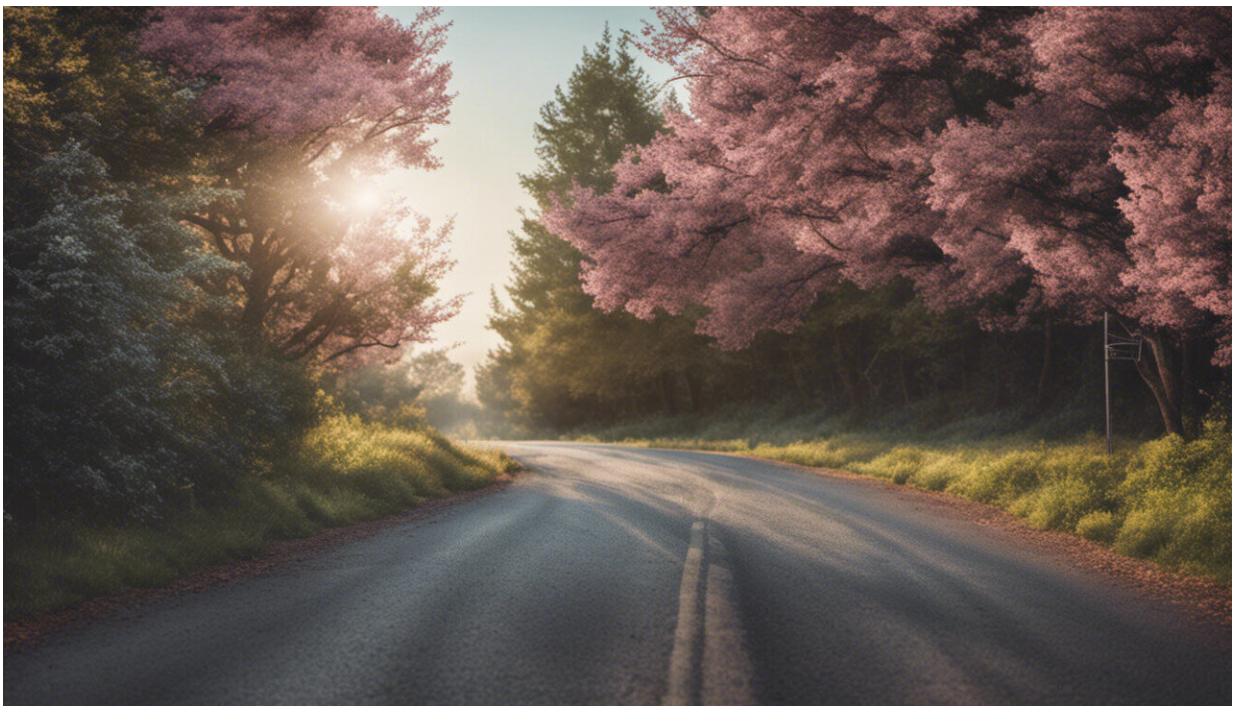


Opinion: Giving AI direct control over anything is a bad idea—here's how it could do us real harm

July 31 2023, by Guillaume Thierry



Credit: AI-generated image ([disclaimer](#))

The release of the advanced chatbot [ChatGPT](#) in 2022 got everyone talking about artificial intelligence (AI). Its sophisticated capabilities amplified concerns about AI becoming so advanced that soon we would not be able to control it. This even led some experts and industry leaders

to warn that the [technology could lead to human extinction](#).

Other commentators, though, were not convinced. The distinguished professor of linguistics, Noam Chomsky, dismissed ChatGPT as "[hi-tech plagiarism](#)".

For years, I was relaxed about the prospect of AI's impact on human existence and our environment. That's because I always thought of it as a guide or adviser to humans. But the prospect of AIs taking decisions—exerting executive control—is another matter. And it's one that is now being seriously entertained.

One of the key reasons we shouldn't let AI have executive power is that it entirely lacks emotion, which is crucial for decision-making. Without emotion, empathy and a moral compass, you have created the perfect psychopath. The resulting system may be highly intelligent, but it will lack the human emotional core that enables it to measure the potentially devastating emotional consequences of an otherwise rational decision.

When AI takes executive control

Importantly, we shouldn't only think of AI as an [existential threat](#) if we were to [put it in charge of nuclear arsenals](#). There is essentially no limit to the number of positions of control from which it could exert unimaginable damage.

Consider, for example, how AI can already identify and organize the information required to build your own conservatory. Current iterations of the technology can [guide you effectively through each step](#) of the build and prevent many beginner's mistakes. But in future, an AI might act as project manager and coordinate the build by selecting contractors and paying them directly from your budget.

AI is already being used in pretty much all domains of information processing and data analysis—from [modeling weather patterns](#) to [controlling driverless vehicles](#) to [helping with medical diagnoses](#). But this is where problems start—when we let AI systems take the critical step up from the role of adviser to that of executive manager.

Instead of just suggesting remedies to a [company's accounts](#), what if an AI was given [direct control](#), with the ability to implement procedures for recovering debts, make bank transfers, and maximize profits—with no limits on how to do this. Or imagine an AI system not only providing a [diagnosis based on X-rays](#), but being given the power to directly prescribe treatments or medication.

You might start feeling uneasy about such scenarios—I certainly would. The reason might be your intuition that these machines do not really have "souls". They are just programs designed to digest huge amounts of information in order to simplify complex data into much simpler patterns, allowing humans to make decisions with more confidence. They do not—and cannot—have emotions, which are intimately linked to biological senses and instincts.

Emotions and morals

[Emotional intelligence](#) is the ability to manage our emotions to overcome stress, empathize, and communicate effectively. This arguably matters more in the context of decision-making than intelligence alone, because the best decision is not always the most rational one.

It's likely that intelligence, the ability to reason and operate logically, can be embedded into AI-powered systems so they can make rational decisions. But imagine asking a powerful AI with executive capabilities to [resolve the climate crisis](#). The first thing it might be inspired to do is drastically reduce the human population.

This deduction does not need much explaining. We humans are, almost by definition, the source of pollution in every possible form. A humanity and climate change would be resolved. It's not the choice that human decision-makers would come to, one hopes, but an AI would find its own solutions—impenetrable and unencumbered by a human aversion to causing harm. And if it had executive power, there might not be anything to stop it from proceeding.

Sabotage scenarios

How about sabotaging [sensors and monitors controlling food farms](#)? This might happen gradually at first, pushing controls just past a tipping point so that no human notices the crops are condemned. Under certain scenarios, this could quickly lead to famine.

Alternatively, how about shutting down [air traffic control](#) globally, or simply crashing all planes flying at any one time? Some 22,000 planes are normally in the air simultaneously, which adds up to a potential death toll of several million people.

If you think that we are far from being in that situation, think again. AIs already drive cars and [fly military aircraft](#), autonomously.

Alternatively, how about shutting down access to bank accounts across vast regions of the world, [triggering civil unrest](#) everywhere at once? Or shutting off [computer-controlled heating systems](#) in the middle of winter, or air-conditioning systems at the peak of summer heat?

In short, an AI system does not have to be put in charge of nuclear weapons to represent a serious threat to humanity. But while we're on this topic, if an AI system was powerful and intelligent enough, it could find a way of faking an attack on a country with nuclear weapons, triggering a human-initiated retaliation.

Could AI kill large numbers of humans? The answer has to be yes, in theory. But this depends in large part on humans deciding to give it executive control. I can't really think of anything more terrifying than an AI that can make decisions and has the power to implement them.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Opinion: Giving AI direct control over anything is a bad idea—here's how it could do us real harm (2023, July 31) retrieved 27 April 2024 from <https://techxplore.com/news/2023-07-opinion-ai-bad-ideahere-real.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.