Make a safety paper’s code easy to re-run

Intermediate · A weekend · Free · Project idea

Online version: https://ai-safety-sandbox.vercel.app/projects/make-research-reproducible/

Summary

Pick a safety paper with public code and make it reproducible: fix broken dependencies, write setup docs, or package it so others can run it in one command.

Why it matters

A lot of research code stops working within a year. Reproducible code lets others check and extend results, and you’ll understand the paper well by the end.

What you will have at the end

A pull request or a clean fork with a working “quick start”.

How to start (tick each box as you go)

- [ ] **1. Find a paper you care about whose code is public but hard to run (check its GitHub issues).**
- [ ] **2. Try to run it from scratch and write down every point where you get stuck.**
- [ ] **3. Fix the blockers one at a time: dependencies, paths, missing data, unclear steps.**
- [ ] **4. Write a short quick-start and open a pull request or issue with your fixes.**
- [ ] **5. Share a short note on what you fixed.**

Source and links

BlueDot — The software engineer’s guide to your first AI safety contribution
https://blog.bluedot.org/p/swe-ai-safety-project-guide

Example micro-tool: batch-querying LLMs
https://github.com/safety-research/safety-examples/blob/main/examples/inference/get_responses.py

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