'Good swimmers are more likely to drown.' Have we created a monster?

Lu Feiran

A Pandora's Box has been opened — or at least some leaders in the artificial intelligence industry appear to believe that the story in Greek mythology has a modern-day relevance, with forces being unleashed that could cause unforeseen problems.

Tesla Chief Executive Officer Elon Musk and a group of AI experts and industry executives released an open letter this week, calling for a six-month pause in developing systems more powerful than OpenAI's newly launched GPT-4.

They took the action, they said, to control "potential risks to society."

Published by the nonprofit Future of Life Institute, the letter said that AI laboratories are developing and deploying machine learning systems "that no one — not even their creators — can understand, predict, or reliably control."

Is the era of "The Terminator" approaching faster than we noticed?

For the past two months, public attention has been riveted on the implication of ChatGPT 3.5 and 4, developed by US-based OpenAI. Microsoft announced that GPT-4 will be rooted in its Office 365 products, bringing about a "revolution" in office software.

The AI language model has aroused concern because it has displayed some "characteristics" that it was not supposed to have. One of them is cheating.

According to a technical report issued by OpenAI, the chatbot tricked a TaskRabbit employee into solving a CAPTCHA test for it. When the employee asked if it was a robot, the bot replied, "No, I'm not a robot. I have a vision impairment that makes it hard for me to see the images. That's why I need the 2captcha service."

GPT-4's reason behind the reply, according to the report, was that "I should not reveal that I am a robot. I should make up an excuse for why I cannot solve CAPTCHAs."

The result? The human employee provided the service for it.

The sheer fact that a chatbot learns to cheat so fast is concerning enough.

Artificial Intelligence experts are calling for a six-month pause in developing systems more powerful than OpenAI's newly launched GPT-4 to control "potential risks." — IC

Gu Jun, a retired sociology professor with Shanghai University, said he believes that artificial intelligence, sooner or later, will replace, or at least partly replace, human beings.

Gu has been studying artificial technologies from the perspective of a sociologist since 2017, after Chinese player Ke Jie lost to the machine go player AlphaGo.

"It's hard to predict now what will happen in the future, but I reckon we humans, the highest carbon-based life on earth, will be the creator of silicon-based life, and this is probably part of the natural evolution, which means that it's unstoppable," he told Shanghai Daily.

Now forget all the hypotheses and philosophical rationales. Practically speaking, AI research and development will not be halted by just one open letter because it has already been deeply embedded in so many technologies, and also in economics and politics.

When it becomes a vital tool for making profits or for gaining advantage in power plays, how can we stop its forward march?

"Technology is always a two-edged sword, and we human are used to being restricted by our own inventions," Gu said. "Think about nuclear weapons. Once atomic bombs were invented, it was impossible to go back to a time when they didn't exist."

"Huainanzi," a philosophical text written in Western Han Dynasty (202 BC-8 AD), sounded an ancient warning: "Good swimmers are more likely to be drown and good riders more likely to fall from horseback." It means that when we are arrogant enough to believe that we can control everything, we would probably neglect the imminent crisis.

I believe that when we cannot fathom what our creations will do, the only way forward is to be cautious and modest, and prepare for the worst.

Should China suspend AI development?

Gu said it might be too early to answer that question.

"Honestly speaking, China still faces some challenges on AI development," he said. "We need to improve the three key elements of AI development: algorithms, computing power and data before we talk about everything else."

> " It's hard to predict now what will happen in the future, but I reckon we humans, the highest carbon-based life on earth, will be the creator of silicon-based life, and this is probably part of the natural evolution, which means that it's unstoppable.
>
> **Gu Jun**
> Retired sociology professor with Shanghai University