Amazon Web Services (AWS)
- Utilize AWS tools like **Bedrock, EC2** and **S3** which helps in creating a specialized AI expert in **toxicology**



Adverse Outcome Pathway (AOP)

AOP: biological frameworks that seek to explain how a molecular level interaction with a stressor can lead to large scale adverse effects



General AOP Framework

Our LLM utilizes the AOP database from the EPA (<https://aopdb.epa.gov>)

AOP Demo from Ilm-ao.com

LLMao

About

Dialyze

Adverse Outcome Pathway Diagram Generator

This work is being done as a shared capstone project in the Department of Chemical Engineering at the University of Washington.

What is an Adverse Outcome Pathway (AOP)?

An Adverse Outcome Pathway (AOP) is a conceptual framework that links a molecular-level perturbation to an adverse health effect. It is used to better understand the mechanisms through which chemicals cause harmful effects and to predict outcomes of chemical exposures. The image below shows an example of visually representing AOPs:

```
graph TD; Chemical[Chemical] --> MIE[Molecule Initiating Event MIE]; MIE --> CR[Cellular Response]; CR --> TR[Tissue response]; TR --> OR[Organ Response]; OR --> ORG[Organism Response]; ORG --> PR[Population Response];
```

The diagram illustrates the General AOP Framework. It shows a sequence of events: Chemical (orange box) leads to Molecule Initiating Event (MIE) (orange box), which leads to Cellular Response (green box), Tissue response (green box), Organ Response (green box), Organism Response (teal box), and finally Population Response (blue box). Below the Cellular Response, Tissue response, and Organ Response boxes is a yellow box labeled 'Key Events'. Below the Organism Response and Population Response boxes is a red box labeled 'Adverse Outcome Pathway'. Arrows indicate the flow from left to right between each step.

General AOP Framework

Why are AOPs Important?

AOPs provide a structured way to organize biological information that can be used in chemical safety assessments and regulatory decisions. They help in identifying key events and processes that can be targeted for testing and intervention.

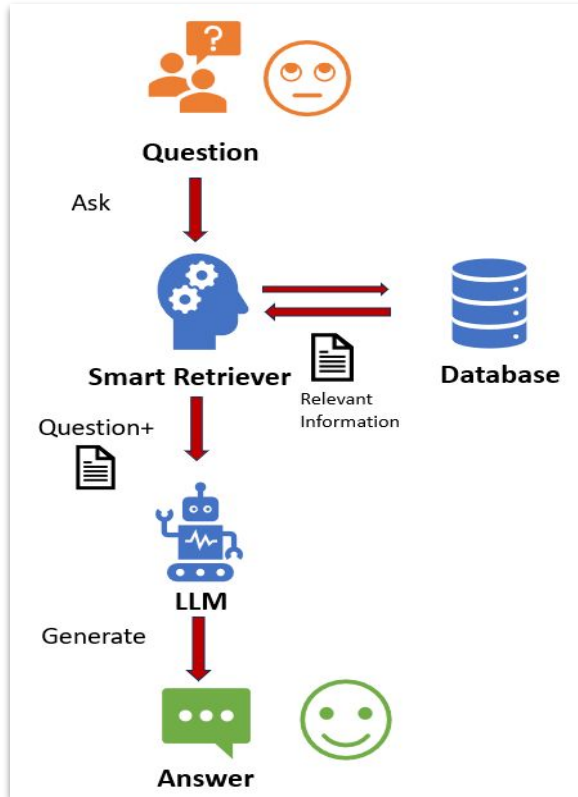
Verified AOPs

The verified AOPs are experimentally validated pathways that have been reviewed and accepted by the scientific community. These pathways provide reliable information on how certain chemicals can lead to adverse health outcomes.

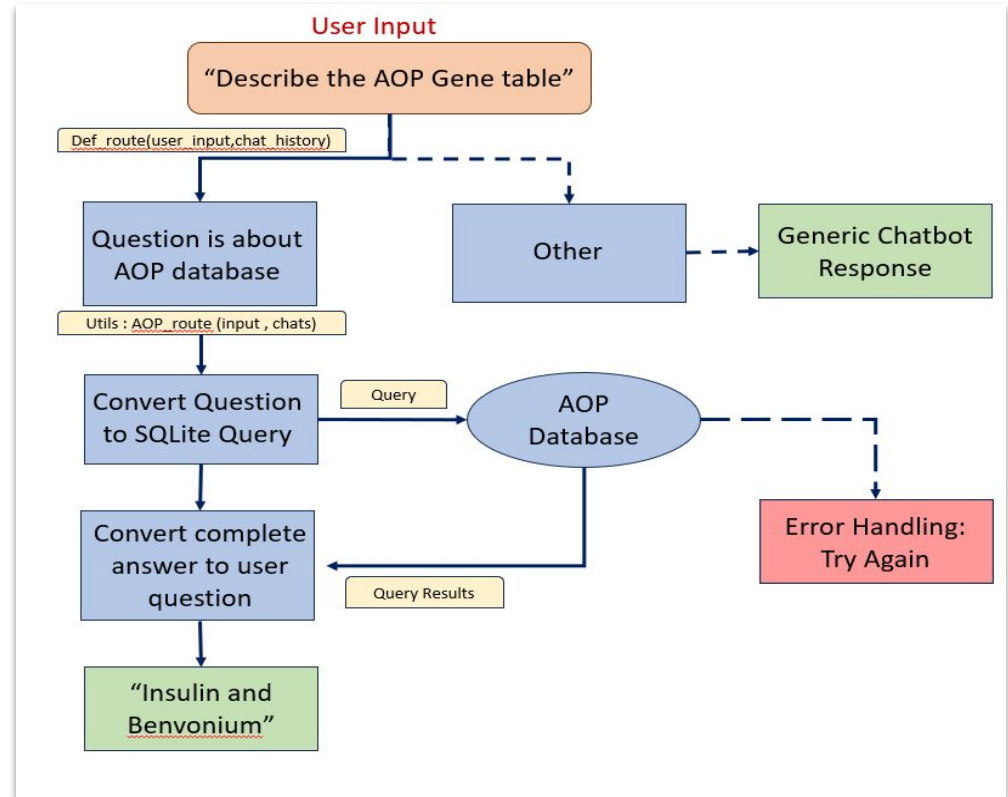
Purpose

Retrieval Augmented Generation (RAG)

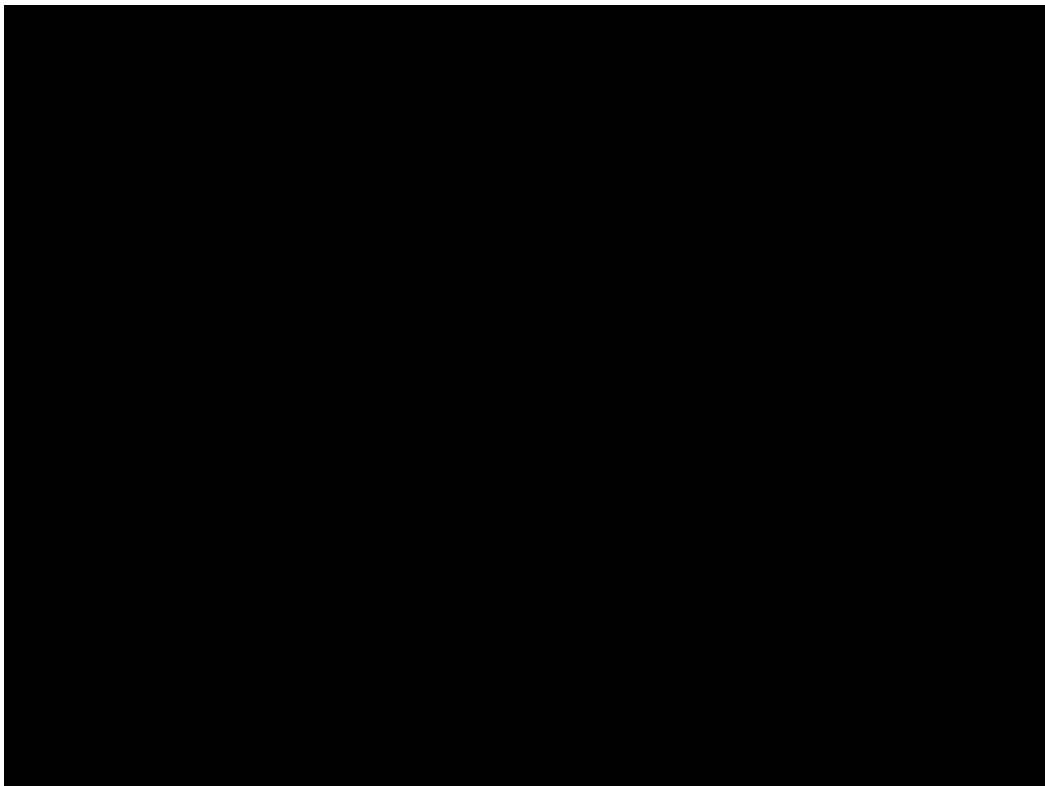
Implementation of RAG:



LLMao using RAG implementation:

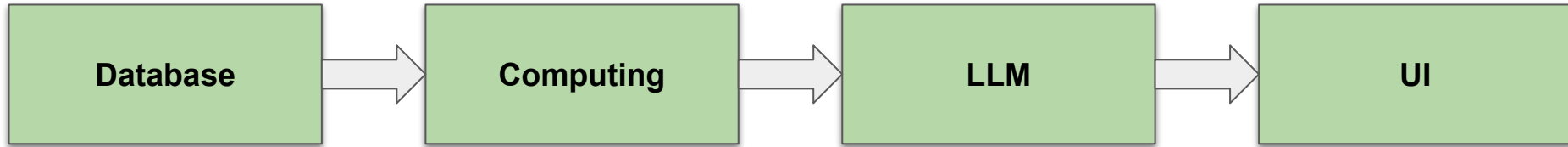


RAG Demo

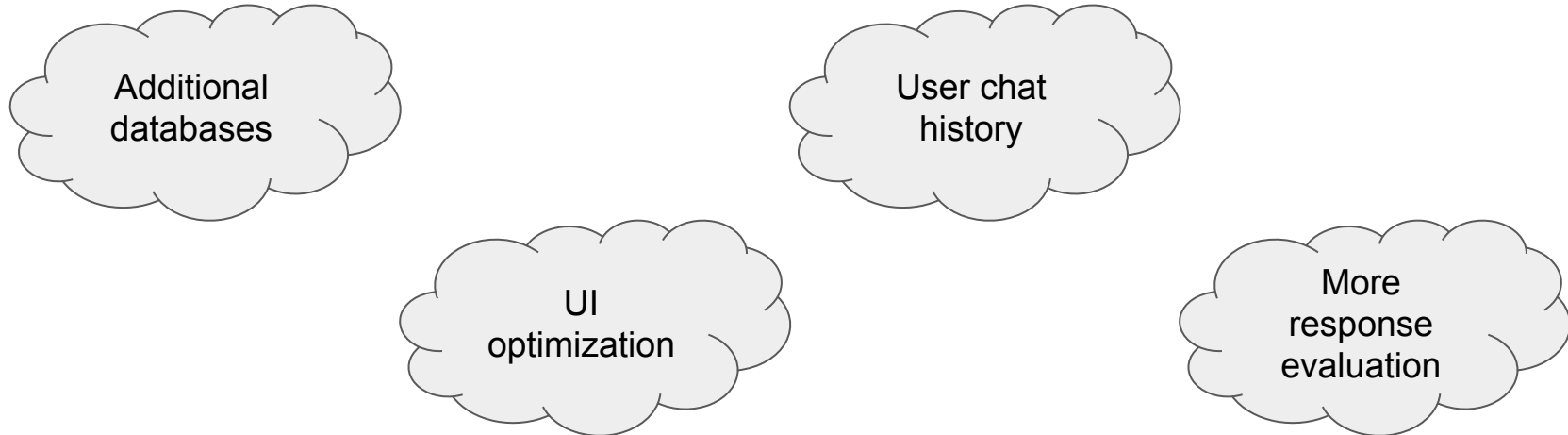


Conclusion & Future Developments

Completed work this quarter:



Future work if this project was continuing:



Try LLMao here:



BE BOUNDLESS

W