Predictability and Lethal Autonomous Weapons Systems (LAWS)

Wendell Wallach

Expert Testimony at the UN, Geneva
12 April 2016

Does predictability provide an overriding concept and perhaps a metric for evaluating when LAWS are acceptable or when they might be unacceptable under international humanitarian law? Arguably, if the behavior of an autonomous weapon is predictable, deploying it might be considered no different from, for example, launching a ballistic missile. This, of course, presumes that we can know how predictable the behavior of a specific autonomous weapon will be.

Over the past two years, this body has focused upon ethical and legal challenges to LAWS. In addition, there are contentions that autonomous weaponry will fail to perform as expected, will behave unpredictably on occasion, and are therefore inherently risky and liable to commit acts that violate IHL, even when this is not the intention of those who deploy the systems. Definitions for "predictability" and "risk," will help shine some light on whether or when autonomous weapons are inherently more hazardous than existing weapon systems.

"Reliability" and similar engineering terms, have traditionally served as a standard for evaluating mechanical systems and weaponry. Testing and experience in the field can provide a degree of certainty about the reliability of automated weapons. Older poorly maintained systems become less reliable.

Software is notoriously unreliable, often released with thousands of known and unknown programming errors. Countries and corporations have spent billions on software applications that were abandoned.

The need for a definition of "predictability" has only arisen because traditional engineering standards of reliability no longer suffice when evaluating the behavior of increasingly autonomous computers and robots, or computational systems with limited but expanding artificial intelligence and learning capabilities.

Nothing less than a law of physics is absolutely predictable. There are only degrees of predictability, which in theory can be represented as a probability. In the evaluation of weaponry, predictability means that within the task limits for which the system is designed, the anticipated behavior will be realized, yielding the intended result. Nevertheless, an unanticipated event, force, or resistance can alter the behavior of even highly predictable systems.

Autonomous systems are best understood as complex adaptive system. Within systems theory, complex adaptive systems act unpredictably on occasion, have tipping points that lead to fundamental reorganization, and can even display emergent properties that are difficult, if not impossible, to explain.

Complex adaptive systems fail for a variety of reasons:
Human error or incompetence offer one reason. Then there are design flaws and vulnerabilities. Third, are "normal accidents" – Charles Perrow's term to designate disasters where no one does anything wrong. Fourth, misconceptions regarding the likelihood of low probability events occurring. And finally "black swans" where a low probability high-impact event is not even recognized.

Unpredictable behavior will not necessarily be lethal. But even a low risk autonomous weapon will occasionally kill non-combatants, start a new conflict, or escalate hostilities.

Machine learning and other forms of artificial intelligence are becoming commonplace as integral components of a wide array of applications. Deep learning algorithms have recently beaten the world's best Go player and solved problems of perception, such as accurately labeling objects in a photograph. Such learning capabilities will be essential if autonomous systems are to demonstrate even a modest degree of situational awareness.

There is no means of ascertaining whether a complex adaptive system will behave in an uncertain manner short of exhaustive testing. Testing, however, is very costly, if not impossible for complex systems. To make matters worse, each software error fixed and each new feature added, can alter a system's behavior in ways that can only be ascertained through additional rounds of extensive testing. No military can support the time and expense entailed in testing systems that are continually being upgraded.

Learning systems are even more problematic. Each new task or strategy learned can alter a system's behavior and performance. Furthermore, learning is not just a process of adding and altering information, it can alter the very algorithm that processes the information. Placing a system on the battlefield that can change its programming significantly raises the risk of uncertain behavior. Retesting dynamic systems that are constantly learning is impossible.

The risk that a bad event will occur is often quantified as the probability of the event multiplied by its consequences. Clearly, an autonomous system that functions as a platform for a machine gun has an immediate destructive impact that pales in comparison to an autonomous weapon system that can launch a ballistic missile.

In summary, while increasing autonomy, improving intelligence, and machine learning can boost the system's accuracy in performing certain tasks, they can also increase the unpredictability in how a system performs overall. Risk will rise relative to the power of the munitions the system can discharge.

Lethal autonomy is a feature that can be added to any weapon systems. Over time, increasingly sophisticated LAWS will be deployed. Therefore it behooves the member states of CCW to not be shortsighted in their evaluation of what will be a very broad class of military applications. CCW must not appear to green light autonomous systems that can detonate weapons of mass destruction. Given the high level of risks, powerful munitions, such as autonomous ballistic missiles or autonomous submarines capable of launching nuclear warheads, must be prohibited. Deploying systems that can alter their algorithms, their programming, is also foolhardy.

Member states may differ on the degree of unpredictability and level of risk they will accept in weapon systems. Hopefully, these comments have convinced you of the importance of reviewing between today and December the degree of unpredictable risk posed by various autonomous weapons configurations. Your decision should not be whether any autonomous systems must be prohibited, but rather how broadly encompassing the prohibition on LAWS must be.

Wendell Wallach
Yale University Institution for Social and Policy Studies
Interdisciplinary Center For Bioethics
P.O. Box 208209
New Haven, Connecticut 06520-8209

Wendell.wallach@yale.edu