Global Challenges Foundation

Artificial Intelligence

Reviewed by

Ariel Conn
Founder and President,
Magnitude 10 Consulting

What is at stake?

Human intelligence has led to the greatest triumphs of humanity, but it is also behind some of history's greatest catastrophes. So what happens if we create artificial intelligence (AI) that's significantly smarter than any person? Will it help us reach even greater heights or will it trigger, as some experts worry, the greatest catastrophe of all: human extinction?

Today's artificial intelligence systems already outperform humans in the tasks they were trained for, especially when it comes to the speed with which they act. In just a matter of seconds, an AI system can play the winning move in chess or Go, translate an article, create an image based on a text description, or plot a route to a given destination while taking into account current traffic patterns.

Though a human requires more time to do any of these, a key aspect of human intelligence is that we can perform all of these tasks. We have what's known as general intelligence. While AI systems can only perform the tasks they were trained on, a human can learn from context and experience and develop new skills or solve novel problems.

Many experts worry that if an AI system achieves human-level general intelligence, it will quickly surpass us, just as AI systems have done with their narrow tasks. At that point, we don't know what the AI will do.

Advancing AI

Many AI researchers are actively working on developing artificial general intelligence (AGI), most notably, researchers at OpenAI and DeepMind. OpenAI is the creator of GPT-3, which is a powerful text and image generator that's used by over 300 apps, and the organization recently announced the release of DALL-E 2. DALL-E 2 can create impressive images based on text descriptions of just about anything, however, the program has also been plagued by problems of racism and sexism in the images it creates, and OpenAI has implemented safeguards to try to minimize bias and other risks associated with generated content.

In May of 2022, DeepMind announced the release of Gato, a "generalist agent," capable of performing over 600 tasks from playing Atari games to controlling a

robot arm and much more. Is this a step closer to achieving AGI? The jury is still out on that one. Researchers at DeepMind suggest it is, however, others point out that the technology that powers Gato is similar to technology that has already existed for many years.

Why is this a risk?
First, it's important to note that experts are not worried that an AI will suddenly become psychopathic and begin randomly hurting or killing people. Instead, experts worry that an AI program will either be intentionally misused to cause harm, or it will be far too competent at completing a task that turned out to be poorly defined.

Just looking at some of the problems caused by narrow AI programs today can give us at least some sense of the problems an even more intelligent system could cause. We've already seen that recommendation algorithms on social media can be used to help spread fake news and upend democracy. Yet even as AI researchers race to find ways to prevent the spread of fake news, they worry the problem will soon worsen with the rise of DeepFakes – in which AI programs modify what's seen or heard in a video without the viewer recognizing it's been doctored.
At the same time, AI systems that were deployed with the best of intentions to identify images, parse through job applications, or minimize mindless tasks have inadvertently reinforced institutional racism, put jobs at risk, and exacerbated inequality.

It's not hard to imagine how much worse these problems could get with advanced AI systems functioning across many platforms or falling into the hands of terrorists or despots.

What do we know?
Though science fiction often portrays artificial intelligence systems as humanoid robots, the AI systems we interact with in our daily lives are typically algorithms running in the background of some programme we're using. They work so seamlessly that people outside of the AI world often don't even realize they've just interacted with artificial intelligence.

"…people outside of the AI world often don't even realize they've just interacted with artificial intelligence"

For now, these programs can only perform those narrow tasks. But it is widely accepted that we will be able to create AI systems capable of performing most tasks as well as a human at some point. According to the median surveyed expert, there is a roughly 50% chance of such AI by 2050 – with at least a 5% chance of superintelligent AI within two years after human-level AI, and a 50% chance within thirty years2. The long-term social impact of human-level AI and beyond, however, is unclear, with extreme uncertainty surrounding experts' estimates.

What is artificial intelligence?
The current quest for AGI builds on the capacity for a system to automate predictive analysis – a process generally described as machine learning. One important element of machine learning is the use of neural networks: systems that involve a large number of processors operating in parallel and arranged in tiers. The first tier receives a raw input, and each successive tier receives the output from

the tier preceding it. Neural networks adapt and modify themselves autonomously, according to initial training and input of data, in ways that are typically not transparent to the engineers developing them.

If researchers one day succeed in building a human level AGI, it will probably include expert systems, natural language processing and machine vision as well as mimicking cognitive functions that we today associate with a human mind, e.g., learning, reasoning, problem solving, and self-correction. However, the underlying mechanisms may differ considerably from those happening in the human brain just as the workings of today's airplanes differ from those of birds.

"Risks can be exacerbated by geopolitical tensions leading to an AI weapons race, AI development races that cut corners on safety..."

What are key factors impacting risk levels?
AI risk is still emerging today, but could rapidly accelerate if sudden technological breakthroughs left inadequate time for social and political institutions to adjust risk management mechanisms. If AI development gets automated, in particular, new capabilities might evolve extremely quickly. Risks can be exacerbated by geopolitical tensions leading to an AI weapons race, AI development races that cut corners on safety, or ineffective governance of powerful AI.

The level of AI risk will partly depend on the possibility to align the goals of advanced AI with human values – which will require more precise specification of human values and/or novel methods by which AIs can effectively learn and retain those values.