What does the Baruch Plan teach compute governance?

Intermediate · A few weeks · Free · Project idea

Online version: https://ai-safety-sandbox.vercel.app/projects/baruch-plan-and-compute/

Summary

Pick three lessons Zaidi and Dafoe draw from the failed attempt at international control of nuclear weapons. For each, argue whether it transfers to AI compute: chips are easier to track than ideas but harder than uranium.

Why it matters

History is one of the few sources of evidence about controlling a powerful technology internationally.

What you will have at the end

A policy-style memo that also takes a position on Bostrom's preventive governance proposals.

How to start (tick each box as you go)

- [ ] **1. Pick a decision-maker and a decision they actually face.**
- [ ] **2. Gather the facts: what exists, what is proposed, who disagrees.**
- [ ] **3. Lay out 2–3 realistic options with their pros and cons.**
- [ ] **4. Recommend one, and say what evidence would change your mind.**
- [ ] **5. Keep it to two pages and get feedback from someone who disagrees.**

Key terms

Compute governance
Using control over AI chips and data centres (who can buy, use or track them) as a lever for governing AI.

Verification
Ways to check, rather than just trust, that someone is following an agreement (for example, confirming a data centre is not secretly training a banned model).

Source and links

AI Safety × Philosophy project ideas (P21) — builds on Zaidi and Dafoe, International Control of Powerful Technology; Bostrom, The Vulnerable World Hypothesis

Bostrom — The Vulnerable World Hypothesis
https://nickbostrom.com/papers/vulnerable.pdf

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