The Consciences of Robot Warriors

Charles Day
Physics Today

This past November, I attended chipmaker Nvidia's GPU Technology Conference, which was held in Washington, DC. Its focus was artificial intelligence. You might be wondering, as I did, what AI has to do with graphical processing units and, by extension, video games. Here's your answer: The calculations needed to track, say, the stream of ignited fuel that shoots from Master Chief's flamethrower in *Halo 3* are similar in character to those needed to run the deep-learning algorithms that underlie one of the most popular variants of AI.

The examples of AI in the opening keynote speech by Greg Estes, Nvidia's VP of developer marketing, were impressive: AI-driven cars, AI-written classical music, AI-translated spoken language, and AI-diagnosed medical images. They stopped well short, however, of what some people might consider the pinnacle of AI: sentient robots. Estes explained one of the obstacles. "It's much easier to train a car to avoid hitting things than it is to train a robot to pick things up it sees for the first time."

Tesla CEO Elon Musk is not waiting for technical obstacles to be surmounted to consider the ethical implications of AI. He was among the signatories of a letter addressed to the United Nations Convention on Certain Conventional Weapons. The letter warned of the dangers of robot soldiers, drones, and other high-tech weapons whose AIs would, in effect, be making life-or-death decisions. If an AI-equipped drone mistook civilians for soldiers and slaughtered them, would anyone be held accountable for the war crime?

The question of military accountability has an evolving history. In the years after World War I, some German military personnel were tried for war crimes. Among the defendants was Karl Neumann, the commanding officer of a U-boat that torpedoed and sank a British hospital ship, the *Dover Castle*. Germany's supreme court acquitted the officer on the grounds that "all civilized recognize the principle that a subordinate is covered by the orders of his superiors." That defense was weakened in World War II by two provisos: First, if the soldier's orders left him with a moral decision to carry them out, he would still be guilty. Second, if the orders themselves were illegal, then their execution would also be illegal.

It's not a stretch to compare a soldier's orders to the program that governs today's autonomous drones. Both are sets of instructions. Indeed, the program is the more binding because the drone has no choice whatsoever. In those circumstances, the programmer and the person who authorized the program are responsible—and culpable in the event of a bug that causes a war crime.

 Published by the IEEE Computer Society 2469-7087/19/$33.00

If a weaponized drone fired on the basis not of programmed instructions but on the criteria it acquired through deep learning, the authors of the algorithms would still be culpable, in my view. If an AI were ever to attain human-level consciousness, then it would presumably be capable of acquiring a conscience of its own and the moral responsibility that goes with it. And somewhere in between, maybe at dog-level intelligence, both the AI and its human owner would share responsibility.

If an AI-equipped drone mistook civilians for soldiers and slaughtered them, would anyone be held accountable for the war crime?

ABOUT THE AUTHOR

Charles Day is *Physics Today*'s editor in chief. The views in this column are his own and not necessarily those of either *Physics Today* or its publisher, the American Institute of Physics.

This article originally appeared in
Computing in Science & Engineering, *vol. 20, no. 1, 2018.*