LLMao: Retrieval Augmented Generation using Large Language Models for Toxicology

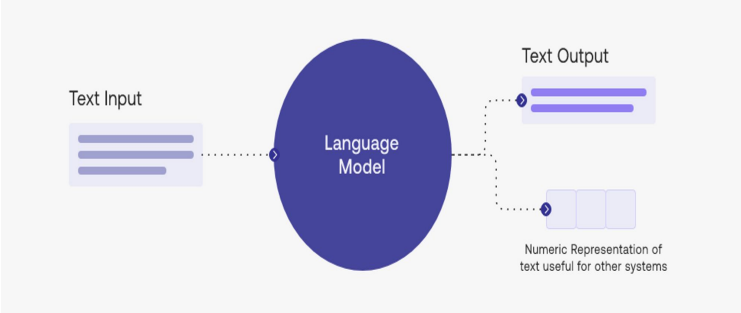
Omkar Chavan, Ashley Fenton, Kenny Lam, Ali Mahmoud, Jhanvi Rana

BE BOUNDLESS

Introduction

What are LLMs?

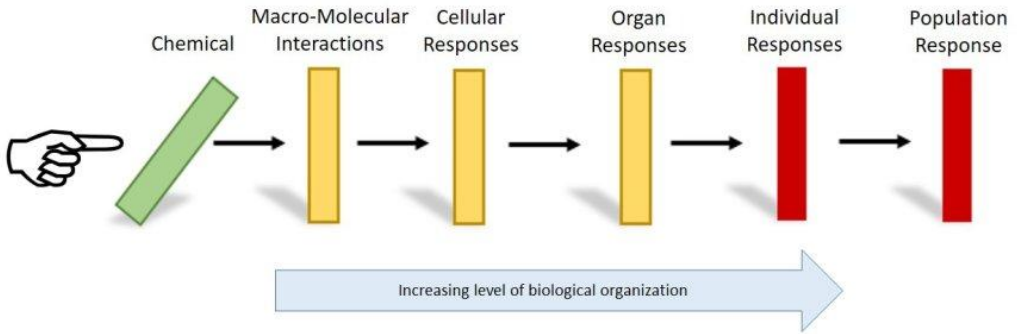
- A large language model is a deep learning algorithm that summarizes, translates and generates text to convey ideas and concepts
- Notable examples include OpenAI's GPT-3 and GPT-4



Challenge

Incompetency of LLMs

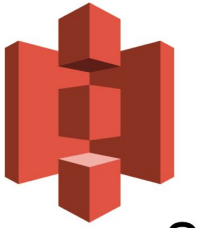
- LLMs often give inaccurate and made-up responses to scientific questions
- This problem is experienced more heavily in toxicology where ChatGPT does not have access to the **Adverse Outcome Pathway** (AOP) database



Objective

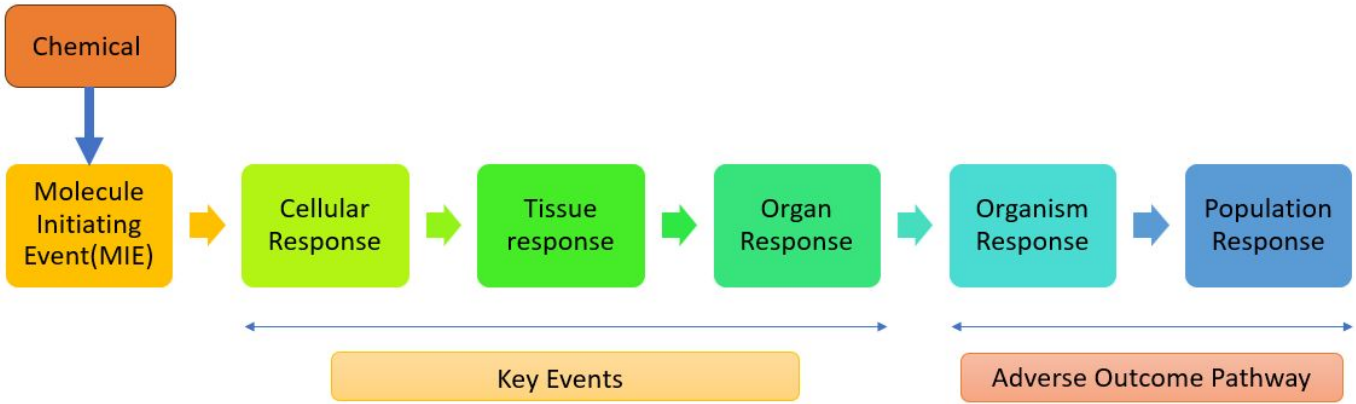
Bridge the gap between academia and general public

- 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)

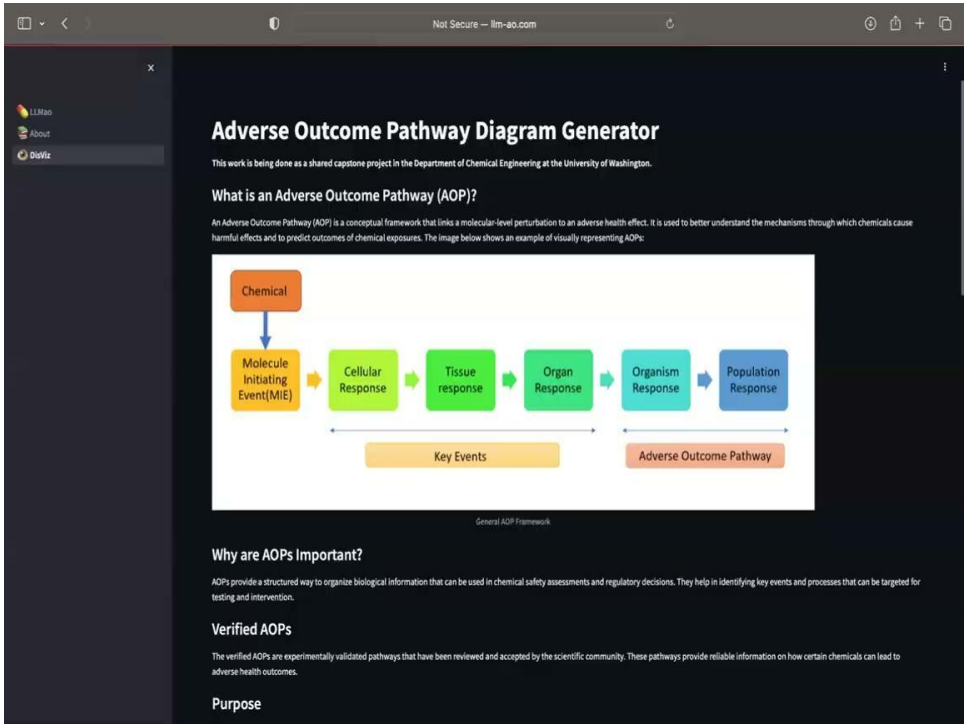
AOP: biological frameworks that seek to explain how a molecular level interaction with a stressor can lead to large scale adverse effects



General AOP Framework

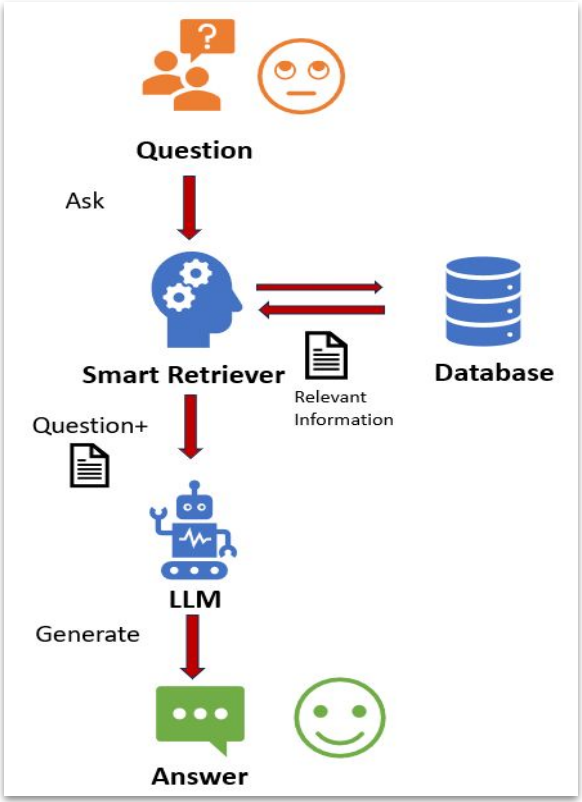
Our LLM utilizes the AOP database from the EPA (https://aopdb.epa.gov)

AOP Demo from llm-ao.com

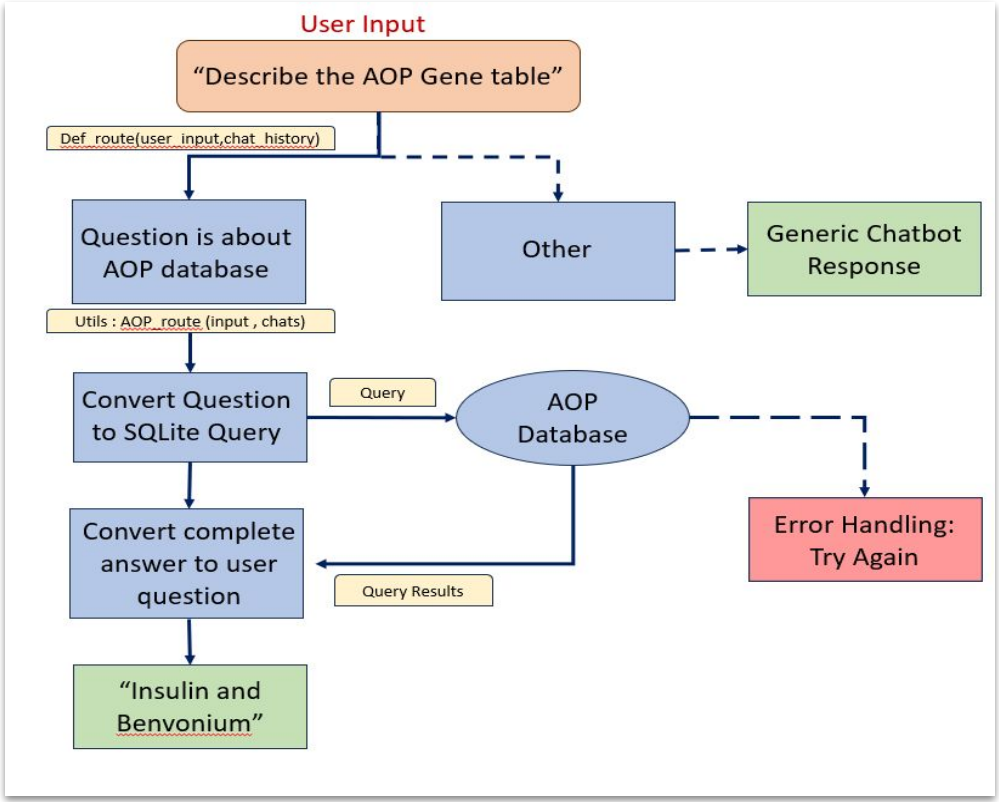


Retrieval Augmented Generation (RAG)

Implementation of RAG:



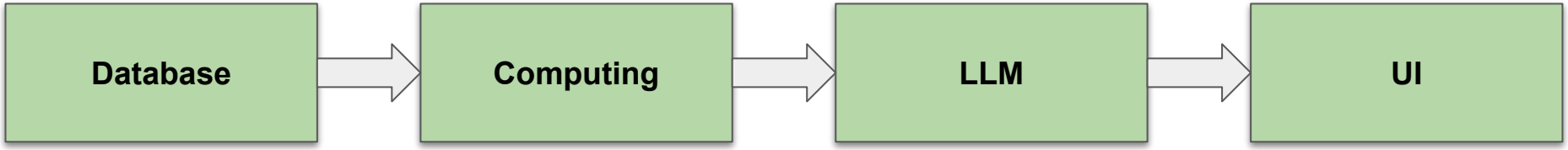
LLMao using RAG implementation:



RAG Demo

Conclusion & Future Developments

Completed work this quarter:



Future work if this project was continuing:

- Additional databases
- User chat history
- UI optimization
- More response evaluation

Try LLMao here:

BE BOUNDLESS