RE: AI Action Plan RFI

While the development of artificial intelligence applications has and will continue to unlock economic value, it's critical that government decisionmakers don't lose sight of the more far-reaching implications. Today, AI is a useful tool for the automation of cognitive tasks, much like any other computer program. In that context, far too many concerns about "AI risk" from misinformation or misuse of existing systems are overblown. Nevertheless, in the limit of arbitrarily powerful AI—machines that can do everything humans can do and more—there's a serious risk of humans losing control of our own creations.

However strange the idea might seem at first, the arguments are so overwhelming that hundreds of qualified experts—including, notably, Turing Award winners Geoffrey Hinton and Yoshua Bengio and the CEOs of frontier AI companies Demis Hassabis, Sam Altman, and Dario Amodei—signed a 2023 statement proclaiming that "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war" (https://www.safe.ai/work/statement-on-ai-risk).

It's hard to judge exactly when in the future trajectory of AI development the truly existential threats will emerge, and harder still to judge what the optimal policy response should be. (A theoretical argument about what should happen "the limit of arbitrarily powerful AI" doesn't say much about how that limit is approached.) But the stunning advances of deep learning over the last dozen years (and, notably, large language models over the past six) present clear evidence that it's not too early to act in some way.

As a mere concerned citizen (a computer programmer by profession, having studied the issue as a private individual), I'm not sure I have the expertise to say exactly what ought to be done, but one obvious suggestion is that the U.S. AI Safety Institute should be well-funded and its recommendations taken seriously. Also, mandatory reporting requirements for frontier AI training runs would provide critical visibility without unduly burdening innovation.

As Paul Christiano of the U.S. AISI points out (https://www.alignmentforum.org/posts/Hw26MrLuhGWH7kBLm/ai-alignment-is-distinct-from-its-near-term-applications), the technical project of figuring out how to build secure AI systems is distinct from the passing political fads of the moment. This is not a partisan issue. I am,

Faithfully yours,

Zack M. Davis
