Investigate a real AI incident

Starter · 1–3 hours · Free · Step-by-step guide

Online version: https://ai-safety-sandbox.vercel.app/projects/investigate-an-ai-incident/

Summary

Pick one case from the AI Incident Database and write it up like an accident investigator: what happened, who was harmed, why, and what rule or practice would have prevented it.

Why it matters

Learning from incidents is how aviation and medicine became safe. AI policy needs the same habit, and each well-documented case helps.

What you will have at the end

A one-page incident report.

Steps (tick each box as you go)

- [ ] **1. Choose an incident (20 min)**
  Browse the AI Incident Database and pick a case with several news reports, so you can cross-check facts.
- [ ] **2. Build the timeline (30 min)**
  List what happened, when, and who was involved. Mark any facts that sources disagree on.
- [ ] **3. Find the causes (30 min)**
  Separate the technical cause, the human and organisational causes, and missing safeguards.
- [ ] **4. Check the rules (20 min)**
  Did any law or standard apply? Would the EU AI Act or your country's rules have changed anything?
- [ ] **5. Recommend one fix (20 min)**
  Propose one realistic change (technical, organisational or legal) and explain why it would help.

What you need

- A web browser
- About two hours

What you will learn

- How AI failures happen in practice
- Root-cause thinking
- Connecting incidents to policy

Key terms

Red-teaming
Deliberately trying to make a system fail or misbehave, so weaknesses are found and fixed before bad actors find them.

Source and links

AI Incident Database
https://incidentdatabase.ai/

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.

Try next: Sort 20 AI news stories with the MIT AI Risk Repository
https://ai-safety-sandbox.vercel.app/projects/sort-risks-with-the-mit-repository/

Try next: Compare whistleblower protections for AI workers
https://ai-safety-sandbox.vercel.app/projects/whistleblower-protections-compared/

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