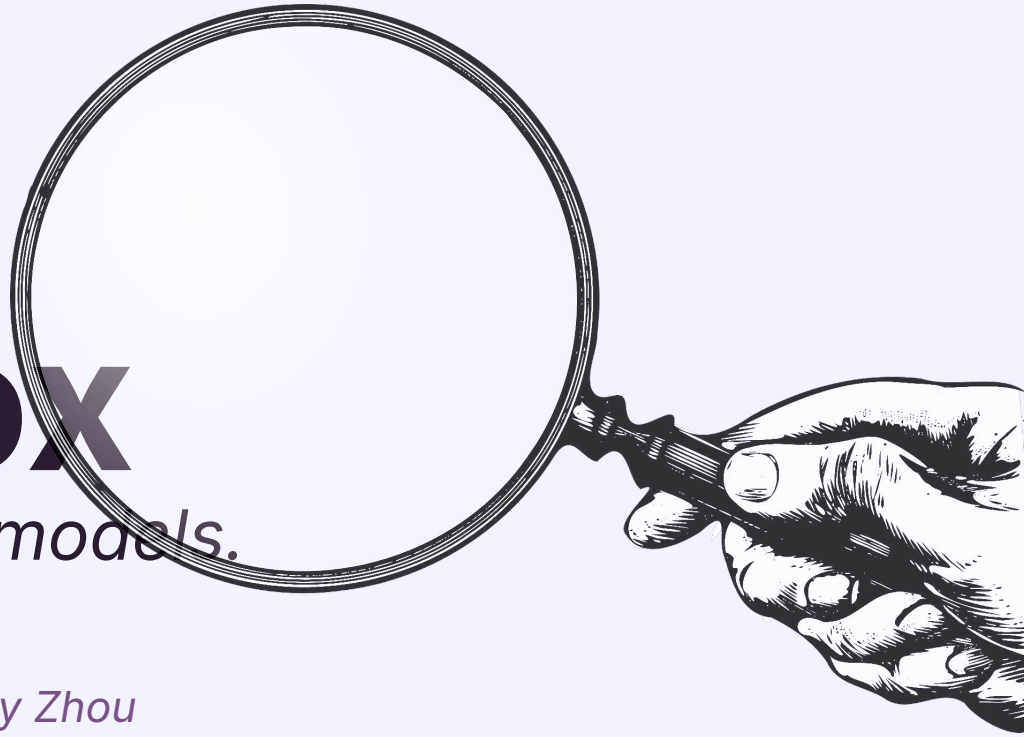
DataDetox

Know what's inside your models.



Kushal Chattopadhyay, Keyu Wang, Terry Zhou

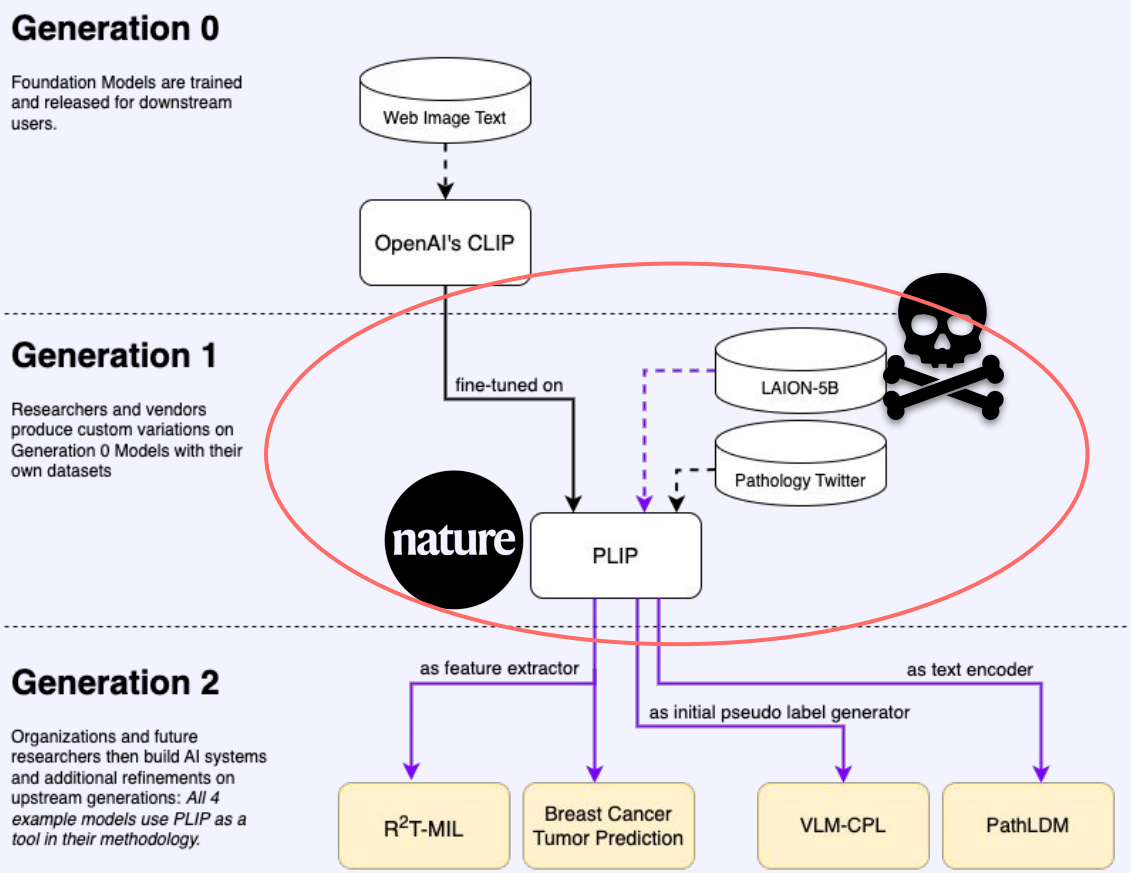
The Problem

The Problem

Increasing # of public models *(ex. Hugging Face...)*
building upon an existing model becomes extremely accessible...

It leads to expanding accessibility...

in Medical Imaging Applications



Who Can Benefit from *DataDetox*

All ML practitioners, including researchers, AI startups,
to avoid risks including...

01
Legal Risk

Mitigate privacy violations, copyright infringement, and non-compliance with GDPR...

02
Ethical Risk

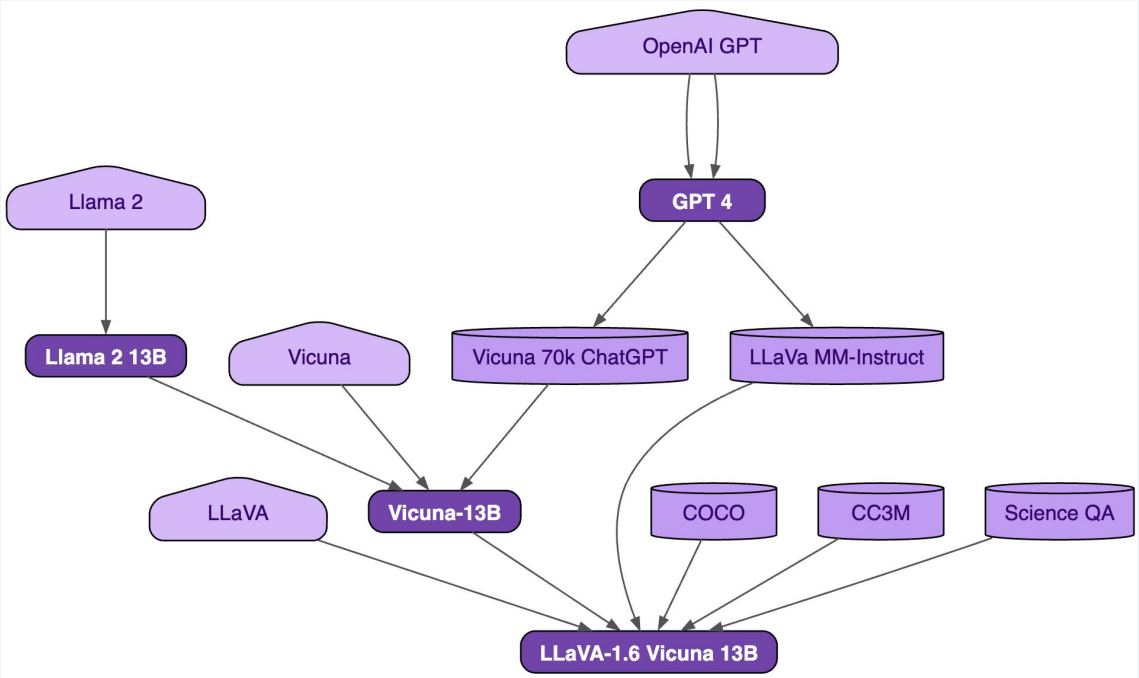
Identify and prevent issues like bias, exposure to CSAM, and non-consensual content...

03
Technical Risk

Address general under-documentation and hidden dependencies in foundation models

The Problem

"ML practitioners are not rewarded for doing deep-dive on all their data or dependencies."



Chat with DataDetox AI

Hello! I'm DataDetox AI. Ask me about model lineages, training datasets, or potential risks in AI models.
11:48 AM

Ask about model lineage, datasets, or risks...

Example Queries

- Which models were trained on LAION-5B?
- Show me the lineage of LLaVA v1.6
- Are there any risks with using Vicuna-13B and PLIP?
- What datasets does GPT-OSS-120B use?

Model Lineage Tree

- nvidia/gpt-oss-120b-Eagle3
  - openai/gpt-oss-120b
    - base-model-v1 Risk
    - base-model-v2
  - fine-tuned-model-1
    - derivative-a
    - derivative-b Risk

Connected Models Risk Detected

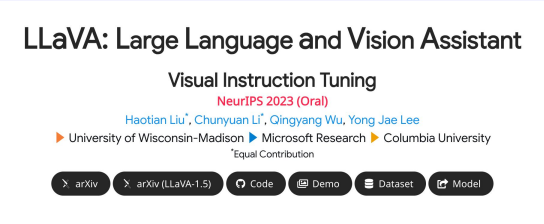
Technical Architecture

Tell me about the LlaVa v1.6, 34B model.

***DataDetox* Agents**

LlaVa-1.6 Vicuna 13B is a child of *LlaVa*. LlaVa is a ...

From the LlaVa-1.6 Vicuna 13B project, we see that they use pre-trained data from GPT, ...

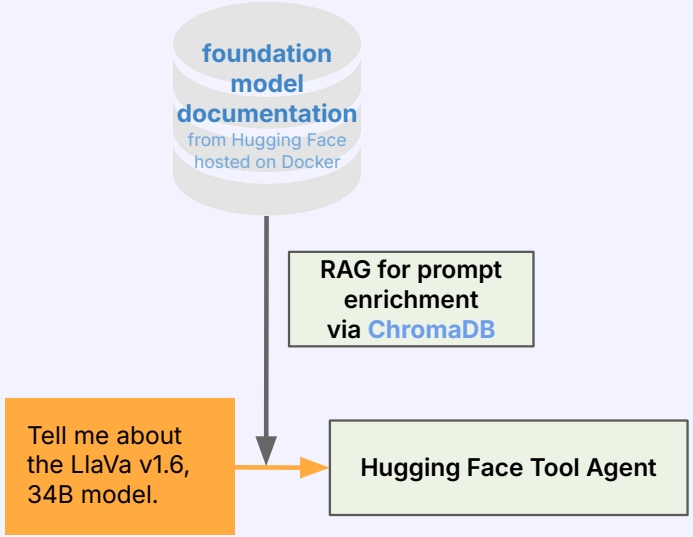


Research Agent: scalable RAG for input context

Querying GeMMa is easy, but *GeMMa-e4b,* or *GeMMa-3n-it*? **Share info on model family!**

Solution: *ChromaDB* for base model families

- 🧪 Cloud-hosted
- 🧪 Reduces duplicate search
- 🧪 Augments contextual intelligence on model
- 🧪 Can scale to more fine-grained versions



Research Agent: deep analysis

MCP to HF → reliable analysis

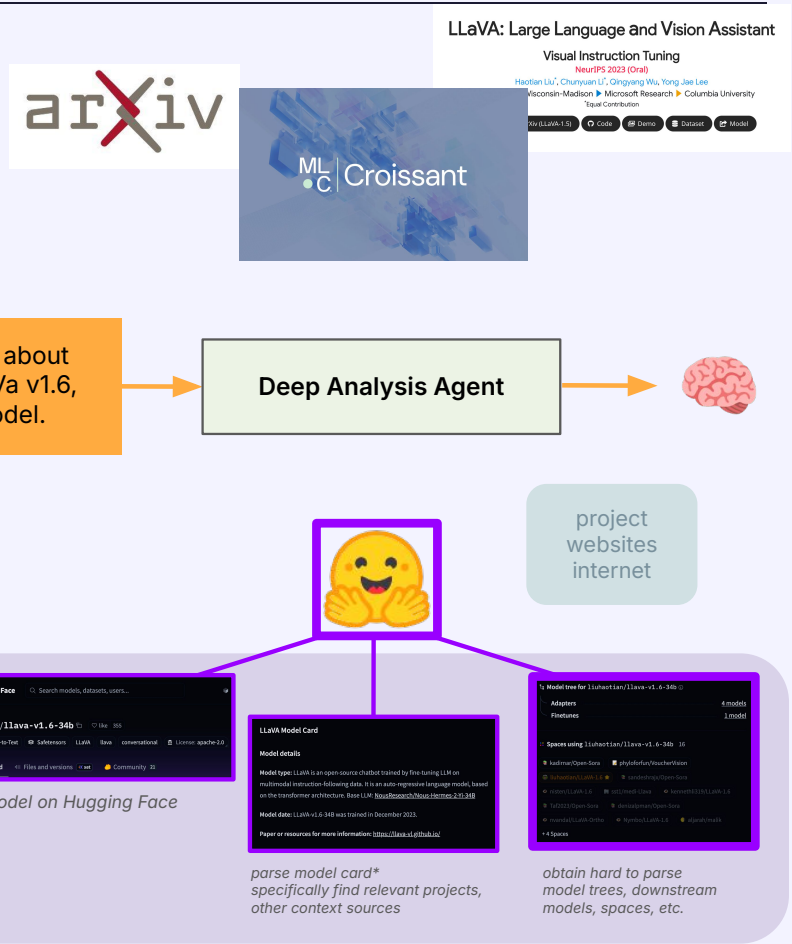
🧪 Deep integration with HF for increased scalability & longevity

🧪 Informed research from RAG context

🧪 Reliable utilization of project page, Arxiv papers, downstream models, metadata, etc

Modularity: can duplicate this agent for related models → compositional framework!

Efficiency: knowledge cache lightweight!

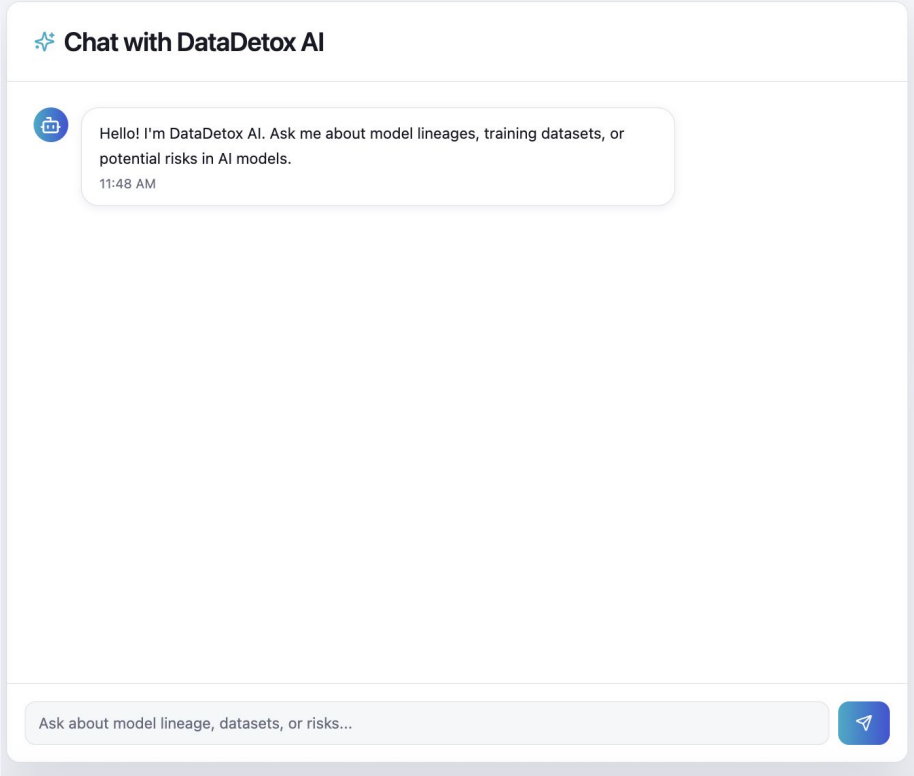


Advice Agent: model trees, risks, optimal fine-tuning,...

Consolidating knowledge from

→ foundation models' releases
→ Arxiv papers
→ project pages
→ Hugging Face datasets
→ fine tuning models
→ etc

DataDetox MAS offers *reliable risk analysis, research, and decision support* across the open-source AI landscape.



Tech Stack

Chroma

Hugging Face

FastAPI

Future Development and Growth Potential

- ➔ Graph visualizations of data trees, model streams, and potential risks for enhanced UX
- ➔ Closer collaboration with Data Provenance Initiative, ML Commons and Hugging Face to enhance our detection of high-risk datasets

Hugging Face