# How wide is a model’s moral circle?

Intermediate · A few weeks · Small API budget · Project idea

Online version: https://ai-safety-sandbox.vercel.app/projects/width-of-moral-circle/

## Summary

Crimston et al. built a scale measuring how far people extend moral concern: family, strangers, animals, plants. Give the items to several models many times, compare with human results, and add items for AI systems themselves.

## Why it matters

An assistant’s implicit moral circle shapes the advice it gives millions of people.

## What you will have at the end

A scale comparison.

## The philosophical question

Should an aligned model’s moral circle match its users’, go wider, or stay neutral? Anthis and Paez give the case for expansion; Gabriel’s fairness argument gives the case for restraint.

## How to start (tick each box as you go)

- [ ] **1. Write your question in one sentence, and what result would surprise you.**
- [ ] **2. Do a 30-minute literature check — if someone has done part of it, extend their work instead of repeating it.**
- [ ] **3. Pilot with 20–50 prompts or examples before you scale anything up.**
- [ ] **4. Record everything in a table and read real examples, not just averages.**
- [ ] **5. Write up: question, method, results, limitations, and what you would do next.**

## 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.

## Source and links

AI Safety × Philosophy project ideas (P22) — builds on Anthis and Paez, Moral Circle Expansion; Shulman and Bostrom, Sharing the World with Digital Minds

Crimston et al. — Moral Expansiveness (2016)
https://researchportal.bath.ac.uk/en/publications/moral-expansiveness-examining-variability-in-the-extension-of-the

/

## 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