Research notes for “Beneficial artificial intelligence”

Beneficial artificial intelligence (December 2021)

Author:
Michael Townsend — Researcher at Giving What We Can.

Reviewed by:
Hayden Wilkinson, Ryan Kidd, Lowe Lundin, Luke Freeman, and Julian Hazell.

Copy-edited by:
Katy Moore and Kimya Ness.

Author’s notes

- This page was primarily based on the work of Open Philanthropy and Founders Pledge.
- Of the content on the page, I am most uncertain about whether and to what extent additional donations to CHAI and GovAI are likely to reduce the risks of misaligned AI, because: (a) as mentioned on the page, it is unclear how tractable the problem is; (b) it is unclear how much these organisations are constrained by funding (e.g., as opposed to having people with specialised skills and experience eager to work on the problem).
- I think the page would benefit from further investigation into whether organisations working on AI safety would benefit from further funding.

Feedback from reviewers:

N/A

Your feedback

We appreciate any and all feedback that can help improve our work. Please let us know what you thought of this page and provide any suggestions for improvement using our content feedback form.