Strategic Research Center for Artificial Intelligence Policy

Project Summary
(Length: 190 words)

We propose the creation of a joint Oxford-Cambridge research center, which will develop policies to be enacted by governments, industry leaders, and others in order to minimize risks and maximize benefit from artificial intelligence (AI) development in the longer term. The center will focus explicitly on the long-term impacts of AI, the strategic implications of powerful AI systems as they come to exceed human capabilities in most domains of interest, and the policy responses that could best be used to mitigate the potential risks of this technology.

There are reasons to believe that unregulated and unconstrained development could incur significant dangers, both from "bad actors" like irresponsible governments, and from the unprecedented capability of the technology itself. For past high-impact technologies (e.g. nuclear fission), policy has often followed implementation, giving rise to catastrophic risks. It is important to avoid this with superintelligence: safety strategies, which may require decades to implement, must be developed before broadly superhuman, general-purpose AI becomes feasible.

This center represents a step change in technology policy: a comprehensive initiative to formulate, analyze, and test policy and regulatory approaches for a transformative technology in advance of its creation.