Compare whistleblower protections for AI workers

Intermediate · A few weeks · Free · Project idea

Online version: https://ai-safety-sandbox.vercel.app/projects/whistleblower-protections-compared/

Summary

If an employee at an AI company sees something dangerous, what protects them if they speak up? Compare protections across two or three jurisdictions.

Why it matters

Insiders are often the first to see problems. The MINT bibliography treats incident reporting and whistleblowing as a core governance topic.

What you will have at the end

A comparison and short recommendations.

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.**

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