This site is primarily about **existential risk from future misaligned advanced AI**. That means we focus on dangers that are:

- at the scale of human extinction, rather than smaller-scale risks (even though many smaller risks are also significant).
- expected to be caused by future, highly-advanced AI systems that can outsmart humans, rather than systems that exist today.
- caused by AI acting in ways that its human designers and users did not intend, rather than by AI following human instructions to do harmful things.

This site aims to inform people about the risk of human extinction due to AI, rather than to advocate for any particular policy to address it. The site's content is intended to reflect the views of AI safety researchers in general. When there is substantial disagreement within the field (which is often the case), we attempt to represent all of the major positions.

Alternative phrasings

-

Related

-

Scratchpad

Current plan is to add another article, probably unlisted but linked from the main page footer, containing information like: who writes this, who funds this, how did this start, link to Rob's YouTube, mention that we have paid editors and volunteers

Murphant also notes similarities to What is Stampy's AI Safety Info?