A conversation with Paul Christiano, May 2, 2015

Participants

- Paul Christiano – PhD student, Theory of Computing Group, University of California: Berkeley
- Dr. Nick Beckstead – Research Analyst, Open Philanthropy Project

Note: These notes were compiled by the Open Philanthropy Project and give an overview of the major points made by Mr. Christiano.

Summary

The Open Philanthropy Project spoke with Mr. Christiano as part of its investigation into risks of artificial intelligence (AI). The key points from the conversion were that:

1. Mr. Christiano believes that speeding up AI progress may make society less prepared for the transition to advanced AI, but that this risk is offset by other considerations (such as the possibility of using AI to address other existential risks), so he is highly uncertain of the overall effect on existential risk.
2. If a funder concerned about AI safety began funding AI work, including work that might speed up the development of very advanced AI, then the funder's entry into the field would probably reduce potential risks from AI. The funder's entry might push the field in a more safety-oriented direction.
3. Mr. Christiano would favor systemic improvements to the field (such as bringing in additional talented researchers). He believes it would generally increase the likelihood of success across projects in the field, including projects that reduce risks from unintended consequences of AI.

All Open Philanthropy Project conversations are available at
http://www.givewell.org/conversations