AI SAFETY SANDBOX · AI POLICY

Who pays when AI causes harm? A liability explainer

Intermediate · A weekend · Free · Project idea

Online version: https://ai-safety-sandbox.vercel.app/projects/ai-liability-explainer/

Summary

Explain how liability and insurance could shape AI companies' behaviour, using one realistic harm scenario to make it concrete.

Why it matters

If developers pay for the harms they cause, safety becomes a business priority. Few people outside law can explain how this would actually work.

What you will have at the end

A 1,500-word explainer.

How to start (tick each box as you go)

- [ ] **1. Pick a specific reader, such as a friend who is not into tech.**
- [ ] **2. Read 2–3 good sources and write the core idea in one sentence.**
- [ ] **3. Draft with one everyday example.**
- [ ] **4. Define every technical term, or remove it.**
- [ ] **5. Test it on a real reader, revise, and publish.**

Source and links

MINT Lab — AGI Governance Bibliography (Seth Lazar, ANU)
https://bibliography.mintresearch.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