Estimate how much compute a famous model needed

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

Online version: https://ai-safety-sandbox.vercel.app/projects/fermi-estimate-training-compute/

Summary

Use the rule of thumb “training compute ~ 6 × parameters × training tokens” to estimate the compute behind a model, then compare your answer with Epoch AI’s database.

Why it matters

Compute thresholds appear in real regulation. Being able to estimate compute, and know how uncertain the estimate is, is a core governance skill.

What you will have at the end

A one-page estimate with your assumptions laid out.

Steps (tick each box as you go)

- [ ] **1. Pick a model (10 min)**
  Choose a model whose parameter count and training tokens have been published (many open models report both).
- [ ] **2. Do the rough estimate (20 min)**
  Multiply 6 × parameters × tokens to get training FLOP. Write the answer in scientific notation, such as 3 × 10²^4.
- [ ] **3. Check against Epoch AI (20 min)**
  Look the model up in Epoch AI’s data. How close were you? If you are off, find out why.
- [ ] **4. Turn it into time and money (30 min)**
  Estimate how long that would take on a cluster of chips and roughly what it would cost. Say which assumption you are least sure about.
- [ ] **5. Explain why it matters (20 min)**
  Find one law or policy that uses a compute threshold. Is your model above or below it?

What you need

- A calculator or spreadsheet
- Comfort with powers of ten

What you will learn

- How training compute is estimated
- Fermi estimation
- Why compute is a governance lever

Key terms

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.

Fermi estimate
A quick, rough calculation that breaks a hard number into easier guesses multiplied together. Good enough to tell whether something is a big deal.

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

Source and links

Epoch AI — data on notable models and compute
https://epoch.ai/data

Example of a great first-pass estimate: could decentralised training solve AI's power problem?
https://epoch.ai/blog/could-decentralized-training-solve-ais-power-problem

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: How much compute did the key AI innovations need?
https://ai-safety-sandbox.vercel.app/projects/compute-behind-key-innovations/

Try next: Are consumer GPUs a problem for export controls?
https://ai-safety-sandbox.vercel.app/projects/consumer-gpus-and-export-controls/

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