Score an AI company with a public safety index

Intermediate · A few weeks · Free · Project idea

Online version: https://ai-safety-sandbox.vercel.app/projects/score-a-lab-with-the-safety-index/

Summary

The Future of Life Institute's AI Safety Index grades leading companies on public evidence. Apply its indicators to a company it does not cover, or re-score one with newer evidence.

Why it matters

Independent scorecards create reputational pressure. Extending them, transparently, is useful and teaches you what good practice looks like.

What you will have at the end

A scorecard with sources for every grade.

How to start (tick each box as you go)

- [ ] **1. Write your question in one sentence.**
- [ ] **2. Spend 30 minutes finding what has already been written about it.**
- [ ] **3. Collect evidence in a simple table: claim, source, how strong.**
- [ ] **4. Draft 1–2 pages: answer first, then reasons, then uncertainties.**
- [ ] **5. Ask one person to poke holes in it, revise, and share.**

Key terms

Frontier safety framework
A company's published plan for which dangerous capabilities it tests for and what it promises to do if a model crosses a risk threshold.

Source and links

Future of Life Institute — AI Safety Index
https://futureoflife.org/index

AI Lab Watch
https://ailabwatch.org/

When you have finished

Write a few paragraphs on what you did, what you found and what you would try next, and share it with a friend, your society or online.

Courses, mentors, small grants and communities:

https://ai-safety-sandbox.vercel.app/next-steps/

Questions? Ask in AI Safety Cardiff University (AISCU): https://aiscu.org/

Notes