THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Minimum levels of human intervention in autonomous attacks

Tim McFarland

The debate about regulating the development and use of autonomous weapon systems (AWS) is far from over, but there is emerging agreement that at least one important factor should guide any regulatory efforts: the degree to which a human operator can directly intervene in an AWS operation to receive information about the activities of the weapon and, if needed, apply manual control measures. International humanitarian law (IHL) does not explicitly state that any specific type or degree of human intervention is needed, but at least one simple limitation can be found: personnel must be able to intervene enough to ensure they can meet whatever IHL obligations they normally bear in relation to an operation. This paper discusses that requirement and some of its implications for AWS adoption.

When discussing AWS, a distinction must be drawn between two aspects of control over weapons. The first aspect is the narrow set of 'direct' measures traditionally associated with manually operating a weapon: assessing a potential target, aiming, pulling a trigger, and so on. The second aspect is the much broader set of control measures that States and their armed forces take to ensure that lethal force is applied in accordance with legal obligations: reviews of weapon systems; training of personnel; formulation of strategic goals, operational objectives and rules of engagement; analysis and vetting of potential targets; and so on. It is the first aspect which is most at issue when discussing AWS and it is those measures which are referred to here as 'human intervention' in operation of a weapon system.

The autonomous technologies which spark the most legal controversy at present are those which promise to displace a human operator from directly performing those tasks. AWS users must understand the extent to which that can be done consistently with the rules of IHL.

Legal limits on weapon autonomy

There are no references to AWS in any IHL treaty or in customary law, nor do AWS fall within the scope of any weapon-specific prohibitions. In fact, IHL contains only minimal references to any particular means of controlling weapons. Legal limits on autonomous capabilities in weapons must therefore be inferred from the principles, rules and goals of general IHL, but the unique nature of autonomous control makes that somewhat challenging. Autonomous control will partly or fully displace some personnel from the roles which the law assumes they will play, and it delves into areas which weapon technologies do not typically touch: specifically, the decision-making processes leading to an act of violence.

Requirements for human intervention

The burden of complying with the rules of IHL rests with people. In some cases, an obligation is formally assigned to individuals: 'those who plan or decide upon an attack' must undertake a range of precautionary measures aimed at minimising civilian harm when preparing to conduct attacks, and those who conduct the attack must also take 'constant care' to minimise civilian harm. Responsible persons must retain the practical ability to meet their legal obligations when operating AWS. They must be afforded the capability to ensure that the weapon systems for which they are responsible behave consistently with the constraints imposed by IHL. That ability rests on two foundations: access to sufficient information about the attack, the behaviour of the weapon and its interaction with targets and the environment; and a sufficient ability to affect the behaviour of the weapon as required by changing circumstances. That is the capacity for human intervention that is required for compliance with IHL.

An approach to describing the requisite capacity for human intervention

A more difficult challenge is how to quantify that requisite capacity for intervention in AWS operations. Some provisions in IHL treaties suggest a very high standard is required. In practice, though, some inability to manually intervene in operation of a weapon is often tolerated: not all missiles can be actively guided in flight; land and sea mines may be left emplaced for considerable time without direct supervision; and so on. AWS further complicate the issue in that they may greatly reduce the need for, or the possibility of, direct human intervention in weapon system activities, without necessarily increasing the chance of a proscribed outcome to an operation.

One way to approach this challenge is to characterise the law's requirement for human intervention as an exercise in risk management.

Where a weapon system is manually operated, the responsibility to ensure it is used in compliance with legal rules is principally a responsibility to operate the weapon in a certain way. Where a weapon system is operating autonomously, though, its control system plays the role of 'operator' to some extent. The task of the responsible person is instead to ensure that the weapon system is operated consistently with the State's overall system of control and, as far as possible, ensure that the actions the weapon system takes during an operation are consistent with the State's legal obligations. That essentially amounts to monitoring for and responding to developments which would unacceptably increase the risk of violations of IHL. In other words, it amounts to managing the risks of operating the AWS: specifically, the risks that the weapon's control system might cause it to act in a way which would fail to meet the obligations borne by the responsible person.

Many sources of such risk also arise in operations with manual weapon systems, but some are unique to AWS: in particular, the behaviour of complex control system software, the tendency of automated systems to fail when faced with challenges which go beyond those for which their control system was explicitly developed and the possibility of 'runaway failure' of an AWS if no person can intervene. Developers and potential operators of AWS should look to risk factors such as those, and the capacity for a human operator to intervene when things go wrong, when making decisions about legal compliance. ●

Tim McFarland, 'Minimum Levels of Human Intervention in Autonomous Attacks' (2022) 27(3) *Journal of Conflict & Security Law* 387–409

This research received funding from the Australian Government's Next Generation Technologies Fund through Trusted Autonomous Systems, a Defence Cooperative Research Centre. The views and opinions expressed are those of the author, and do not necessarily reflect the views of the Australian Government or any other institution. They also do not constitute legal advice. Cover photo by Spc Carlos Cuebas Fantauzzi. The appearance of US Department of Defense (DoD) visual information does not imply or constitute DoD endorsement.

Law and the Future of War
School of Law
The University of Queensland
law.uq.edu.au/future-war

TRUSTED AUTONOMOUS SYSTEMS
DEFENCE CRC

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA