Did they keep their promise? Check the Seoul commitments

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

Online version: https://ai-safety-sandbox.vercel.app/projects/seoul-commitments-check/

Summary

At the 2024 AI Seoul Summit, frontier AI companies signed voluntary safety commitments, including publishing a safety framework. Pick five signatories and check what they actually published.

Why it matters

Voluntary commitments only matter if someone checks them, and simple, well-sourced tracking is one way to do that.

What you will have at the end

A small tracker: company, commitment, evidence, date, status.

Steps (tick each box as you go)

- [ ] **1. Read the commitments (20 min)**
  Open the Frontier AI Safety Commitments page on GOV.UK. List the specific things companies agreed to do.
- [ ] **2. Choose five companies (10 min)**
  Mix large and smaller signatories so the picture is fair.
- [ ] **3. Hunt for the evidence (1 hr)**
  For each company, find its published framework or statement. Record the link and date, or “not found”.
- [ ] **4. Rate each one (20 min)**
  Use three simple labels: done, partly done, not found. Write one sentence justifying each.
- [ ] **5. Share it carefully (20 min)**
  Note your search method and date so others can update it. Share it with your group.

What you need

- A web browser
- A spreadsheet
- About two hours

What you will learn

- How voluntary AI commitments work
- Accountability through public evidence
- Careful, fair sourcing

Key terms

Frontier safety framework
A company's published plan for which dangerous capabilities it tests for and what it promises to do if a model crosses a risk threshold.

Frontier AI
The most capable AI models at any given time, built by a small number of companies with very large amounts of computing power.

Source and links

GOV.UK — Frontier AI Safety Commitments, AI Seoul Summit 2024
https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024

METR — Common Elements
https://metr.org/common-elements

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: Compare two AI companies' safety promises
https://ai-safety-sandbox.vercel.app/projects/compare-lab-safety-promises/

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