Purpose: inform and convince individuals of the possibility of superintelligence
Audience: General audiences who are starting to learn about superintelligence but have their doubts, also technical individuals who want a high level summary
Length: This is slightly on the longer side but feels important to have a bit of a nuanced and strongly argumentative stance

The rise of superintelligent AI systems poses significant potential risks that are increasingly being recognized by experts in the field. Leaders from prominent AI research organizations, such as OpenAI, Google DeepMind, and Anthropic, have expressed concerns over the potential for AI to pose an existential risk to humanity. They point out that it's conceivable that within the next decade, AI systems could exceed human skill level in most domains and carry out productive activity comparable to today's largest corporations. This power, while having the potential to create a significantly prosperous future, also carries risks that need to be managed proactively rather than if/when things get out of hand..

The Centre for AI Safety has listed several disaster scenarios that could occur due to the rise of superintelligent AI. These scenarios include the weaponization of AI, AI-generated misinformation that could destabilize society, the concentration of AI power into fewer hands, and human over-dependence on AI, leading to a situation similar to the one portrayed in the film Wall-E. The Centre has emphasized that mitigating the risk of extinction from AI should be a global priority, on par with other societal-scale risks such as pandemics and nuclear war.

However, it's important to note that there's a diversity of opinions on the subject. Some experts, like Arvind Narayanan, a computer scientist at Princeton University, argue that current AI is not capable enough for these risks to materialize and that the focus on these distant risks can distract from more immediate AI-related harms. Other concerns raised include biased automated decision-making, the spread of misinformation, and increased inequality, particularly for those on the wrong side of the digital divide.

In response to these risks, OpenAI's leaders have called for the regulation of "superintelligent" AI, arguing for an international regulator to inspect systems, require audits, test for compliance with safety standards, and place restrictions on degrees of deployment and levels of security. They also call for some degree of coordination among companies working on cutting-edge AI research to ensure that the development of ever-more powerful models integrates smoothly with society while prioritizing safety.

Despite these risks, OpenAI's leaders believe that the continued development of powerful AI systems is worth the risk, as they could lead to a much better world than what we can imagine today. They caution that pausing development could also be dangerous, as the cost to build AI decreases each year, the number of actors building it is rapidly increasing, and it's inherently part of the technological path we are on. They suggest that stopping it would require something like a global surveillance regime, which isn't guaranteed to work. Therefore, the focus should be on getting it right rather than stopping it.

If you're curious and want to keep reading:
https://www.theguardian.com/technology/2023/may/24/openai-leaders-call-regulation-prevent-ai-destroying-humanity
https://www.bbc.com/news/uk-65746524
https://www.amazon.com/Superintelligence-Dangers-Strategies-Nick-Bostrom/dp/1501227742

Alternative phrasings

-

Related

-

Scratchpad