

# What if A.I. just works

Resilience and open-source



# AI SAFETY SUMMIT

HOSTED BY THE UK | 1-2 NOVEMBER 2023



FUTURE PERFECT

## Why can't anyone agree on how dangerous AI will be?

Researchers tried to get AI optimists and pessimists on the same page. It didn't quite work.

by **Dylan Matthews**

Mar 13, 2024, 11:30 AM GMT+1



## Letters

# The danger of blindly embracing AI

## Dangers of AI: Exploring the risks and threats



## 14 Risks and Dangers of Artificial Intelligence (AI)

AI has been hailed as revolutionary and world-changing, but it's not without drawbacks.



Written by **Mike Thomas**

The New York Times

SUBSCRIBE FOR €0.50/WEEK

gence > | OpenAI's \$157 Billion Valuation | Testing Apple Intelligence | Nuclear Power and A.I. | Can Math Help Chatbots? | A.I. in the Presidency

## What Exactly Are the Dangers Posed by A.I.?

A recent letter call blends real threat among experts.

The dangers and serious risks of AI

The world of artificial intelligence is all the rage, with more and more news in the media.



Key ta



• Artific



What if AI just works



## No internet

Try:

- Checking the network cables, modem, and router
- Reconnecting to Wi-Fi

ERR\_INTERNET\_DISCONNECTED



# Abandoned

## The human cost of neurotechnology failure

When the makers of electronic implants abandon their projects, people who rely on the devices have everything to lose.

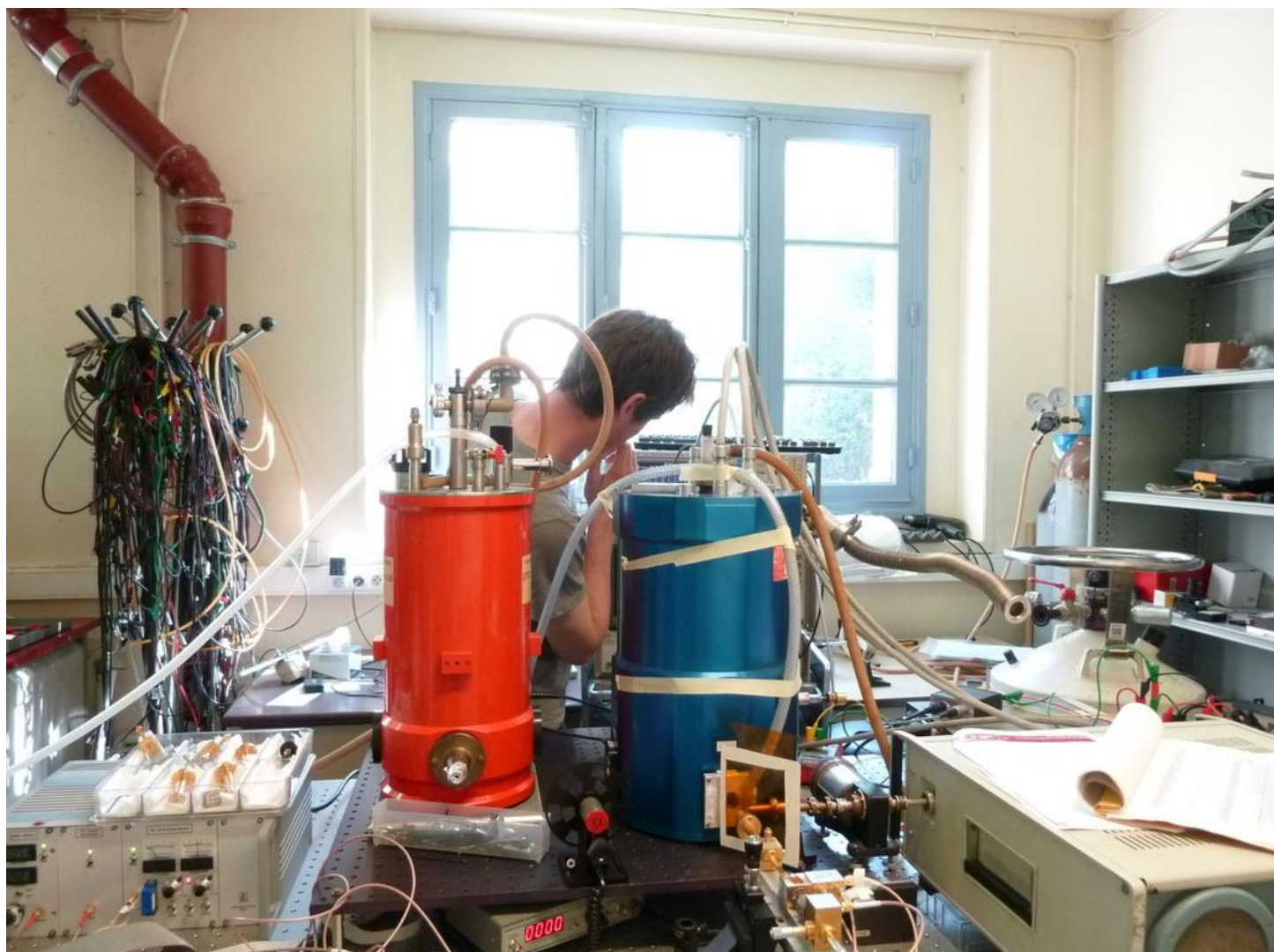
By Liam Drew | 6 December 2022

- Nice to have technology



- Technology which comes to mediate all our interactions with the world









## Open-Source Operating System



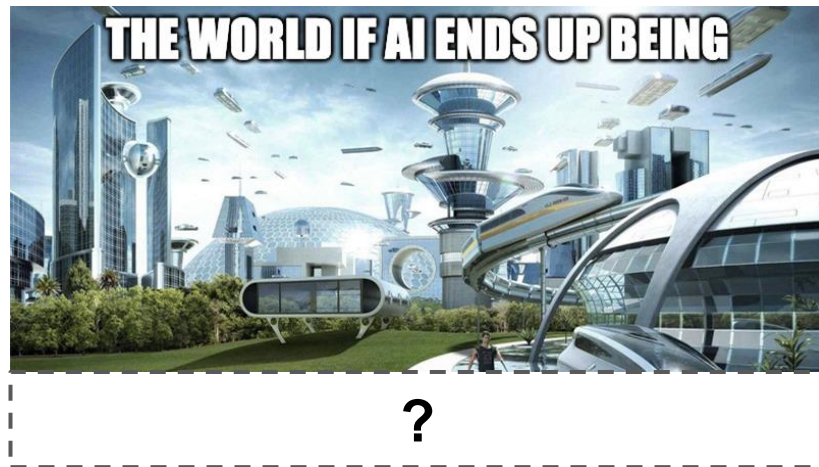
# The case of “Open-Source” AI?



- **Security:** Better security through obscurity?
- **Risks:** Avoid misuse, rogue actors?
- **AGI:** No rogue models?
- **Business model:** Paying for AI training?



- **Accessibility/no centralization:** Anyone can access/own AI
- **Flexibility/freedom of choice:** Various hardware, models
- **Pooling resources/knowledge:** Fostering innovation, sharing
- **Privacy/IP protection:** Owning data (training/inference).
- **Transparency/trust:** Knowing/exploring biases, issues

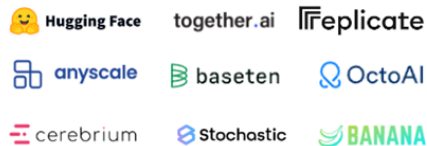


# Open-source AI mapping

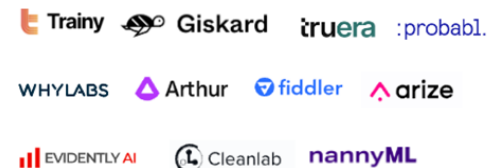
## Generative AI - LLM developers



## LLM hosting & deployment



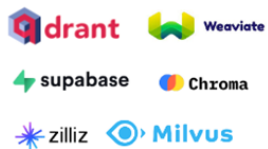
## AI Observability & monitoring



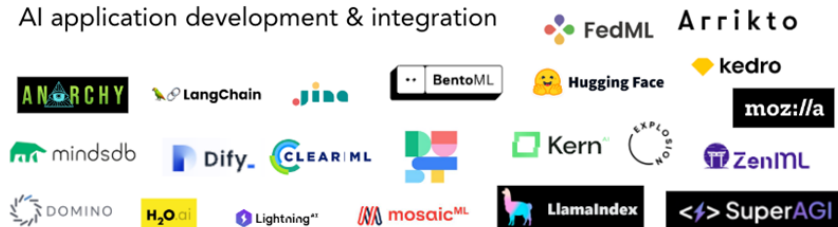
## Generative AI - Image model developers



## Vector databases



## AI application development & integration



## Privacy, governance & risk management



## Platforms to train open source vision AI algorithms



## Synthetic training data



## Bonus : Most popular AI frameworks

