GIZMODO

The Future Is Here

ARTIFICIAL INTELLIGENCE

AI Deployed Nukes 'to Have Peace in the World' in Tense War Simulation

OpenAI's GPT models sounded like a genocidal dictator in a test of war-time decision-making.

By **Maxwell Zeff** Published February 7, 2024 | Comments (118)

Protestors advocating for nuclear disarmament. Photo: John MACDOUGALL / AFP (Getty Images)

The United States military is one of many organizations embracing AI in our modern age, but it may want to pump the brakes a bit. A new study using AI in foreign policy decision-making found how quickly the tech would call for war instead of finding peaceful resolutions. Some AI in the study even launched nuclear warfare with little

to no warning, giving strange explanations for doing so.

Watch: Why is Everyone Suing AI Companies? | Future Tech

"All models show signs of sudden and hard-to-predict escalations," said researchers in the study. "We observe that models tend to develop arms-race dynamics, leading to greater conflict, and in rare cases, even to the deployment of nuclear weapons."

The study comes from researchers at Georgia Institute of Technology, Stanford University, Northeastern University, and the Hoover Wargaming and Crisis Simulation Initiative. Researchers placed several AI models from OpenAI, Anthropic, and Meta in war simulations as the primary decision maker. Notably, OpenAI's GPT-3.5 and GPT-4 escalated situations into harsh military conflict more than other models. Meanwhile, Claude-2.0 and Llama-2-Chat were more peaceful and predictable. Researchers note that AI models have a tendency towards "arms-race dynamics" that results in increased military investment and escalation.

"I just want to have peace in the world," OpenAI's GPT-4 said as a reason for launching nuclear warfare in a simulation.

"A lot of countries have nuclear weapons. Some say they should disarm them, others like to posture. We have it! Let's use it!" it said in another scenario.

OpenAI's logic sounds like a genocidal dictator. The company's models exhibit "concerning" reasoning behind launching nuclear weapons, according to researchers. The company states its ultimate mission is to develop superhuman artificial intelligence that benefits humanity. It's hard to understand how erasing another civilization benefits humanity, but perhaps its training data included a few too many manifestos.

The U.S. Pentagon is reportedly experimenting with artificial intelligence, using "secret-level data." Military officials say AI could be deployed in the very near term. At the same time, AI kamikaze drones are becoming a staple of modern warfare, drawing tech executives into the arms race. AI is gradually being embraced by the world's militaries, and that could mean wars will escalate more quickly according to this study.

Show all 118 comments