How DeepMind Plans to Build Safe AI

Originalartikel

Backup

<html>
 From Medium: <blockquote> At a high level, safety research at DeepMind focuses on designing systems that reliably function as intended while discovering and mitigating possible near-term and long-term risks. Technical AI safety is a relatively nascent but rapidly evolving field, with its contents ranging from high-level and theoretical to empirical and concrete. The goal of this blog is to contribute to the development of the field and encourage substantive engagement with the technical ideas discussed, and in doing so, advance our collective understanding of AI safety. In this inaugural post, we discuss three areas of technical AI safety: specification, robustness, and assurance. Future posts will broadly fit within the framework outlined here. While our views will inevitably evolve over time, we feel these three areas cover a sufficiently wide spectrum to provide a useful categorisation for ongoing and future research. </blockquote> Read more <div class=„ctx-subscribe-container ctx-personalization-container ctx_default_placement ctx-clearfix“/> <div class=„ctx-social-container ctx_default_placement ctx-clearfix“/> <div class=„ctx-module-container ctx_default_placement ctx-clearfix“/> </html>