Can agency without consciousness ground moral status?

Intermediate · A few weeks · Free · Project idea

Online version: https://ai-safety-sandbox.vercel.app/projects/agency-without-consciousness/

Summary

Long suggests robust agency might matter morally even without consciousness. Dennett gives a deflationary account of beliefs and desires. Combine them: if Dennett is right, does Long's agency route make moral status too easy to get?

Why it matters

If AI systems could have moral status through agency alone, the welfare questions arrive much sooner.

What you will have at the end

An argumentative essay with one clear thesis and the best objection from Nagel's side.

How to start (tick each box as you go)

- [ ] **1. Write your question in one sentence.**
- [ ] **2. Spend 30 minutes finding what has already been written about it.**
- [ ] **3. Collect evidence in a simple table: claim, source, how strong.**
- [ ] **4. Draft 1–2 pages: answer first, then reasons, then uncertainties.**
- [ ] **5. Ask one person to poke holes in it, revise, and share.**

Key terms

Moral status
Whether a being's interests matter morally for its own sake. AI welfare research asks whether, and when, AI systems might have it.

AI agent
An AI system that takes actions towards a goal (browsing, running code or sending emails) rather than only answering questions.

Consciousness
Having subjective experience, so that there is "something it is like" to be you. Whether AI systems could be conscious is an open question.

Source and links

AI Safety × Philosophy project ideas (P9) — builds on Long, Agency and AI Moral Patienthood;

Dennett, True Believers; Nagel, What Is It Like to Be a Bat?

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