Re-check “catch-up algorithmic progress”

Advanced · A few weeks · Free · Project idea

Online version: https://ai-safety-sandbox.vercel.app/projects/catch-up-algorithmic-progress/

Summary

MIRI TGT argued that catching up to a given capability level gets cheaper by roughly 60× per year. Re-run the analysis with newer models and test how robust that number is.

Why it matters

The MIRI team describes this post as roughly the minimum standard of rigour they expect. The figure also affects how quickly export controls lose their effect.

What you will have at the end

A replication note with updated estimates.

How to start (tick each box as you go)

- [ ] **1. Read the original work and write down the exact claim you will test.**
- [ ] **2. Get the code or setup running on the smallest possible version.**
- [ ] **3. Reproduce one key result before changing anything.**
- [ ] **4. Change one thing — that is your extension — and compare carefully.**
- [ ] **5. Write up what matched, what did not, and why you think so.**

Key terms

Replication
Re-running someone else’s experiment to check you get the same result. It’s a common way to start research, especially if you then extend it.

Compute
The raw computing power (specialised chips running for a long time) used to train and run AI models. Easier to measure and track than data or ideas.

Source and links

Technical AI Governance resource pack (MIRI TGT reading lists)
https://www.lesswrong.com/posts/GRPswd5v7haob97fA/recommendations-for-people-getting-into-technical-ai

Catch-up algorithmic progress might actually be 60× per year
https://techgov.intelligence.org/blog/catch-up-algorithmic-progress-might-actually-be-60x-per-year

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