AI SAFETY SANDBOX · AI POLICY

Should AI companion apps be regulated? A short brief

Intermediate · A weekend · Free · Project idea

Online version: https://ai-safety-sandbox.vercel.app/projects/ai-companions-brief/

Summary

Millions of people talk to AI companions. Review the evidence on benefits and harms and propose proportionate safeguards.

Why it matters

Companion apps raise questions about vulnerable users, dependence and manipulation, and rules are only starting to appear.

What you will have at the end

A two-page brief.

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.**

Source and links

MINT Lab — AGI Governance Bibliography (Seth Lazar, ANU)
https://bibliography.mintresearch.org/

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