

RSIS Commentary is a platform to provide timely and, where appropriate, policy-relevant commentary and analysis of topical and contemporary issues. The authors' views are their own and do not represent the official position of the S. Rajaratnam School of International Studies, NTU. These commentaries may be reproduced with prior permission from RSIS and due recognition to the author(s) and RSIS. Please email to Mr Yang Razali Kassim, Editor RSIS Commentary at RSISPublications@ntu.edu.sg.

AI: Could It Be More Ethical than Humans?

By Richard A. Bitzinger

SYNOPSIS

Artificial intelligence in autonomous systems (i.e., drones) can address human error and fatigue issues, but also, in the future, concerns over ethical behaviour on the battlefield. Installing an algorithmic “moral compass” in AI, however, will be challenging.

COMMENTARY

A COMMON theme among many discussions concerning the military uses of artificial intelligence (AI) is the “Skynet” trope: the fear that AI will be self-aware and decide to turn on its masters. Inherent in this argument is the contention that AI does not share the same ethical constraints that humans do.

While almost certainly an over-exaggeration, the Skynet scenario does highlight the problem of ensuring that the ethical behaviour we believe is incumbent on humans in combat is not lost as we increasingly devolve battlefield decision-making to autonomous systems. In fact, we may have it the wrong way around: rather than being less ethical than humans, AI might be programmed to be more ethical. And that could have both positive and negative repercussions.

Ubiquity of Drones on the battlefield

AI on the battlefield is generally tied in with autonomous systems, in particular the use of drones and other robotic systems. Traditionally, drones have been assigned to replacing humans in roles and missions synonymous with the “three Ds”: dull, dirty, or dangerous. Originally, drones – particularly unmanned aerial vehicles (UAVs), were used for surveillance and reconnaissance over enemy territory.

Today, aerial drones are also employed as communications relays, or for laser-targeting or even electronic warfare (jamming). Other functions include policing duties, border patrol, and bomb disposal, all of which have expanded the utility and importance of robots.

The drawback, of course, is that currently such drones do not operate totally autonomously. There is always a human in the loop, usually controlling a drone remotely. This raises its own set of built-in challenges. The command and control (C2) of UAVs, let alone armed drones, is quite demanding. The support network behind drone use is enormous, especially for long-distance, long-endurance operations.

Drones often require satellites for target acquisition and strike-control, as well as secure datalinks; without satellites, drones need line-of-sight datalinks or relay aircraft to remain in contact with remote operators. Pilots do not come cheap, either: they have to be as skilled as a pilot of a manned aircraft, and the fatigue factor of remotely operating a drone can be high.

Moreover, drones have an incredibly high loss rate. A US Air Force (USAF) report in March 2009 showed that it had lost 70 Predators in air crashes during its operational history up to that time. Fifty-five were lost to equipment failure, operator error, or weather, while four had been shot down by enemy forces and 11 more were lost to accidents on combat missions. According to another USAF report, this time from 2015, the Predator, Reaper and Global Hawk drones were “the most accident-prone aircraft in the Air Force fleet”.

AI on the Battlefield

AI promises to be the solution to these problems of human fatigue and error. AI can take over many of the more monotonous tasks of remote piloting. In fact, the US Army already has in place such manned-unmanned teaming. The intention of such teaming, according to a 2012 report by the US Defence Science Board, is not to replace humans, “but rather to extend and complement human capability”.

It is only a few steps further, however, before AI takes over the lion’s share of such operations. This could mean AI ultimately replacing the final, still-human-centric decision as to when to launch a strike against a supposed military target.

The concern here is self-apparent: besides dealing with human fatigue, how do we ensure that the embedded AI is sophisticated enough to differentiate between civilian and military targets, and or is able to minimise sufficiently civilian collateral damage? In other words, how do we avoid AI-initiated massacres?

One answer is to always keep a human involved, operating a “kill switch” to prevent such tragedies. However, that negates much of the reason for devolving operations to AI in the first place, that is, to relieve humans of the possibility for error due to emotion, fatigue, or other shortcomings.

Are Humans Really that Ethical?

We like to think of ourselves as ultimately moral people, but in fact humans are capable of engaging unethical behaviour, almost on a daily basis. We run red lights, cheat on our taxes, or call in sick when we just want to stay home. More to the point, we are quite capable of forgiving ourselves for our ethical lapses. This is perhaps alright for small offences, but humans are quite good at rationalising large ethical failures.

Nowhere is this perhaps more seen than on the battlefield. Looting, the shooting of prisoners of war, and brutalisation of civilians is sadly too widespread. Worse, even in societies where one would think that ethics would be held in high esteem, there is an enormous capacity for excusing, validating, and condoning such behaviour. Look at the My Lai massacre in Vietnam, Abu Ghraib in Iraq, and more recently, President Trump's pardon of three soldiers convicted of war crimes. In these cases and more, few people are punished, the punishments slight or even reduced, and the behaviour rationalised.

Could AI be *More Ethical* than Humans?

AI, of course, is no more ethical than the people who program it. However, that could mean that such programs could be more rigidly ethical because it is hard-wired in their systems. In other words, some kind of algorithmic "moral compass" could be installed in autonomous systems, intended to look for and avoid ethical dilemmas.

Not that any of this would be easy. Such AI would require human checks and readjustments to machine-learning. In particular, the algorithms for AI intelligence-gathering and processing would have to be highly reliable, to ensure that autonomous systems can clearly differentiate between military and civilian targets.

Such AI might be even "*too ethical*", if it develops a "zero-tolerance" for error and refuses to ever attack a target for fear that civilians might be hurt. And we have to keep in mind that machine-learning can also be susceptible to bias and to adversarial attacks.

The point is, however, that AI could conceivably deal with the problems of human error and fatigue as well as ethics in the battlespace. It is a technical fix, but a philosophical one as well.

Richard A. Bitzinger is a Visiting Senior Fellow with the Military Transformations Programme at the S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU), Singapore. He is formerly with the RAND Corp. and the Centre for Strategic and Budgetary Assessments.

Nanyang Technological University

Block S4, Level B3, 50 Nanyang Avenue, Singapore 639798
Tel: +65 6790 6982 | Fax: +65 6794 0617 | www.rsis.edu.sg