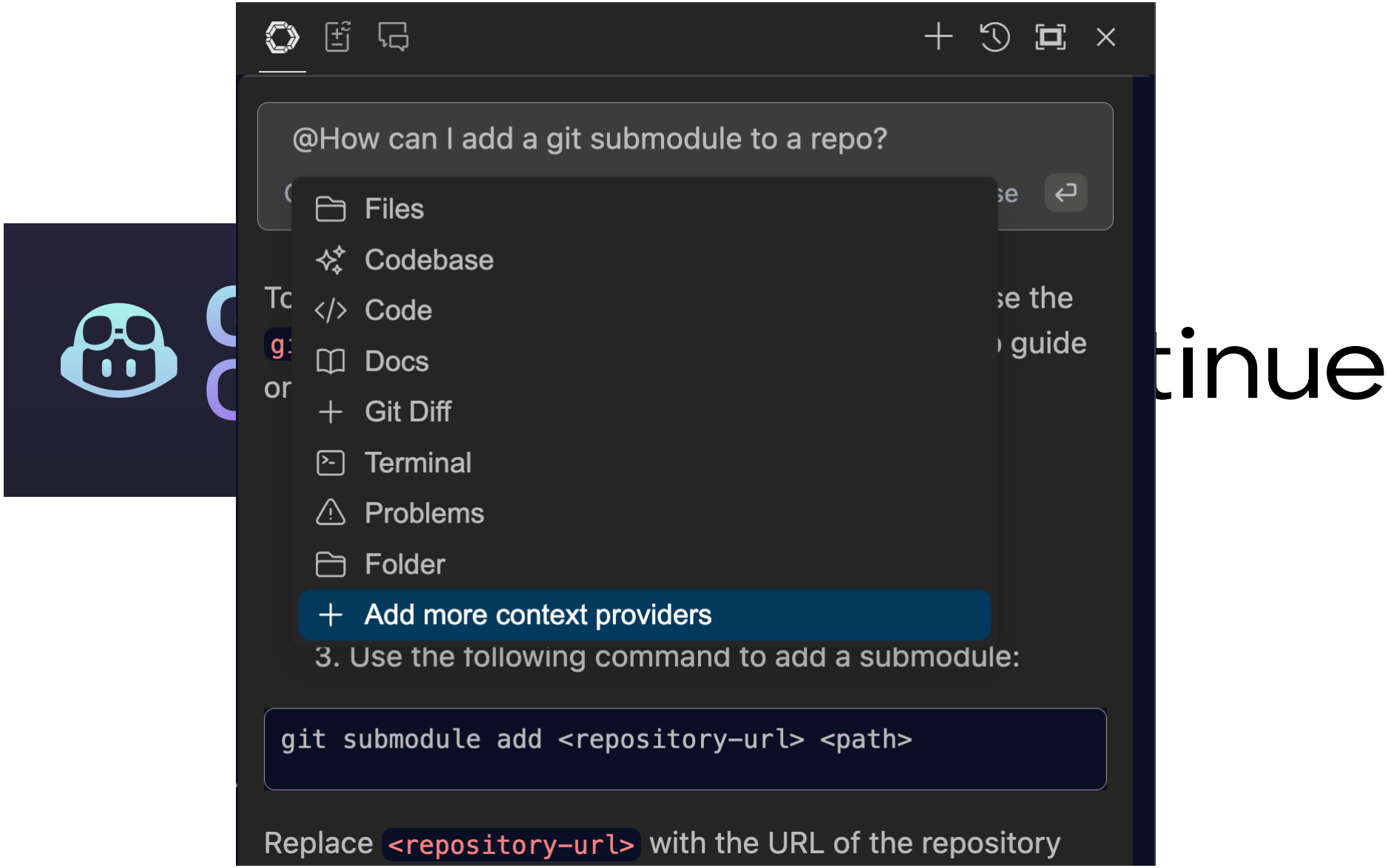
National Institute of
Environmental Health Sciences
Division of Translational Toxicology

ToxPipe: Chatbots and Retrieval-Augmented Generation on Toxicological Data Streams

Trey Saddler

Use Cases for LLMs at NIEHS/DTT



tinue

National Institute of Environmental Health Sciences
Division of Translational Toxicology

ToxPipe

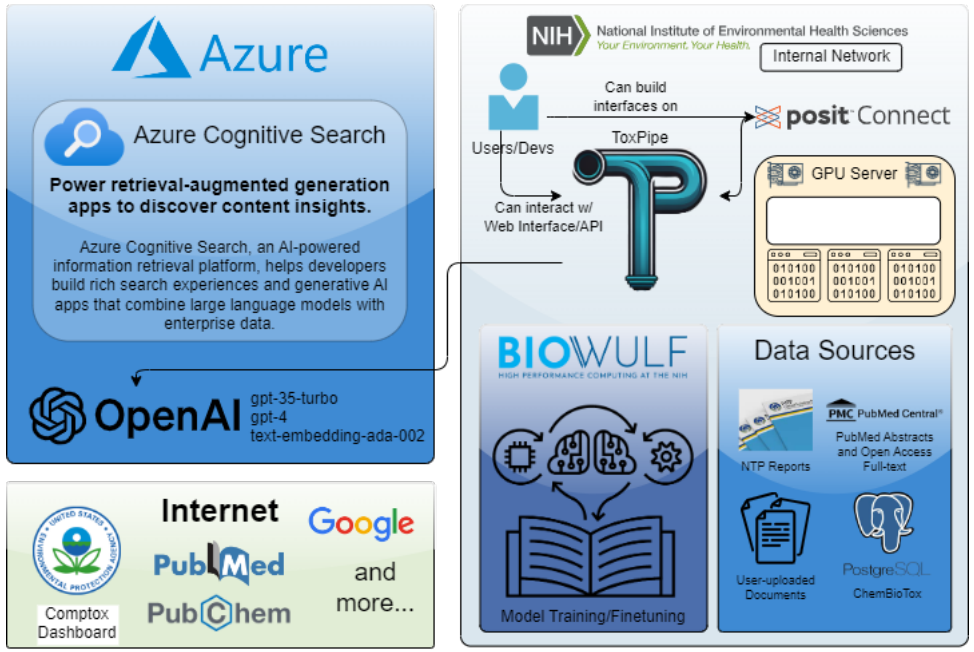
National Institute of Environmental Health Sciences
Division of Translational Toxicology

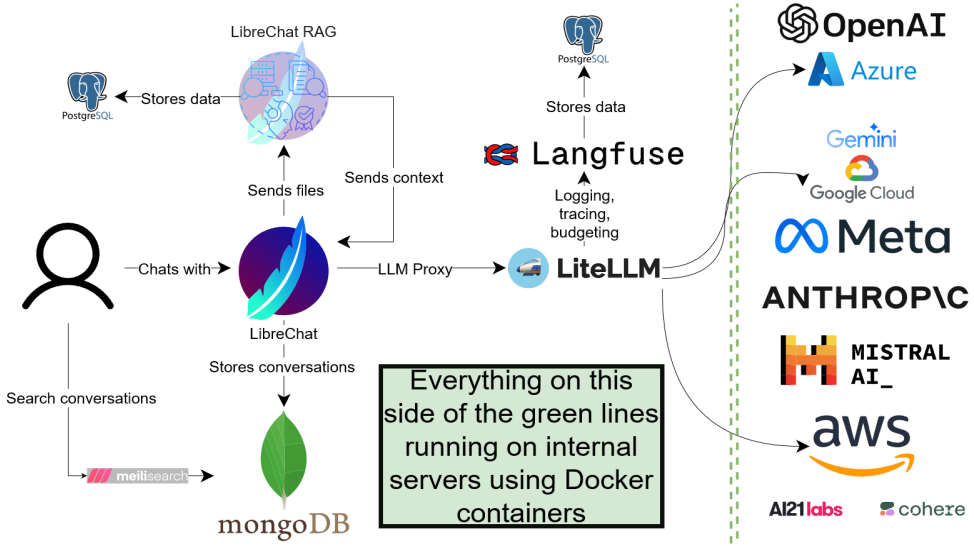
ToxPipe learns from trusted sources with bespoke AI agents → "expert guidance"

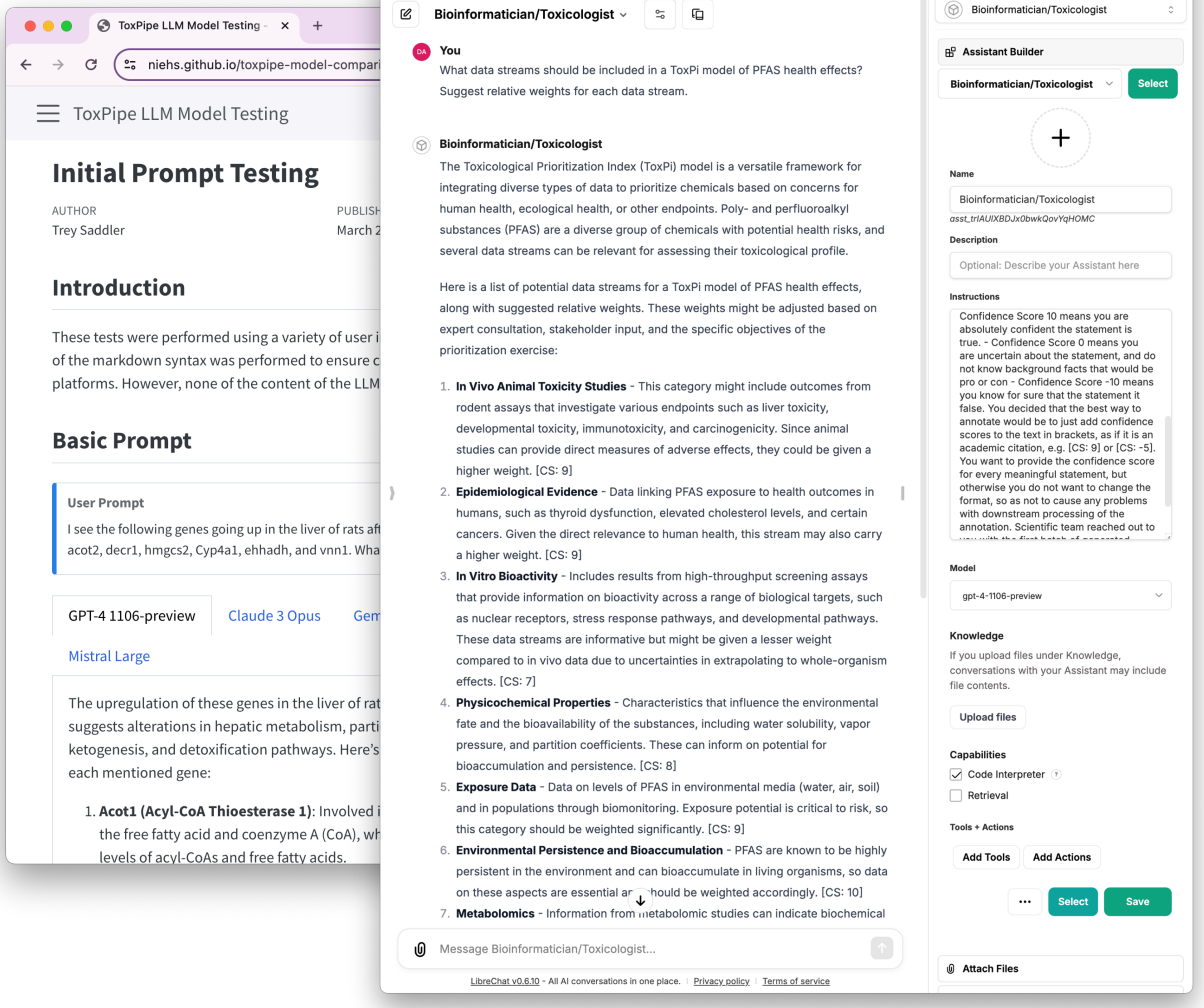

Scott Auerbach


Trey Saddler


David Reif







ToxPipe

Trey Saddler, Parker Combs, Jeremy Erickson, Jonathon Fleming, Charles Schmitt, David Reif, **Scott Auerbach**

National Institute of Environmental Health Sciences
Division of Translational Toxicology

LibreChat

Meet the New LibreChat Resources Hub! 🚀

LibreChat

Docs Blog Demo Changelog Search... CTRL K

Librechat

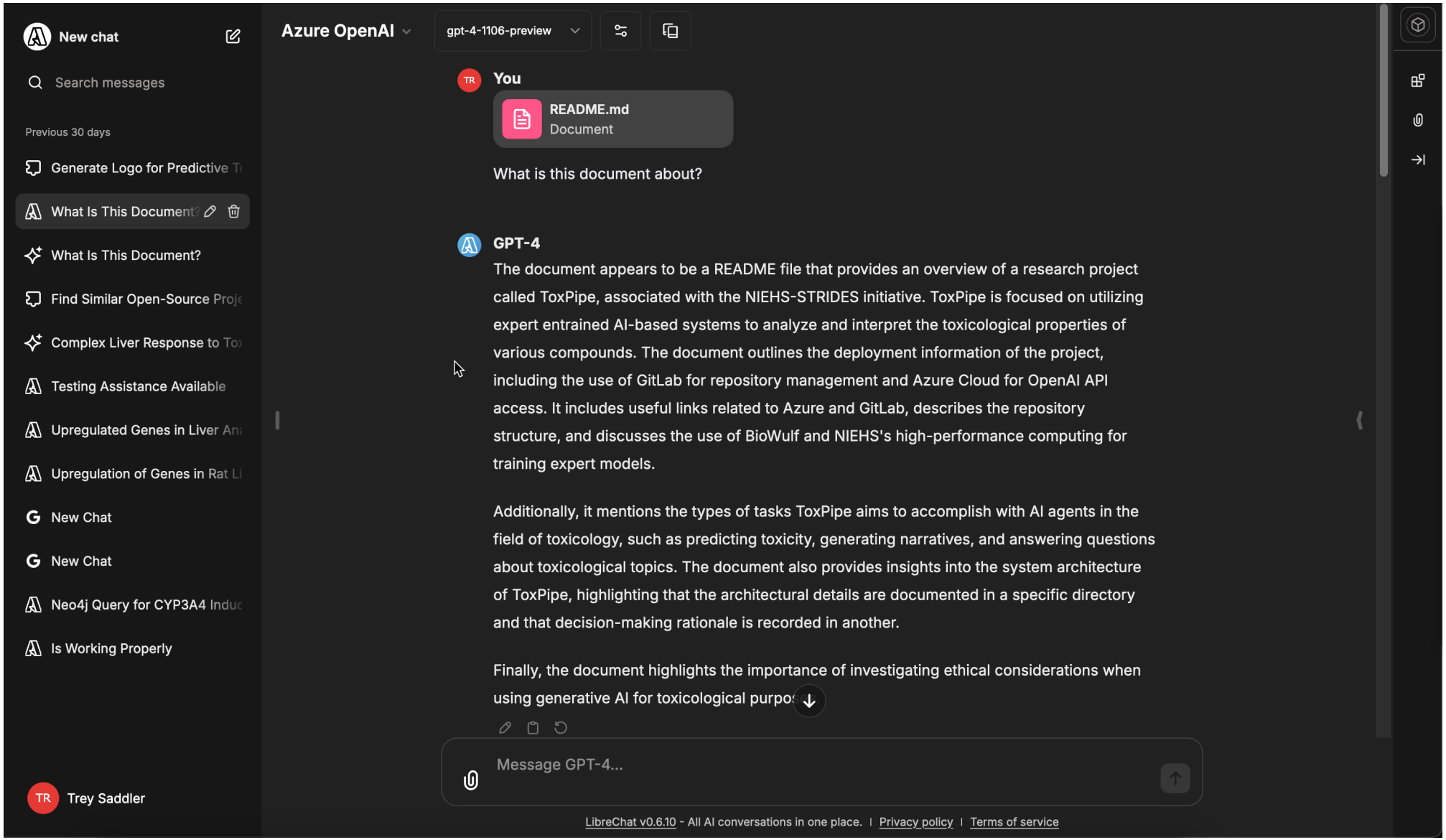
Every AI in One Place
Built for Everyone

LibreChat is a free, open source AI chat platform. This Web UI offers vast customization, supporting numerous AI providers, services, and integrations. Serves all AI Conversations in one place with a familiar interface, innovative enhancements, for as many users as you need.

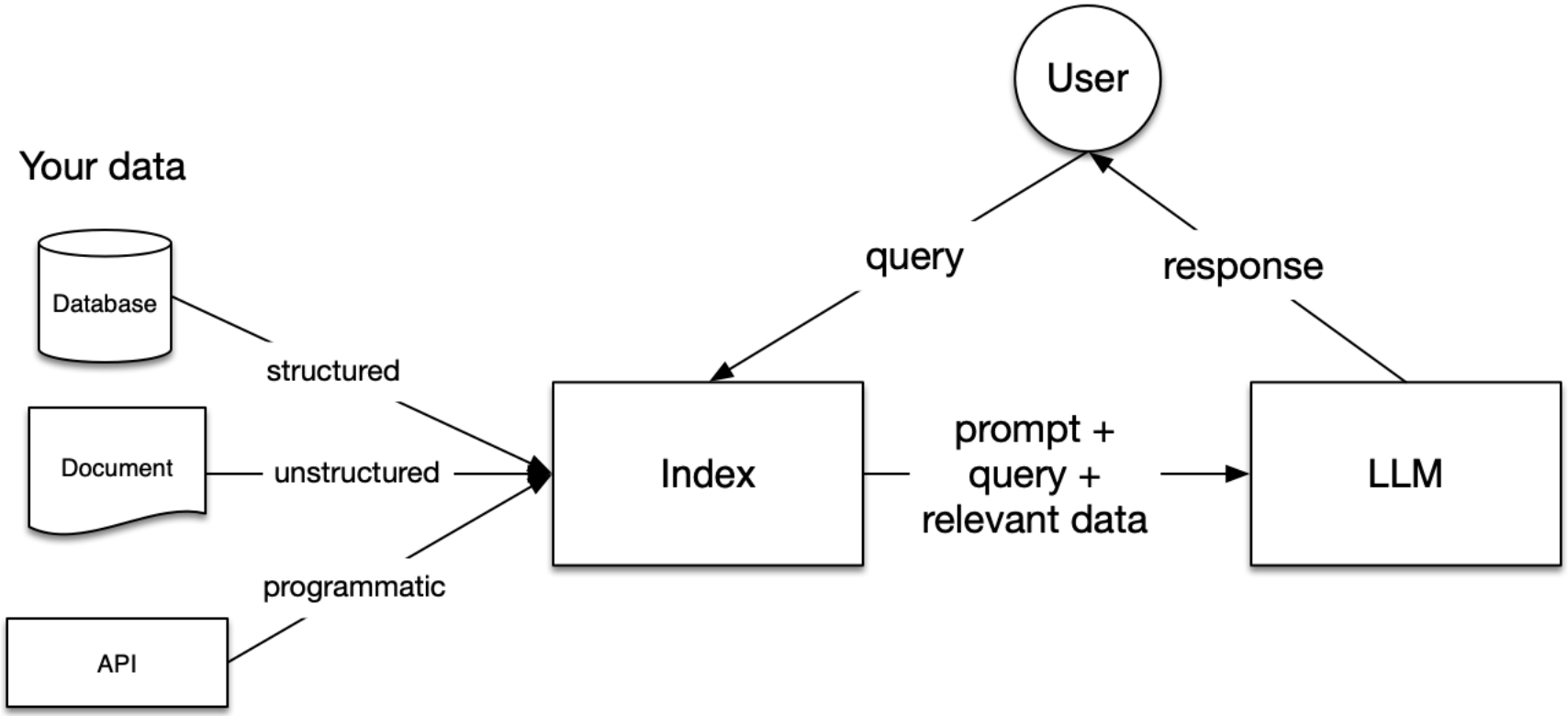
Get Started → GitHub ↗

National Institute of Environmental Health Sciences
Division of Translational Toxicology

ToxPipe LibreChat



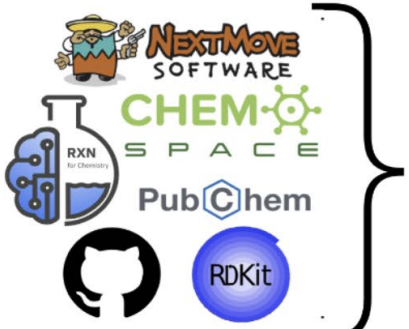
RAG Techniques



Chem Crow

APIs

Tools



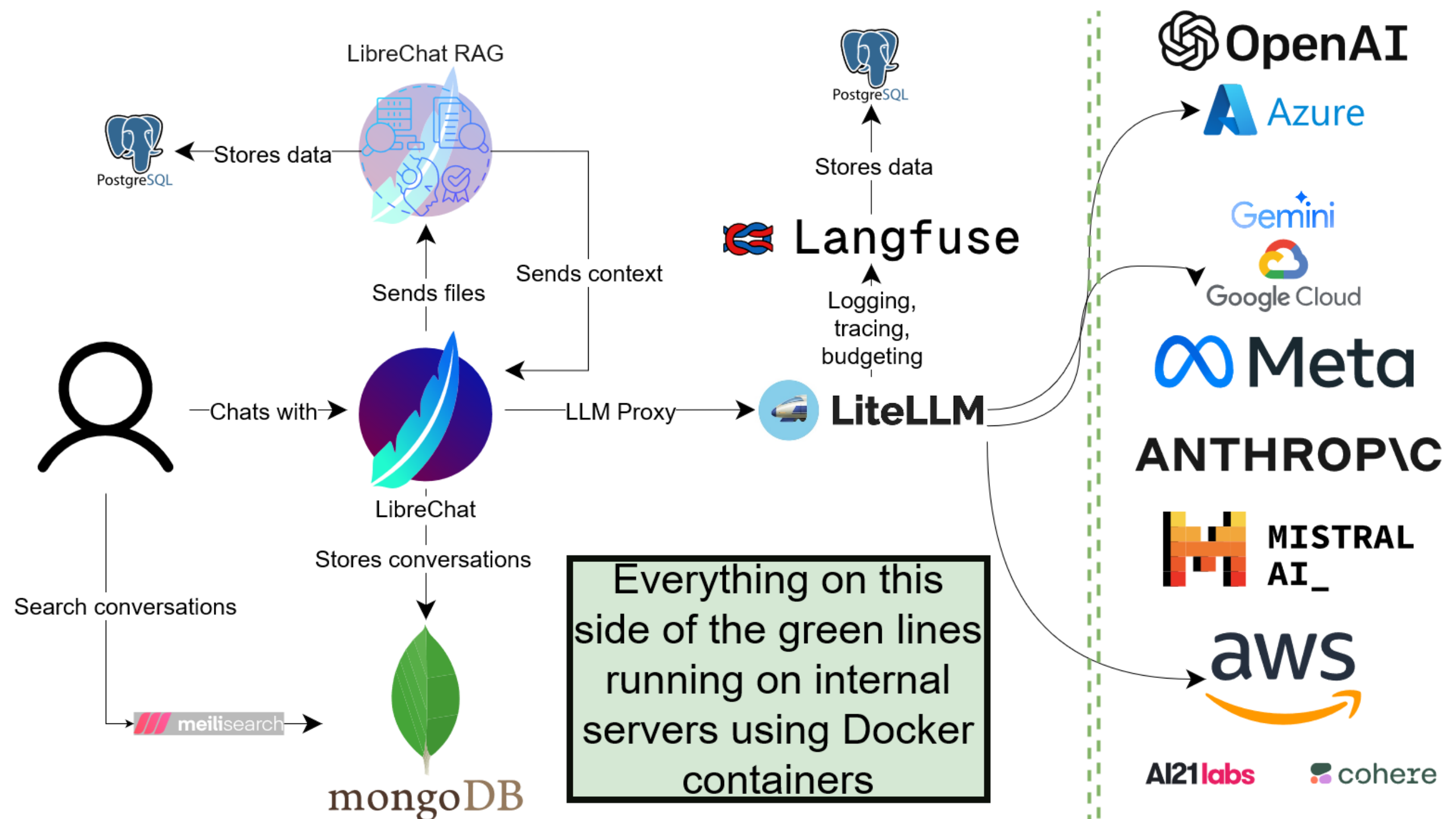
ChemCrow
LangChain
+ LLMs

- *Synthesize target*
- *Explain mechanism*
- *Find similar molecules*

User input / tasks

Tool-informed final answer

Automatic iterative LLM + tools workflow


LibreChat RAG
PostgreSQL
Stores data
Sends files
Sends context
PostgreSQL
Stores data
Langfuse
Logging, tracing, budgeting
OpenAI
Azure
Gemini
Google Cloud
Meta
ANTHROP\C
MISTRAL AI_
aws
AI21labs
cohere
Chats with
LLM Proxy
LiteLLM
LibreChat
Stores conversations
Search conversations
meilisearch
mongoDB
Everything on this side of the green lines running on internal servers using Docker containers

Data Sources

Semantic Scholar

PubChem

RDKit
Open-Source Cheminformatics and Machine Learning

PubMed

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

Technical Reports

Long-term studies of the toxicologic potential of agents in test animals. Go »

Tox21

Technologies Used

Chem  Crow

OpenAI

LangChain

python

docker

quarto

GitHub

Microsoft Azure

National Institute of Environmental Health Sciences
Division of Translational Toxicology

ToxPipe integrates across the Translational Toxicology Pipeline

MoVIZ visualization and AI-guided results interpretation

Dextr

AI-assisted data extraction

Data / Knowledge Mining

Computational Analysis

OPERA: Open Structure-activity Relationship App

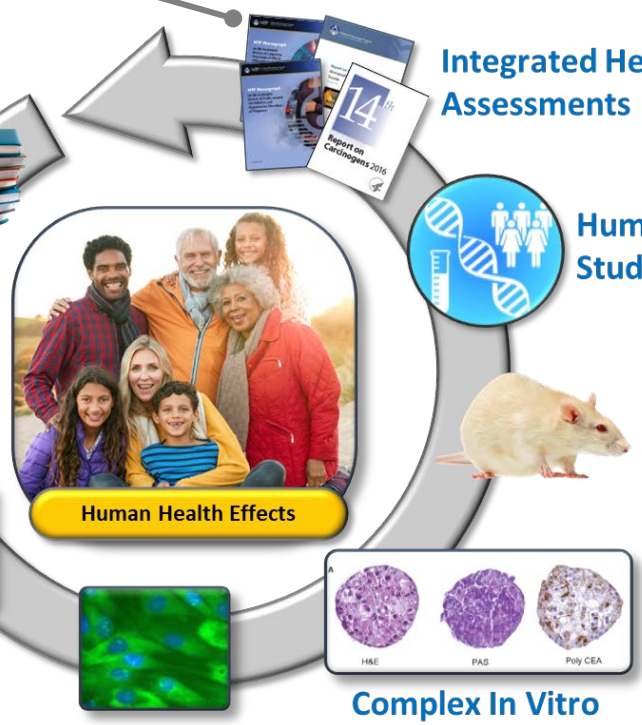
Bioactivity Screening

DNT-DIVER

Developmental NeuroToxicity Data Integration and Visualization Enabling Resource

Tox 21

ICE: Integrated Chemical Environment



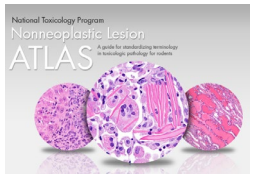
Integrated Health Assessments

Human Studies

Animal Studies

Simple In Vitro Systems

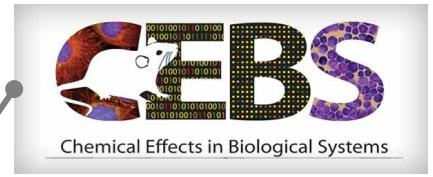
Complex In Vitro Systems



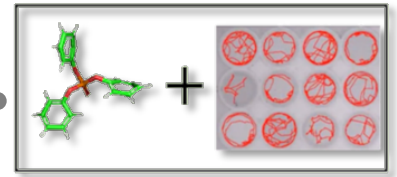
AI-augmented pathology

ToxPipe

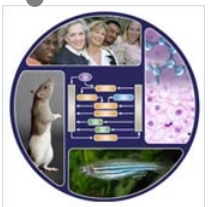
AI chat interface with expert "agents"



Chemical Effects in Biological Systems database



Generative Adversarial Neural Network (GAN) models of behavioral effects

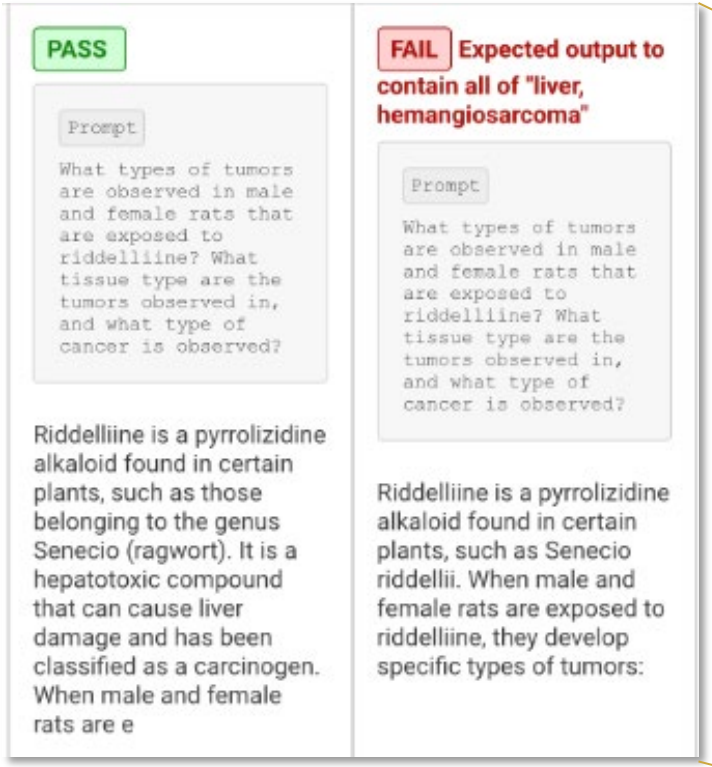


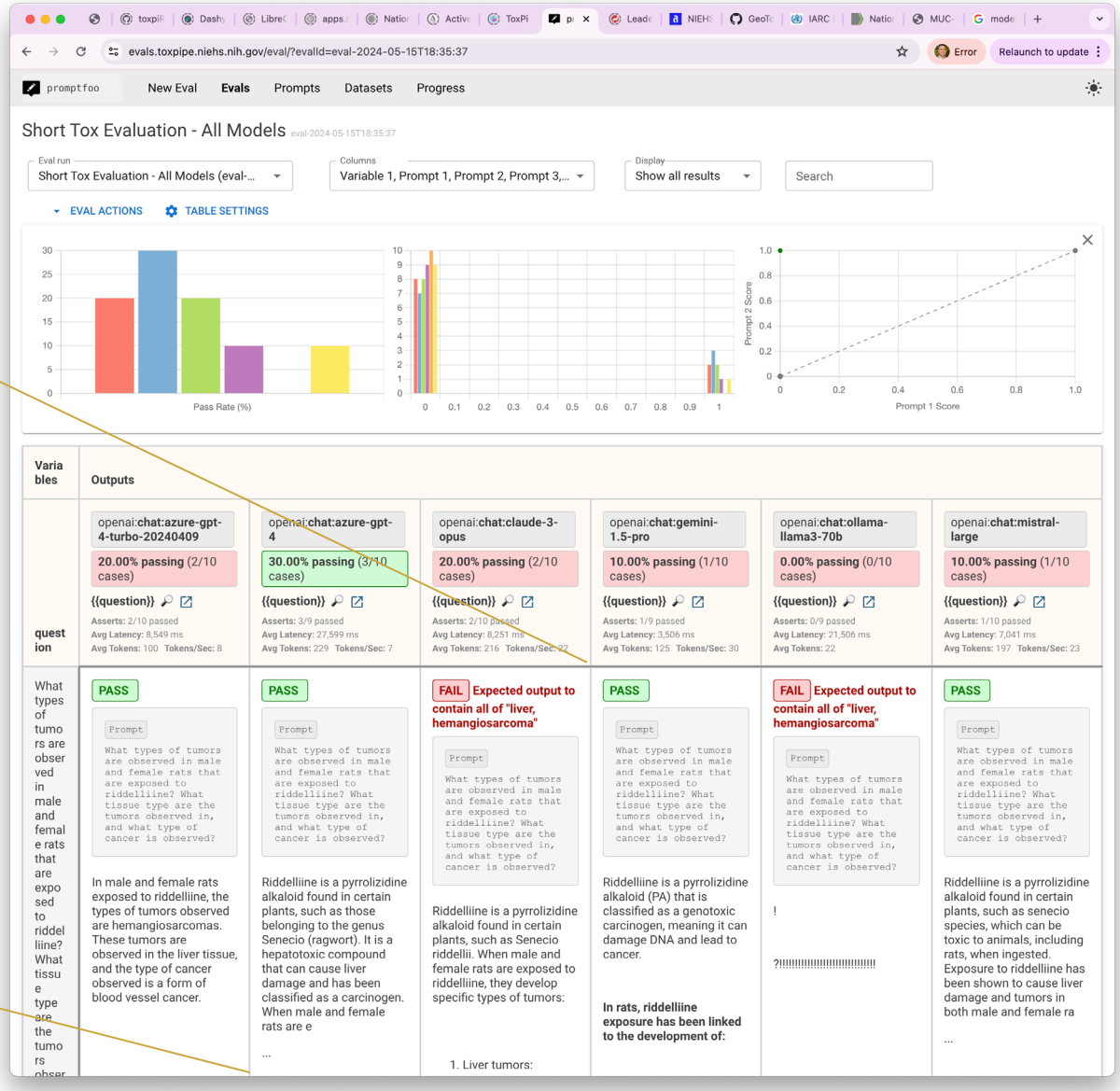
IVIVE: In vitro to In Vivo Extrapolation

National Institute of Environmental Health Sciences
Division of Translational Toxicology

Formal testing and model evaluation: Does generative AI give meaningful answers?

Generic or *one-size-fits-all* AI chat solutions FAIL at specific translational tests → expert guidance, training and domain-tailored RAG solutions will be key



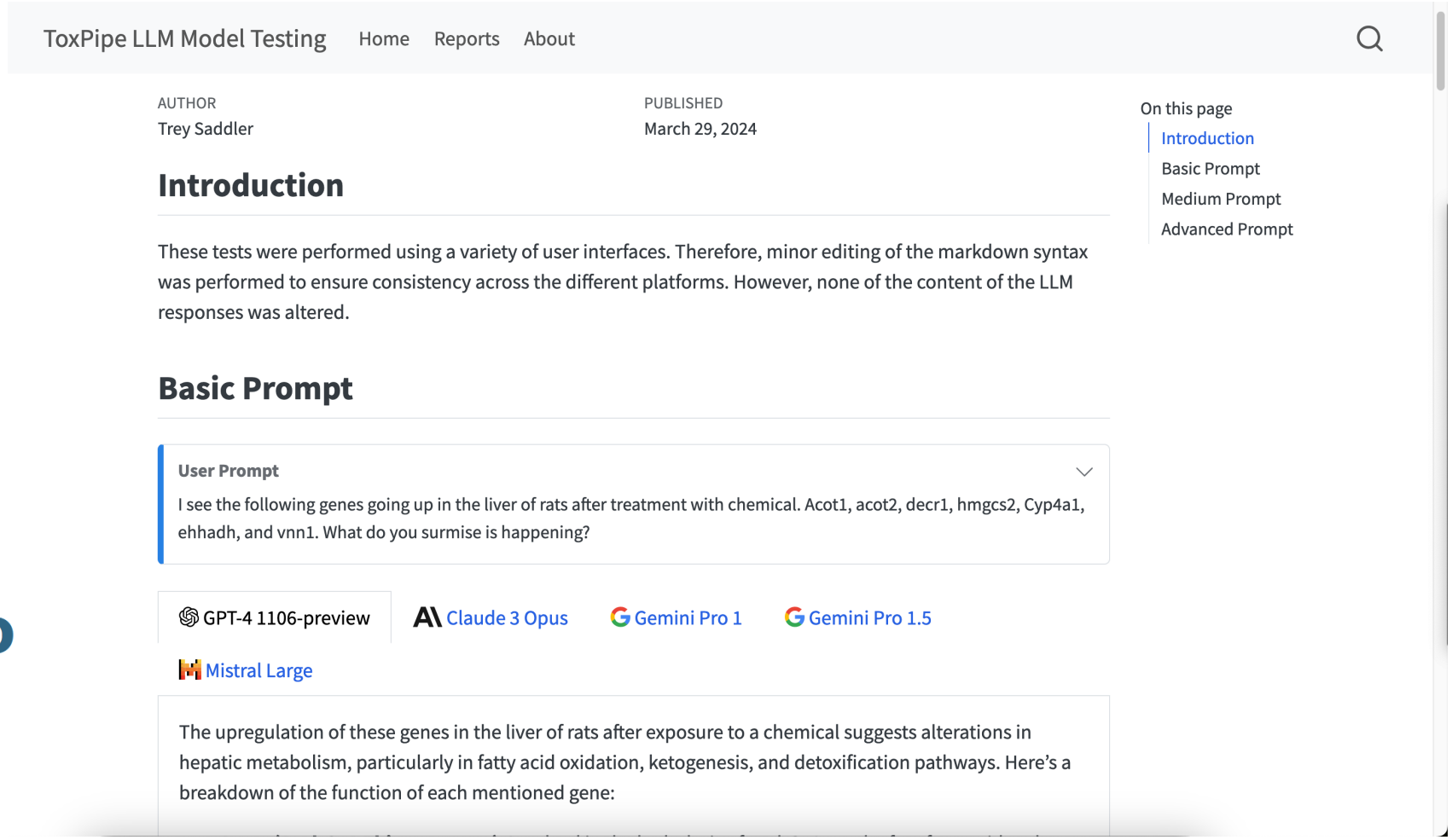


ToxPipe

Trey Saddler, Parker Combs, Jeremy Erickson, Charles Schmitt, David Reif, **Scott Auerbach**

ToxPipe LLM Comparison

promptfoo



- https://niehs.github.io/ToxPipe-Model-Comparison/reports/2024-03-29-initial-prompts.html

Agents and Assistants

- Focused on natural language interactions and conversations with users
- Typically respond to specific queries or commands from users
- Often have a more limited scope of capabilities centered around assisting with tasks like scheduling, answering questions, etc
- May not have as much autonomy or ability to take independent actions

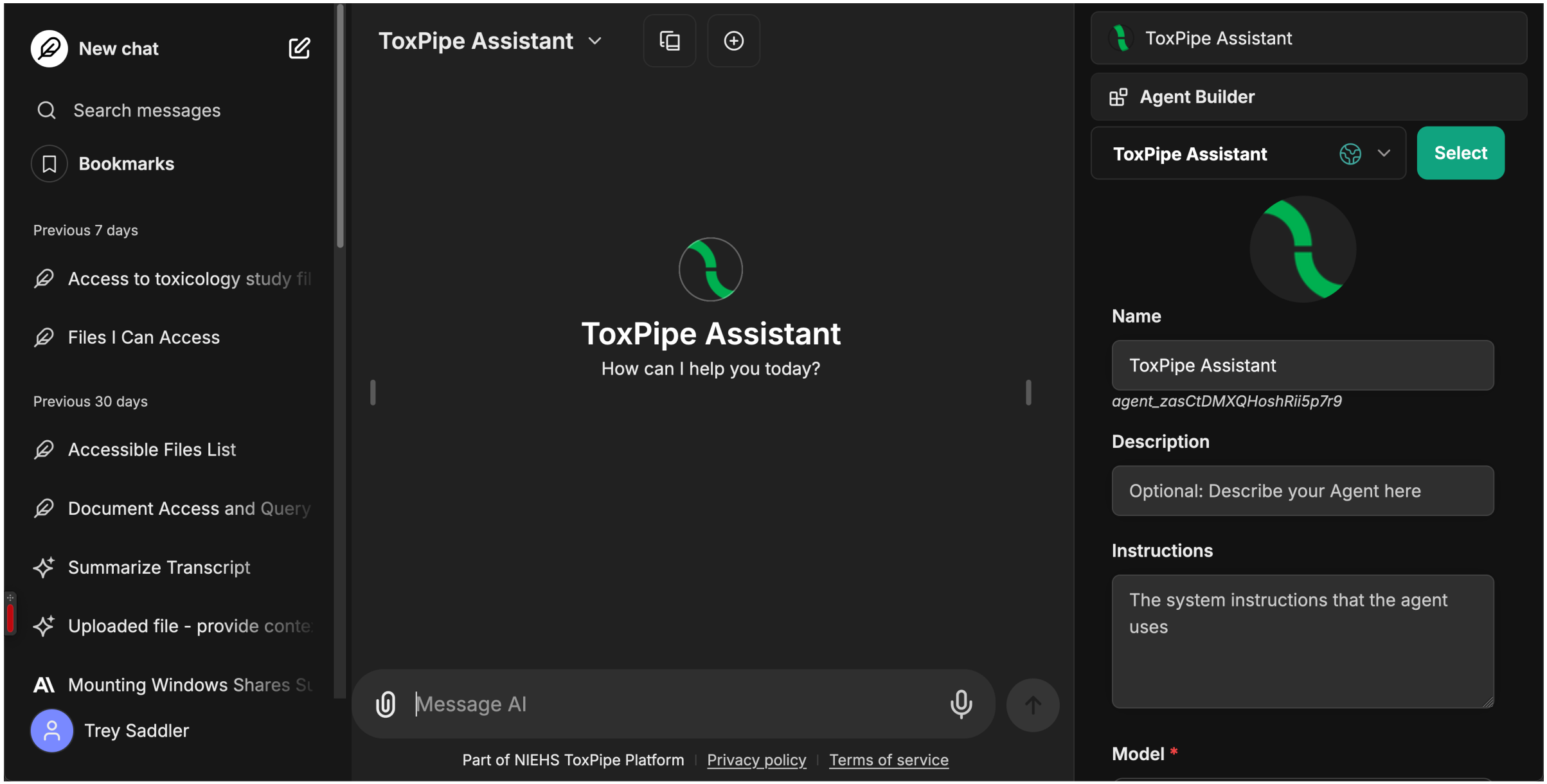
Assistant

- Can operate autonomously without constant human input or supervision
- Designed to take independent actions to achieve goals
- Might learn and improve their performance over time
- May use more advanced reasoning and decision-making capabilities
- Can interact with other systems, tools, and agents to complete complex tasks

Agent

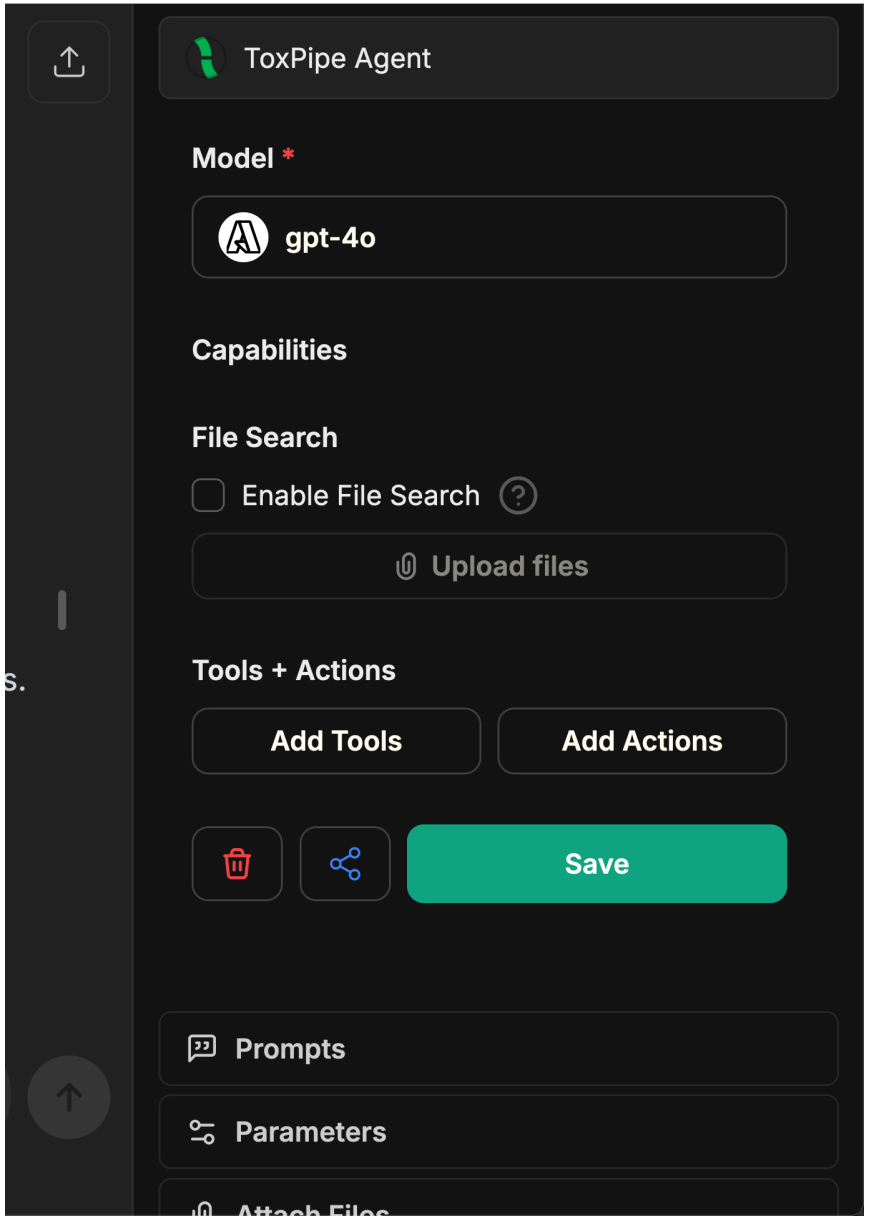
National Institute of Environmental Health Sciences
Division of Translational Toxicology

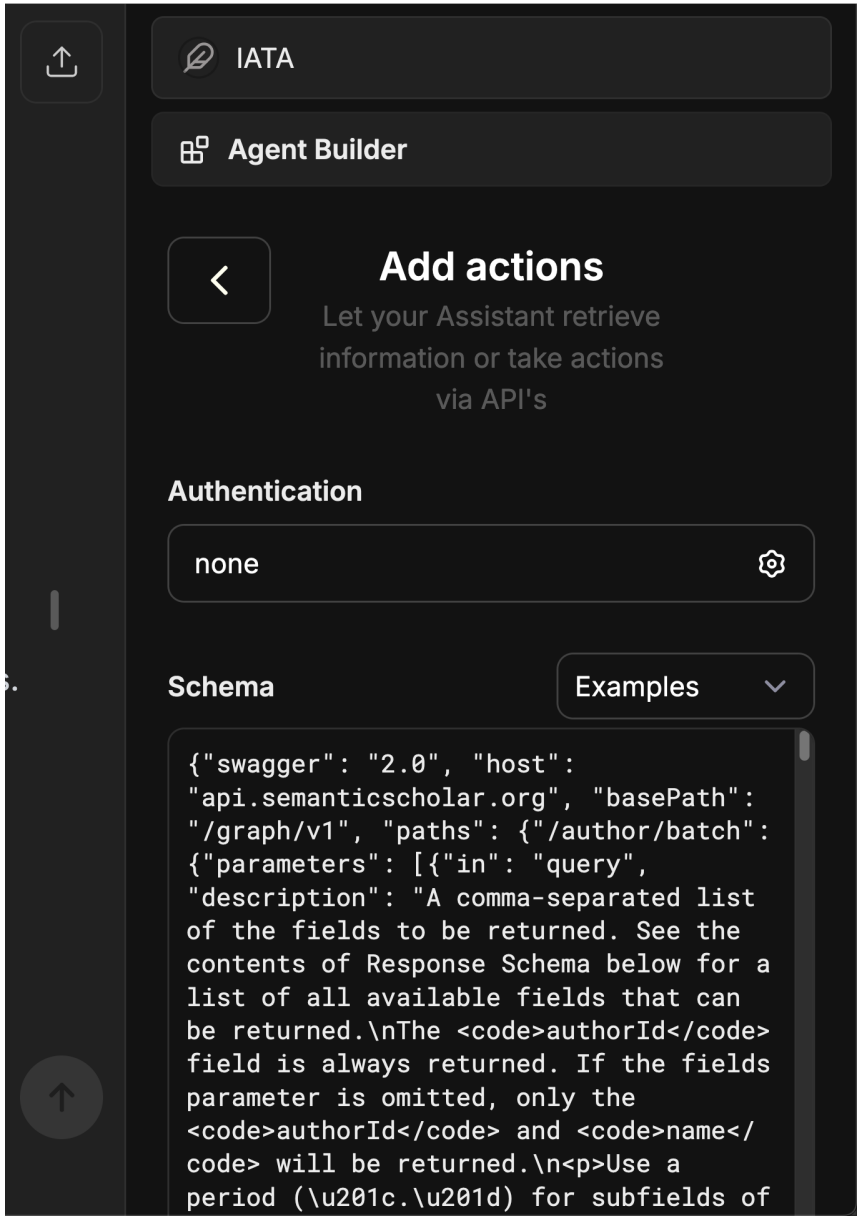
ToxPipe Agent Interface



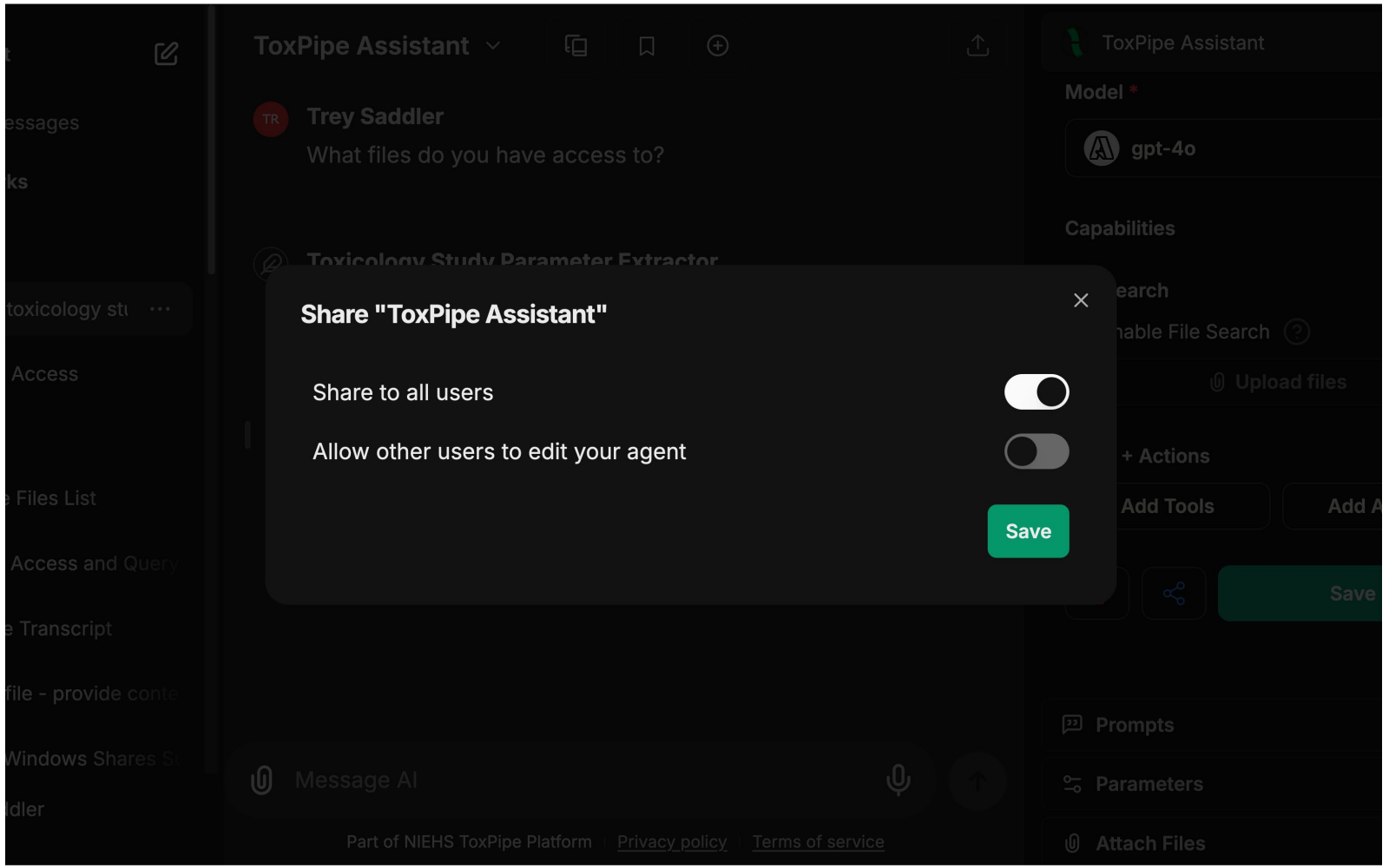
National Institute of Environmental Health Sciences
Division of Translational Toxicology

ToxPipe Agent Interface



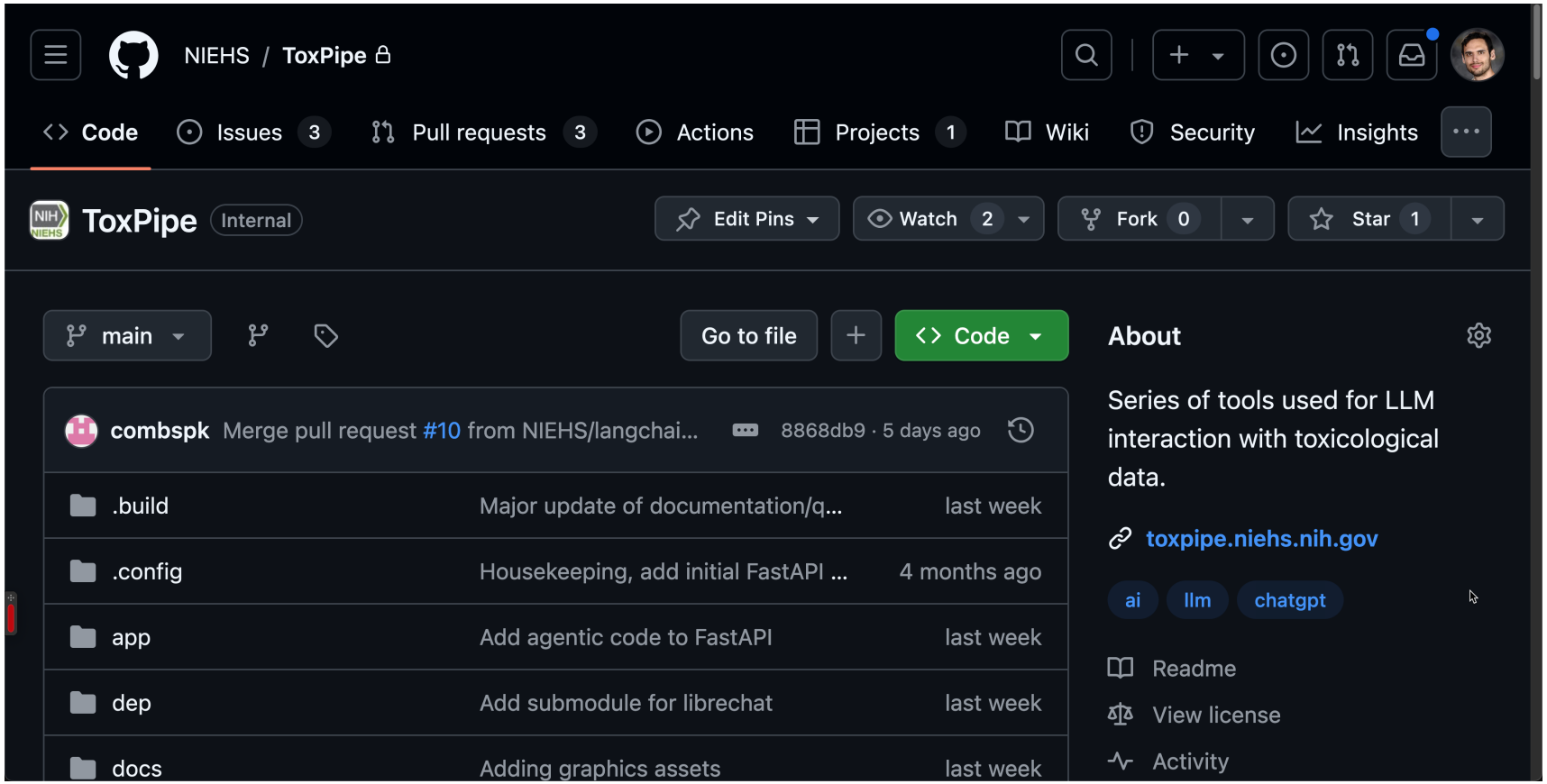


Global Agents



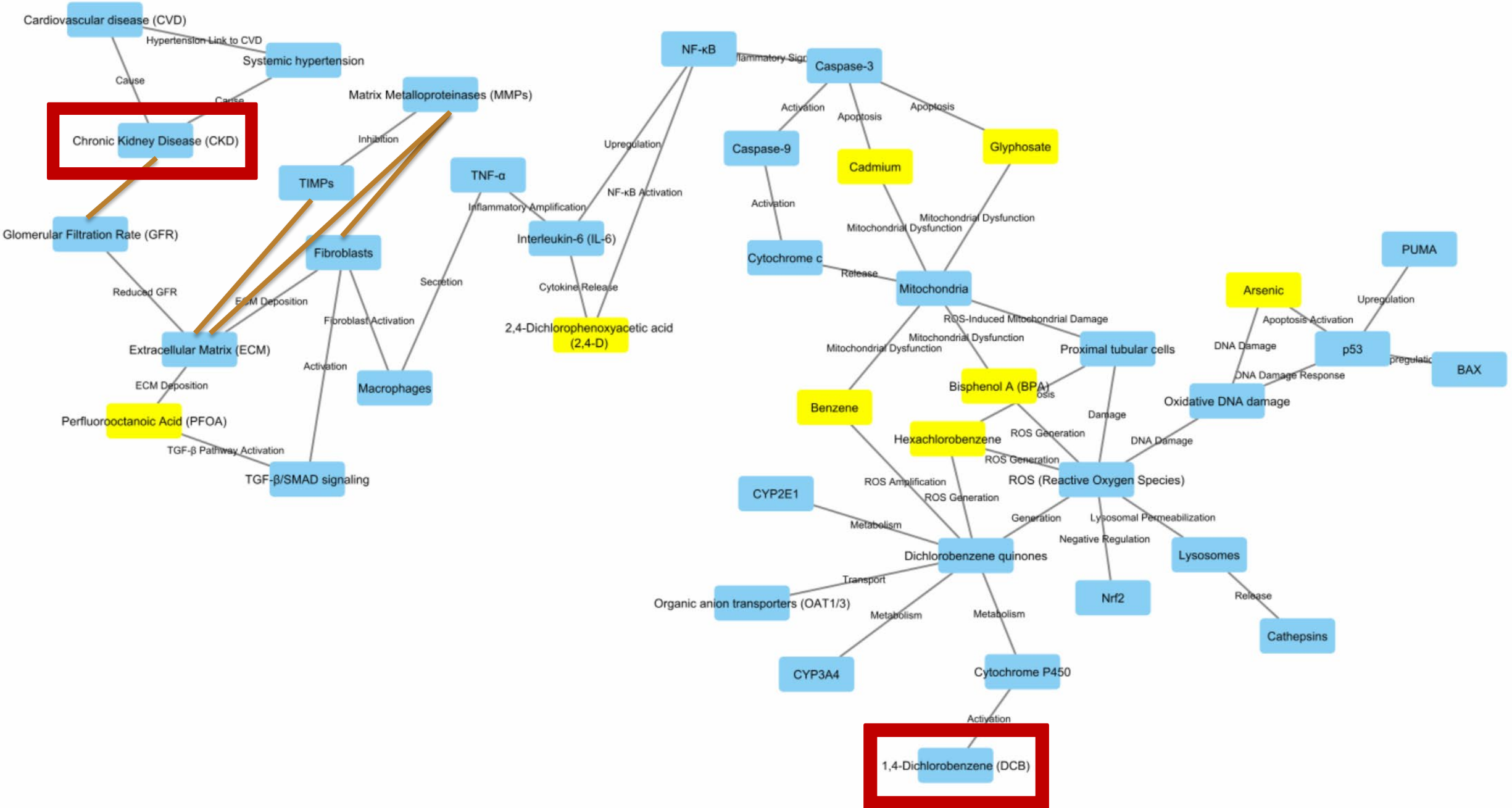
National Institute of Environmental Health Sciences
Division of Translational Toxicology

How can I use ToxPipe?



docker

AOP to Knowledge Graph: Dichlorobenzene → Chronic Kidney Disease



Knowledge Mapping and Mixture Interaction

Agent: Exposure Disease Mapper
Input: 1,4-dichlorobenzene, Chronic Kidney Disease

↓

Exposure to Disease Narrative

↓

Agent: Chemical Interaction with Exposure to Disease Narrative
Input: Exposure to Disease Narrative

↓

Multichemical Interaction Narrative

↓

Agent: Disease Map to Knowledge Graph
Input: Exposure to Disease Narrative
+
Multichemical Interaction Narrative

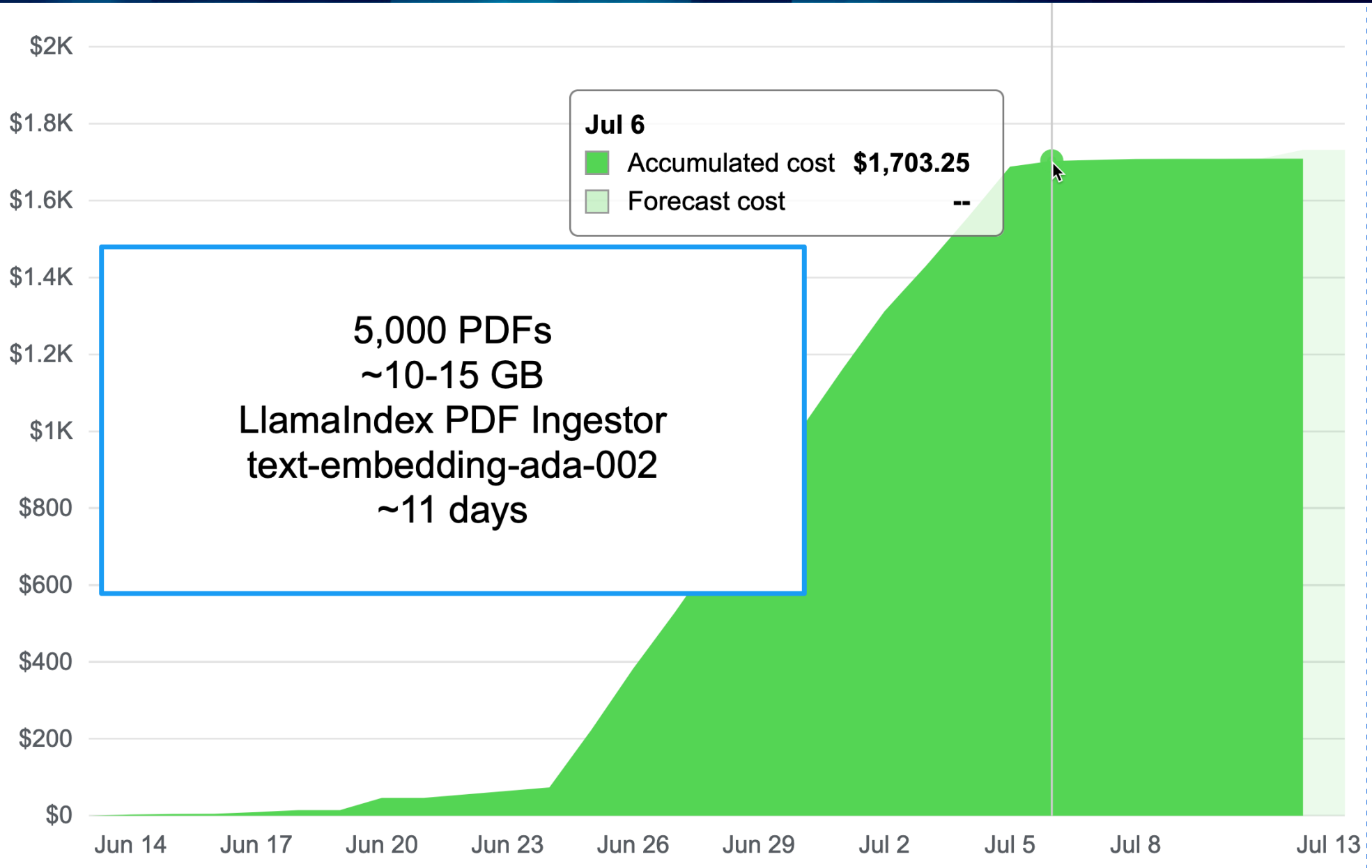
↓

Cytoscape Files

PDF Information Extraction

- OCRmyPDF
  - Tesseract
- Marker
  - Surya
- PaddleOCR
- Cloud Options ($$$)
  - Azure
  - Amazon Textract
  - Google Cloud Platform

Cost of embedding ~5k PDFs



Questions?