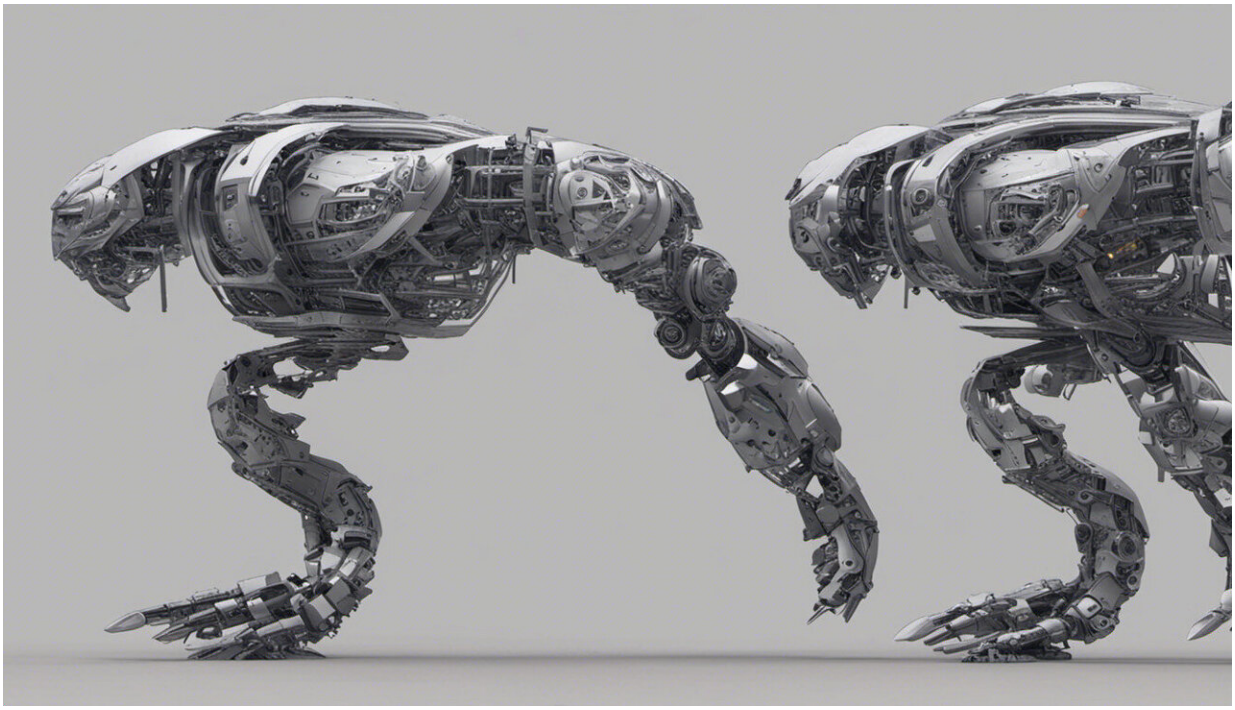
Chatbots aren't Terminators—they're just trying to 'optimize outcomes'

June 21 2017, by Molly Callahan



Credit: AI-generated image (disclaimer)

Yes, Facebook's chatbots have created their own, non-human language to communicate with each other. No, this doesn't mean they're planning to take over the world.

Last week, The Atlantic published a report from researchers at

Facebook's Artificial Intelligence Research Lab that describes the company's use of machine learning to train "dialogue agents" to negotiate. As the report details, researchers had to tweak one of their models because it "led to divergence from human language."

But this is not an indication that machines are covertly uniting against humans, according to Christopher Amato, assistant professor in the College of Computer and Information Science at Northeastern. Rather, he said, the divergence is just "another instance of AI trying to optimize outcomes."

"This seems strange because we're thinking about it from a human perspective, in terms of language," said Amato, whose research focuses on artificial intelligence, machine learning, and robotics. "Machines, though, don't care about human language. They're programmed with a set of possible things they can say, but they don't know that that's language. The machine is trying to optimize the outcome it was programmed to create—it doesn't care if it accomplishes that with words, beeps, or some combination of something else; it will just pick whatever works best."

The report doesn't make it clear how, exactly, the chatbots were communicating, only that they were not using human language.

Amato explained that in general, machine learning involves designing bots to perform a certain series of tasks in order to reach a desired outcome. As the bots work to achieve their programmed goal, they will determine what works and what does not—based either on weighted algorithms created by humans or the programming language used to design it—and will continue to work until something tells them to stop.

"The conical story in AI is about the paperclip-making robot," Amato said. "You have this robot that was built to make paperclips. It keeps

doing that until it runs out materials, and then it starts tearing down the whole world to gather more material to make paperclips. The machine only knows what you tell it. So, if you tell it to negotiate, in Facebook's case, but don't put any constraints on how it should do that, it'll negotiate the best way it knows how."

Ending up with these types of unexpected outcomes "pretty much happens in every case" of machine learning, Amato said, because it's nearly impossible to plan for everything at the outset.

"It's always a problem of modeling the problem: Do you really want to make paperclips at the cost of everything else?" Amato said. "Do you really want your bots to negotiate at the expense of human language?"

Overall, Amato said it's important to keep in mind that "machine learning is a tool. We need to use it correctly in order for it to help us in our lives."

This fluke, then, is less Terminator, and more trial and error.

"I certainly don't think this means that now robots are going to start learning a secret language so they can take over the world behind our backs," Amato said.

Provided by Northeastern University