

THE COMPATIBILITY OF AUTONOMOUS WEAPONS WITH THE PRINCIPLE OF DISTINCTION IN THE LAW OF ARMED CONFLICT

ELLIOT WINTER*

Abstract The law of armed conflict requires ‘distinction’ between civilians and combatants and provides that only the latter may be targeted. However, for proper implementation, distinction requires advanced observation and recognition abilities as well as the capacity to exercise judgement based on situational awareness. While the observation and recognition abilities of machines may now surpass those of humans, the capacity of machines to exercise judgement remains significantly more limited than our own. Consequently, this article contends that the deployment of ‘autonomous weapons’ based on current levels of technological sophistication would be incompatible with distinction and that, as such, their use in conflict would be unlawful.

Keywords: public international law, law of armed conflict, international humanitarian law, distinction, autonomous weapons, machine observation, machine recognition, machine judgement, artificial intelligence.

I. INTRODUCTION

The law of armed conflict (LOAC) is responsible for regulating the conduct of hostilities and protecting people in situations of both international and intranational violence. The pre-eminent principle of the regime is that of ‘distinction’: the notion that one must discern between civilians and combatants and only direct attacks against the latter. The logic behind distinction is that while it is militarily necessary for a combatant to attack enemy personnel and materiel in order to achieve victory; it is inhumane to attack civilians or their property as their destruction would cause suffering without getting the combatant any closer to victory. In this sense, distinction is a simple concept concerned with seeing people and objects and categorising them. However, application of the principle is complicated by contextual considerations. For example, it is possible for any person, or object, in war to move back and forth between ‘targetable’ and ‘untargetable’ status depending on whatever they happen to be doing, or being used for, at any

* Lecturer, Newcastle University Law School, elliott.winter@newcastle.ac.uk. A debt of gratitude is owed to Dr Conall Mallory and Dr Elena Katselli for their support.

given time. A civilian may work at a toy factory in the morning then pick up arms and fight in the evening. A soldier may be actively scouting an enemy area one minute but then become wounded and thus *hors de combat* (out of action) the next. A building may go from housing troops one day, to refugees the next. A hospital may be commandeered and used as a weapons depot. Hitherto, these inherent difficulties in distinction have been resolved by humans who can detect contextual shifts.

However, ‘autonomous weapons’—machines that are capable of waging war independently after deployment—are now on the horizon. There is much debate over whether these machines, if deployed, would comply with various aspects of LOAC. To date, much focus has been placed on the ‘humanity’ of autonomous weapons¹ or on who may be held accountable when things go wrong.² This article will instead tackle the issue of whether such machines could comply with the principle of distinction. To do this, it will first consider the principle of distinction and the nature of autonomous weapons. It will then consider the intersection of these two phenomena—and for this a framework is needed.

According to Singer (one of the world’s leading experts on changes in twenty-first-century warfare) robots are ‘man-made devices with three key components ... “sensors” that monitor the environment and detect changes in it, “processors” or “artificial intelligence” ... that decide how to respond and “effectors” that act upon the environment in a manner that reflects the decisions’.³ This useful starting point will be adopted with a number of amendments. First, it unnecessary to consider ‘effectors’, or hardware generally, as it is ‘decisions’ with which we are concerned. Secondly, for the current analysis it is more helpful to separate ‘artificial intelligence’ into two strands, as this allows for more detailed consideration of its components. Thus, after providing an exposition of the fundamentals, the article will analyse: (i) the extent to which machines can ‘observe’ (ie the extent to which they can be equipped with adequate sensors); (ii) the extent to which machines can ‘recognise’ that which they have observed (the first component of their artificial intelligence) and (iii) the extent to which machines can make appropriate ‘judgements’ on action (the second, higher, component of their artificial intelligence). These are the three abilities that autonomous weapons would need to master if they were to comply with distinction.

As will be seen, ‘observation’ is an area in which technology developers have made huge strides. ‘Recognition’ too has seen significant development and the

¹ O Ulgen, ‘Human Dignity in an Age of Autonomous Weapons: Are We in Danger of Losing an “Elementary Consideration of Humanity”?’ (2017) 17 *Baltic Yearbook of International Law* 167.

² D Amoroso and B Giordano, ‘Who Is to Blame for Autonomous Weapons Systems’ Misdoings?’ in E Carpanelli and N Lazzarini (eds), *Use and Misuse of New Technologies: Contemporary Challenges in International and European Law* (Springer 2019) 211.

³ PW Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (Penguin 2009) 67.

systems are now highly sophisticated, but the technology has been highly controversial in some areas such as facial recognition. Machine ‘judgement’ is the Holy Grail of artificial intelligence, but more limited advancement has been made there. Owing to the present restrictions on such intelligence, it will be argued that technology is not currently capable of delivering a fully autonomous machine able to wage war while satisfying the principle of distinction.

II. AN OVERVIEW OF DISTINCTION

The principle of distinction encapsulates the fundamental divide in conflict between armed actors (who may be targeted) and civilians (who may not). The modern expression of distinction, in the context of international armed conflict (IAC), can be found in Additional Protocol I which states that parties to a conflict ‘shall at all times distinguish between the civilian population and combatants ... and accordingly shall direct their operations only against military objectives’.⁴ In the context of non-international armed conflict (NIAC), Additional Protocol II states that ‘the civilian population ... shall not be the object of attack [and] acts or threats of violence [designed] to spread terror among the civilian population are prohibited’.⁵ In terms of status, distinction is evidently a ‘rule’ of LOAC owing, for example, to its inclusion in Additional Protocols I and II. However, that does not necessarily mean that it is a ‘principle’ or that it has crystallised into ‘customary international law’ (with the effect that it binds all States regardless of whether they are bound by the relevant treaties).

Whether distinction is a principle matters because, as Kolb stated, principles provide ‘gravitational points ... for understanding and correctly applying the law’.⁶ However, there is a difficulty in that LOAC has no single, conclusive, list of principles. Rather, there are manifold, often contradictory, pronouncements emanating from different institutions, uttered at different times and in the pursuit of different ends. This inconsistency has proven to be especially problematic in the context of ‘humanity’ and ‘military necessity’ and it prompted this author’s previous finding that those concepts ought to be regarded as ‘pillars’ rather than ‘principles’ of LOAC.⁷ Happily, the difficulty is not so acute here, as distinction is always included in statements of the principles of LOAC. In terms of judicial pronouncements, the *Nuclear*

⁴ Protocol Additional to the Geneva Conventions of 12 August 1949 and relating to the Protection of Victims of International Armed Conflicts (adopted 08 June 1977, entered into force 7 December 1978) 1125 UNTS 3 art 48.

⁵ Protocol Additional to the Geneva Conventions of 12 August 1949 and Relating to the Protection of Victims of Non-International Armed Conflicts (adopted 08 June 1977, entered into force 7 December 1978) 1125 UNTS 609 art 13(2).

⁶ R Kolb, *Advanced Introduction to International Humanitarian Law* (Edward Elgar 2014) 78.

⁷ E Winter, ‘Pillars not Principles: The Status of Humanity and Military Necessity in the Law of Armed Conflict’ (2020) 25 *JC&SL* 1.

Weapons judgment of the International Court of Justice (ICJ) brands distinction as a ‘cardinal principle’.⁸ In terms of State pronouncements, the UK Ministry of Defence has said that distinction is a ‘fundamental principle’,⁹ as has Denmark’s Ministry of Defence¹⁰ and New Zealand’s Defence Force¹¹. As regards academic opinion, one might consider Solis who observed that distinction is a ‘core principle’ of LOAC,¹² or Kolb who opined that ‘without general principles of law such as ... distinction [LOAC] would be largely blind’.¹³ Through a synthesis of these assertions, it can be seen without doubt that distinction is a ‘principle’ of LOAC.¹⁴

In terms of customary international law status—achieved through a combination of general State practice and *opinio juris*¹⁵—it can be said with confidence that distinction is also firmly established. The International Committee of the Red Cross (ICRC) confirmed this in its Customary Law Study¹⁶ which provides distinction as ‘Rule 1’ (of one hundred and sixty one) and mirrors the language of Additional Protocol I in stating ‘the parties to [a] conflict must at all times distinguish between civilians and combatants [and] attacks may only be directed against combatants’.¹⁷ Customary status is important because not all States are party to the treaties that set out distinction in its modern form, such as Additional Protocols I and II, with notable examples being the USA, India, Iran, Turkey and Israel.¹⁸ Its customary status ensures that distinction binds all States regardless of treaty participation.

In summary, distinction involves differentiating between civilians on one hand and combatants (or other ‘fighters’) on the other—with the corollary that only the latter may be targeted. Furthermore, distinction is firmly enshrined in treaty law as well as forming both a principle and customary rule of LOAC. It is important to bear in mind that these points will remain true even as new technologies, such as autonomous weapons, appear on the battlefield. As Schmitt and Widmar point out, ‘while the weaponry and

⁸ *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion) [1996] ICJ Rep 226 para 78.

⁹ United Kingdom Ministry of Defence, *The Manual of the Law of Armed Conflict* (Oxford University Press 2004) 21.

¹⁰ Danish Ministry of Defence, *Military Manual on International Law Relevant to Danish Armed Forces in International Operations* (Defence Command Denmark 2016) 145–55.

¹¹ New Zealand Defence Force, *Manual of Armed Forces Law*, vol 4 (DM69, 2nd edn, New Zealand Defence Force 2019) 4.6.1.

¹² GD Solis, *The Law of Armed Conflict* (2nd edn, Cambridge University Press 2016) 309.

¹³ Kolb (n 6) 77.

¹⁴ J Pictet, *Development and Principles of International Humanitarian Law* (Martinus Nijhoff 1985).

¹⁵ Statute of the International Court of Justice (adopted 26 June 1945, entered into force 24 October 1945) UKTS 67 (1946) art 38(1)(b).

¹⁶ JM Henckaerts and L Doswald-Beck, *Customary International Humanitarian Law, Volume I: Rules* (Cambridge University Press 2005).

¹⁷ *ibid* 3.

¹⁸ MJ Matheson, ‘The United States Position on the Relation of Customary International Law to the 1977 Protocols Additional to the 1949 Geneva Conventions’ (1987) 2 *American University Journal of International Law and Policy* 419.

tactics of targeting continue to evolve with unprecedented advances in technology and innovation, the fundamental principles of targeting law will remain binding rules for the foreseeable future'.¹⁹ Nonetheless, it would be naïve to suggest that no challenges to our *interpretation* of the law will be posed by new technologies. Therefore, it is important to look ahead and anticipate those challenges by understanding the technologies that will cause them to arise. It is to that effort that this article now turns. In particular, autonomous weapons technology will be explored and its potential significance assessed.

III. AN OVERVIEW OF AUTONOMOUS WEAPONS

In 2012, the US Department of Defense adopted a working definition providing that an autonomous weapon is a 'weapon system that, once activated, can select and engage targets without further intervention by a human operator'.²⁰ This definition was widely cited in the earlier days of autonomous weapons discourse and, as will be seen shortly, continues to influence the discussion today. It captures the core of what is meant by an autonomous weapon: namely a machine comprised of hardware and software that might be released into a battlespace to perform its function independently. It is the absence of direct human involvement in their operation that separates 'autonomous weapons' from the more familiar technology found in drones which, while 'unmanned', are still piloted by humans—albeit from distant military bunkers.²¹ This critical difference led the ICRC to state that the deployment of autonomous weapons would represent a 'paradigm shift' in the way hostilities are conducted.²²

Since those early days, international efforts have been directed towards finding a suitable definition for autonomous weapons, principally under the auspices of the Convention on Conventional Weapons (CCW).²³ The decision to discuss the issue was taken by the parties in 2013²⁴ but there has been limited progress. Perhaps the most tangible step thus far has been the establishment of a Group of Governmental Experts (GGE) to consider the matter. The GGE has not yet settled on a definition for the technology, but

¹⁹ MN Schmitt and E Widmar, 'The Law of Targeting' in PAL Ducheine *et al.* (eds), *Targeting: The Challenges of Modern Warfare* (Springer 2016) 121.

²⁰ United States Department of Defense, 'Autonomy in Weapons Systems' (2012) Directive 3000.09, Glossary Part II <<https://bit.ly/2UCP4fc>>.

²¹ S Casey-Maslen, 'Pandora's Box? Drone Strikes Under *jus ad bellum, jus in bello*, and International Human Rights Law' (2012) 94/886 International Review of the Red Cross 597.

²² International Committee of the Red Cross, 'Autonomous Weapon Systems - Q&A' (ICRC, 12 November 2014) <<http://bit.ly/2ixib2p>>.

²³ Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (adopted 10 October 1980, entered into force 2 December 1983) 1342 UNTS 137.

²⁴ Convention on Conventional Weapons, 'Meeting of the High Contracting Parties: Final Report' (16 December 2013) UN Doc CCW/MSP/2013/10 para 32.

the chairperson has defended this by claiming that ‘while a definition would be eventually essential, the absence of an agreed definition should not prevent the Group from moving forward with the discussions’.²⁵ Similarly, the GGE has asserted that a definition based on technological attributes alone would be of limited utility as technology develops so quickly that any definition agreed upon would soon be rendered redundant. Instead, as the chairperson observed, the GGE favours focussing on the extent of the link between machines and human beings: the ‘human-machine interface’.²⁶ This approach is in line with the opinion of many individual States. For example, the UK favours a ‘technology-agnostic’ approach which emphasises the importance of human control rather than a definition based on technical characteristics.²⁷ In summary then, it is the extent of human–machine interaction—rather than specific hardware, software or mission functions—that should be used to define an autonomous weapon. Only such an approach will be robust enough to deal with all candidate devices, irrespective of their physical form, processing capacity or operational capabilities.

In terms of the degree of human–machine interaction that is actually permitted, the position of some key actors is that a ‘truly’ autonomous weapon must be able to act without the need for any further input (such as data, decisions or approvals) at all from human beings after deployment. It was observed above that the early US definition applied to weapons that could operate ‘without further intervention by a human operator’.²⁸ France agrees and is equally as demanding as the US when it comes to the level of independence required of the machine before it will be considered autonomous, stating that ‘LAWS [lethal autonomous weapons systems] should be understood as implying a total absence of human supervision’.²⁹ The UK has gone even further, asserting that a truly autonomous weapon would be ‘capable of understanding, interpreting and applying higher level intent and direction based on a precise understanding and appreciation of what a commander intends to do and perhaps more importantly why’.³⁰ These pronouncements give a sense of just how high the bar can be set for autonomy: absolute independence. Only States really know their motivations for setting such high thresholds. However, to a cynic, it might seem that

²⁵ Group of Governmental Experts, ‘Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems’ (23 October 2018) UN Doc CCW/GGE.1/2018/3, Annex III (Chair’s Summary) para 2.

²⁶ *ibid* paras 2 and 5.

²⁷ United Kingdom, ‘Statement to the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems’ Plenary Meeting of the Group of Governmental Experts (25–29 March 2019) para 3.

²⁸ United States Department of Defense (n 20).

²⁹ France, ‘Characterization of a LAWS’ Informal Meeting of the Group of Governmental Experts (11–15 April 2016).

³⁰ United Kingdom, ‘Working towards a Definition of LAWS’ Informal Meeting of the Group of Governmental Experts (11–15 April 2016) para 4.

these definitions are merely a legal wheeze: they allow States to make impressive claims concerning how robust they will be in regulating ‘autonomous weapons’, while ensuring that those regulations would in fact only apply to super-intelligent machines that are unlikely to exist for decades. Even then, such machines may only form a tiny subset of what most people would refer to in common parlance as an ‘autonomous weapon’.

While the general trajectory for a definition seems broadly to have been set (ie a human–machine interaction test based on the absence of contact after deployment) the specifics remain disputed by the GGE. A number of contenders have been submitted by delegates and those were amalgamated into a list of ten by the chairperson.³¹ For present purposes, the candidate that will be relied on is ‘a system that can select and attack targets without human intervention, in other words a system that self-initiates an attack’.³² This option has been selected in part because it does not demand the total lack of human ‘supervision’ (ie humans ‘on the loop’) that some do. Such approaches are too narrow, as militaries will probably try to give humans override capability where possible. Further, this definition does not require the system to possess ‘higher level intent’ in the way that States such as the UK would prefer. Their approaches would again be too restrictive given the current state of artificial intelligence (considered below). Nor does it require that the system is capable of self-learning—or, indeed, actually ‘lethal’—as other proposals require, as neither of those criteria seem to be of any particular necessity. A machine that is capable of wounding is in as much need of regulation as one that can kill.

Still, the definition adopted for this piece is strict in the sense of demanding ‘full’ autonomy from humans. It is conceded that this approach is not without its critics and that autonomous weapons are unlikely to be utterly autonomous. Some engagement with external elements is likely to remain—such as the sharing of information with combat soldiers, intelligence scouts or, indeed, other machines.³³ Furthermore, as Bradshaw put it, autonomy is not a ‘unidimensional concept’ (which, at its simplest, could be said to be comprised of self-direction and self-sufficiency) and it has a broad range of potential meanings.³⁴ As a result, some have suggested that it would be more accurate to say that such machines will bear ‘autonomous characteristics’ rather than possessing full autonomy.³⁵ For example, Van Rompaey has criticised the trend of ‘persistent anthropomorphism’ in this field and takes the view that weapons systems, albeit with increased independence, will merely form part

³¹ Group of Governmental Experts (n 25), Annex III (Chair’s Summary) para 6.

³² *ibid.*

³³ Defense Science Board, *Task Force Report: The Role of Autonomy in DoD Systems* (United States Department of Defense 2012) 21 and 59.

³⁴ JM Bradshaw *et al.*, ‘The Seven Deadly Myths of “Autonomous Systems”’ (2013) 28 *IEEE Intelligent Systems* 54, 54.

³⁵ L Suchman and J Weber, ‘Human–Machine Autonomies’ in N Bhuta *et al.* (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016).

of broader ‘network-centric sociotechnical systems’.³⁶ More particularly, he argues that:

LAWS are perceived as the replacement for human soldiers, and that makes us believe they possess some of the inherent features of a human soldier. Those irrelevant features include physical embodiment, mental individualization, and weaponization. This makes the CCW’s discussions ... underinclusive [so we should consider] taking a networks perspective instead [of focussing on the] interactions between different systems.³⁷

This may indeed be a more accurate view in the near term, given the current state of technology. Furthermore, there is some support for this view from States. For example, the UK has intimated that it does not anticipate the level of autonomy outlined above coming to pass imminently.³⁸ Instead, it notes that there is a broad spectrum of technological capabilities leading up to that point—exhibiting varying degrees of autonomy—and that these also need to be considered. Nonetheless, to preclude the notion of fully autonomous machines is to ignore the inexorable and exponential technological leaps that will likely continue to be seen.³⁹ Further, it ignores the vast military advantages that would come from possessing a genuinely autonomous weapon, such as immunity from ‘jamming’ (which involves blocking radio and other communications to disrupt operations; but which is ineffective against systems not reliant on external communications) and general rapidity of action.

In fact, sophisticated *defensive* weaponry bearing limited autonomy already exists. There are sentry guns and missile interception technologies that repel incoming threats without the need for human authorisation such as ‘Phalanx’⁴⁰, ‘Iron Dome’⁴¹ and ‘Super aEgis-II’⁴². Admittedly, when it comes to *offensive*, advanced and mobile technologies (which are the focus of this article) development has been slower. However, one can see early efforts here in the form of projects such as ‘Taranis’⁴³, an aerial combat vehicle being developed by BAE Systems (a UK-based aerospace manufacturer), or ‘Atlas’⁴⁴, a humanoid-like machine being developed by Boston Dynamics (a US-based private robotics company)—although the latter is presently being designed for general purpose duties rather than the conduct of war.

³⁶ L Van Rompaey, ‘Shifting from Autonomous Weapons to Military Networks’ (2019) 10 Journal of International Humanitarian Legal Studies 111, 111. ³⁷ *ibid* 115.

³⁸ United Kingdom (n 27) para 4.

³⁹ M Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (Allen Lane 2017).

⁴⁰ Raytheon Missiles & Defense, ‘Phalanx Weapon System’ (Raytheon Missiles & Defense) <<https://bit.ly/2UEy4Fw>>.

⁴¹ Raytheon Missiles & Defense, ‘Iron Dome System and SkyHunter Missile’ (Raytheon Missiles & Defense) <<https://bit.ly/3dTYTNz>>.

⁴² Dodaam Systems, ‘Super aEgis II: The Best Mobile Remote Controlled Weapon Station’ (Dodaam Systems) <<http://bit.ly/2G0Hlhi>>.

⁴³ BAE Systems, ‘Taranis’ (BAE Systems) <<http://bit.ly/2uamk2j>>.

⁴⁴ Boston Dynamics, ‘Atlas’ (Boston Dynamics) <<http://bit.ly/2sP5pwi>>.

Therefore, although we are yet to see the deployment of any independent and offensive autonomous weapons; the regulation of such technology remains worthy of study.

The final matter to consider in this overview of autonomous weapons is the form that any future regulation might take. The GGE has, helpfully, distilled States' suggestions on this into four categories. The first is for a 'legally-binding instrument' to be agreed that would, *inter alia*, 'ensure human control over the critical functions' (ie targeting decisions) of these machines.⁴⁵ The second is for a 'political declaration' that would set out key principles in the field.⁴⁶ The third is to focus on existing international law and to discuss its application to this new technology.⁴⁷ The fourth is to proceed on the basis that LOAC is capable of regulating autonomous weapons satisfactorily in its existing form.⁴⁸ Given that States (through the GGE) have been unable to agree even a broad direction of travel for regulation after five years, it seems unlikely that the first, second or even third categories will develop into tangible proposals. This is against a backdrop of wider international security tensions and dissensus.⁴⁹ Consequently, the most viable approach for regulating autonomous weapons falls into the fourth category: taking LOAC as it stands and ensuring this emergent technology complies. Indeed, States are already bound to do so under Additional Protocol I.⁵⁰

This article now will focus on *existing* LOAC. As noted above, a tripartite approach, modelled on Singer's original three components,⁵¹ will be used to address the intersection of the 'distinction' and 'autonomous weapon' phenomena: looking in turn at (i) machine observation; (ii) machine recognition and (iii) machine judgement.

IV. THE INTERSECTION OF DISTINCTION AND AUTONOMOUS WEAPONS: MACHINE OBSERVATION

A. Machine Observation in Computer Games

The first element to be considered is machine observation. 'Observation' here simply means seeing or perceiving without any attendant processing or cognition—those matters will be considered later. LOAC does not stipulate precisely how observations are to be conducted. It does not, for example, require that specific equipment be used to monitor the battlefield nor does it set minimum requirements for matters such as the resolution of imagery used in making observations or the amount of time devoted to such exercises.

⁴⁵ Group of Governmental Experts (n 25) para 28(a).

⁴⁷ *ibid*, para 28(c).

⁴⁹ R Gowan, 'Muddling Through to 2030: The Long Decline of International Security Cooperation' (2018) 42 *The Fletcher Forum of World Affairs* 55.

⁵⁰ Additional Protocol I (n 4) art 36.

⁴⁶ *ibid*, para 28(b).

⁴⁸ *ibid*, para 28(d).

⁵¹ Singer (n 3) 67.

Rather, the obligation is framed more loosely and it is simply stipulated that those who plan or decide upon attacks must ‘do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects ... but are military objectives’.⁵² Inevitably, the requirement to do ‘everything feasible’ is vague and open to interpretation. Schmitt, for his part, has argued that it would require ‘full use of onboard ... sensors that could boost the reliability of target identification’.⁵³ That is almost certainly correct—though Thurnher disagrees on some of the finer details such as the need to incorporate the use of observations gathered by other, external, units.⁵⁴ The real question for present purposes is whether the sort of sensor technology that could be mounted on autonomous weapons is up to the challenges of modern warfare.

Even in the relatively recent past, observation was a tricky task for robots. Difficulties were found, for example, in the context of computer games. From the 1970s, programmers have been designing games such as Pong and Space Invaders in which humans and machines battle each other, with the results projected onto a screen in real-time to allow for the consequences to be displayed and future choices to be made. Of course, the screen is used by the human player only. A key problem with getting computers to play using imagery is that ‘making sense of the screen is a visual task that computers have never really taken to’.⁵⁵ Indeed, ‘looking at the monitor and judging actions accordingly ... has always been a special human skill’.⁵⁶ In practice, developers simply work around this problem by permitting computers to play games using ‘direct inputs’ from the system rather than indirectly via a monitor.⁵⁷ Really, this is a cheat as it allows game outputs to go straight into a computer’s processing systems and frees machines from the difficult task of dealing with imagery.

In 2013, a team led by Mnih at UK developer DeepMind Technologies began to tackle this imagery challenge head-on by putting together a system that worked with visual inputs from games such as Pong on the Atari 2600 console.⁵⁸ Crucially, the system was ‘not privy to the internal state of the emulator [i.e. the console]’ but instead was compelled to train itself using only RGB (red, green and blue) video imagery in the same way that humans must do.⁵⁹ The results of the project are relevant to the autonomous weapons debate as the developers had introduced ‘a new deep learning model for

⁵² Additional Protocol I (n 4) art 57(2)(a)(i).

⁵³ MN Schmitt, ‘Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics’ (2013) Harvard National Security Journal 23 <<https://bit.ly/2WCnpwb>>.

⁵⁴ JS Thurnher, ‘Feasible Precautions in Attack and Autonomous Weapons’ in WH von Heinegg, R Frau and T Singer (eds), *Dehumanization of Warfare: Legal Implications of New Weapon Technologies* (Springer 2018) 109.

⁵⁵ The Physics arXiv Blog, ‘Neural Net Learns Breakout Then Thrashes Human Gamers’ (*The Physics arXiv Blog*, 23 December 2013) para 9 <<https://bit.ly/2swuqCf>>.

⁵⁷ *ibid*, para 9.

⁵⁸ V Mnih *et al.*, ‘Playing Atari with Deep Reinforcement Learning’ (2013) Cornell University arXiv <<https://bit.ly/38h18b8>>.

⁵⁶ *ibid*, para 2.

⁵⁹ *ibid* 2 and 7.

reinforcement learning, and demonstrated its ability to master ... computer games, using only raw pixels as input'.⁶⁰ Thus, the system is a watershed example of a machine perceiving visual imagery rather than relying on 'direct' input. That said, it should be noted that Mnih's team had simplified the task for the system by reducing the amount of visual information it was required to process. The Atari 2600 ordinarily produces frames with a resolution of 210 x 160 pixels and uses a pallet of 128 different colours. For this project, the developers reduced the resolution to 84 x 84 pixels and cut the pallet to only 4 colours. The result was that the computer was required to process far fewer data than humans.⁶¹ Machines had been given the ability to see—but only a diminished world.

B. Machine Observation in Mapping Drones

Of course, there has been much advancement in 'machine observation' even since DeepMind's 2013 attempt. One company that embodies this is Exyn Technologies which emerged in 2014 as a spin-off from the University of Pennsylvania and which develops multi-purpose drone technology with varying degrees of flight automation ranging from pilot-assisted to fully-autonomous.⁶² In terms of the mechanics, the technology uses a variety of different sensors to perceive its environment including visual cameras, LIDAR (light detection and ranging), radar and RGBD (red green blue depth) sensors. The resulting data are then synthesised in real time for the purposes of simultaneous localisation and mapping. In other words, the system can look around to observe its location, use that information to construct a map of the area and then discern its position within that location in much the same way that a human would. There is no pre-programming to tell the system what its environment looks like or where it is situated therein.

The Exyn system has seen rapid commercialisation. A recent commission came from Ascot Resources which was considering exploitation of the long-abandoned Big Missouri Ridge mine. To survey the mine using human geologists would have been prohibitive as many areas were inaccessible or unsafe. Crowe reported on the deployment and noted that Exyn's technology allowed for this task to be performed by drone-like robots 'without the need for a pilot or prior map'.⁶³ John Kiernan, Chief Operating Officer of Ascot, stated that 'Exyn came to our site to show us the autonomous capabilities of their drone technology, and [we] were very impressed with the timeliness and quality of the data acquired'.⁶⁴ In fact, Ascot now plans to explore further uses of autonomous technology in this context because the Exyn system proved to be

⁶⁰ *ibid* 8. ⁶¹ *ibid* 5–6. ⁶² GRASP, 'Research Projects' (GRASP) <<https://bit.ly/2xJV7pw>>.

⁶³ S Crowe, 'Exyn Drone Maps Inactive Mine on the Fly' (The Robot Report, 19 November 2019) para 3 <<https://bit.ly/35R3YRo>>.

⁶⁴ *ibid*, para 10.

safer, cheaper and faster than human surveyors and delivered a more complete map.⁶⁵ Another mining firm with which Exyn has worked, Dundee Precious Metals, revealed how the fully autonomous aerial robots are transforming their monitoring systems with increased safety and efficiency.⁶⁶ According to Theophile Yameogo, Vice President of Digital Innovation at Dundee, ‘the Exyn [machines] allow frequent and hi-resolution mapping of underground environments ... we are very excited at the results of the maps we are seeing’.⁶⁷ Indeed, Leotaud reports that both firms expect to be more efficient in the future as a result of having access to better maps of intended mines.⁶⁸ In summary, Exyn has demonstrated that modern machines can be equipped with high-performance capabilities in ‘observation’ and that this technology has already been deployed to great effect in the mining sector.

C. Machine Observation beyond Visual Line of Sight

While the relatively self-contained nature of mines provided an ideal starting point for the trialing of ‘seeing’ robots, it is not the end point. There has been frenetic development aimed at a more general ability to go ‘beyond visual line of sight’ (BVLOS)—to go past the reach of operators.⁶⁹ Naturally, this is a critical step for any fully ‘autonomous’ system. BVLOS capability is of acute importance for machines being developed for the automated delivery sector by companies such as Amazon, UPS and Google Wing (a sibling company of Google). One of the more notable developments in this context was the permission given in 2019 by the US Federal Aviation Administration (FAA) to a University of Alaska project—operated in collaboration with technology companies Iris (visual systems), Echodyne (tracking) and Skyfront (drone hardware)—to inspect an oil pipeline. In terms of Iris and its visual (ie ‘observation’) systems, the company took the process of automated inspection a step further in November 2019 in a project for Kansas Department of Transportation and Kansas State University. Its detect-and-avoid system allowed a robot to undertake over one hundred miles of power line inspections while flying BVLOS. Naturally, this required a high degree of observational capability. The operation marked the first BVLOS autonomous drone flight under the FAA’s small unmanned aircraft system rules, known as ‘Part 107’, which did not require visual observers or ground-based radar.⁷⁰ Of course, Iris is not the only company active in this area and the number of rivals is growing.⁷¹

⁶⁵ *ibid.*, para 9.

⁶⁶ VR Leotaud, ‘Exyn Technologies Introduces Robots into Dundee Precious Metals’ Gold Mines’ *Mining.Com* (28 February 2019) <<https://bit.ly/2vCBEWO>>. ⁶⁷ *ibid.*, para 4.

⁶⁸ *ibid.*, para 5.

⁶⁹ Nanalyze, ‘How Autonomous Drone Flights Will Go Beyond Line of Sight’ (Nanalyze, 31 December 2019) <<https://bit.ly/3aGR7o8>>.

⁷⁰ United States, ‘Electronic Code of Federal Regulations’ Title 14 Chapter I Subchapter F Part 107. ⁷¹ Nanalyze, ‘7 Startups Using Drones for Inspections & Monitoring’ (Nanalyze, 18 July 2017) <<https://bit.ly/2R7eJuo>>.

European companies have shown interest in giving even passenger-carrying aircraft the ability observe the world around them. On 16 January 2020, the European aircraft manufacturer Airbus executed the first fully autonomous ‘vision-based take-off’ using a test aircraft at Toulouse-Blagnac airport.⁷² The test was a success and the plane launched autonomously eight times in less than five hours. There are now plans for the development of similar vision-guided taxi and landing capabilities. Of course, planes with the ability to fly on autopilot are not a novelty as ‘fly-by-wire’ has been around for a long time.⁷³ However, existing systems navigate by radio-navigation including, for example, the ‘instrument landing system’ which ‘provides aircraft with horizontal and vertical guidance just before and during landing and, at certain fixed points, indicates the distance to the reference point of landing’.⁷⁴ In essence, traditional fly-by-wire is reliant on ground-based *radio* signals to operate. On the contrary, the Airbus tests were based on machine *observation*.

D. Machine Observation in Security Guard Systems

Development in machine observation is now extending beyond games, mapping and flying and into potential security applications, thus bringing its potential relevance to war and LOAC into sharper focus. One example is a system from Toronto-based Patriot One Technologies.⁷⁵ Patriot One was founded in 2016 and aims to provide ‘a single threat detection product’ for weapons-screening at public places with a view to preventing gun and knife crime and even terrorist incidents.⁷⁶ This issue appeared on the radar after a spate of mass shootings in the United States and the accompanying realisation—highlighted by academics such as Rocque and Duwe—that such events were occurring with increasing frequency.⁷⁷ Patriot One’s principal product in this area, PatScan, effectively functions as an automated security guard by using sensors to identify threats. The company boasts that the system provides ‘multi-sensor, layered security ... that [can] identify threats ... from parking lots to entry access and beyond’.⁷⁸

⁷² Airbus, ‘Airbus Demonstrates First Fully Automatic Vision-Based Take-Off’ (Airbus, 16 January 2020) para 1 <<https://bit.ly/2vjMWyK>>.

⁷³ GH Hunt, ‘The Evolution of Fly-By-Wire Control Techniques in the UK’ (1979) 83 *The Aeronautical Journal* 165.

⁷⁴ International Telecommunication Union, *Radio Regulations* (International Telecommunication Union 2012) 16 (art 1.104).

⁷⁵ Patriot One Technologies, ‘About’ (Patriot One Technologies) <<https://bit.ly/3dT3UWH>>.

⁷⁶ *ibid*, para 3.

⁷⁷ M Rocque and G Duwe, ‘Rampage Shootings: An Historical, Empirical, and Theoretical Overview’ (2018) 19 *Current Opinion in Psychology* 28, 30.

⁷⁸ Patriot One Technologies, ‘Introducing the PatScan Multi-Sensor Covert Threat Detection Platform’ (Patriot One Technologies) <<https://bit.ly/2x4Jucx>>.

There are four components to PatScan: 'PatScan Video', 'PatScan Radar', 'PatScan Magnetic' and 'PatScan Chemical'. First, as explained in a promotional video, PatScan uses video imagery taken from CCTV cameras to identify threats.⁷⁹ Second and third, radar and magnetic systems work together to scan people passing through a designated area, such as turnstiles, to determine if they are carrying any weapons.⁸⁰ In terms of operation, microwaves are generated and operate using resonance frequency patterns which make it possible for accompanying radar sensors to detect shapes. Simultaneously, magnetic fields are generated which can detect disturbances as objects pass through the field. Fourth, the system can also detect 'explosives and chemical hazards such as gunpowder and C4 in the air with 'parts-per billion sensitivity'.⁸¹ The visual, radar, magnetic and chemical inputs combine to paint a very clear picture of what is happening within the area of interest and thus allow the system to 'observe' what is happening. Indeed, it is even asserted that the ability of PatScan to observe is not limited to what is in plain sight and that it can also observe weapons concealed 'on body or in bag'.⁸² PatScan is thus an example of a highly perceptive system that could easily be turned to a military support role. Indeed, Patriot One took the Award for Anti-terrorism and Force Protection at the International Security Conference & Exposition ('ISC West') in 2017.⁸³

E. Machine Observation in Military Drones

Even more on-point in terms of the military applications of machine observation is the recent work of US defence manufacturer Raytheon. Raytheon, working with Exyn, has developed 'mapping autonomous drones' that are able to perceive their surroundings without access to GPS or mapping data.⁸⁴ According to the company, it has developed 'a fully autonomous aerial robot, that ... can operate in GPS-denied environments to map dense urban environments in 3-D [and] can dig deep to reveal tunnels, urban undergrounds and natural cave networks'.⁸⁵ It does this using 'a combination of sensors, including cameras and lidar [which is] similar to radar, but using pulsed, infrared laser light'.⁸⁶ The company boasts that the system collects 300,000 data points per second in order to map its environment and that it is sensitive enough to detect even dangling wires. In essence then, the same technology is at play here as was discussed above in the context of mineral exploration and, again, machines have developed a remarkable ability to not only 'observe' in incredibly high detail but also to record what they see for posterity. Raytheon recognised the potential that high-accuracy machine

⁷⁹ Patriot One Technologies (n 75).

⁸⁰ *ibid.*

⁸¹ *ibid.*

⁸² *ibid.*

⁸³ Patriot One Technologies, 'Patriot One Wins Best in Category at Security Industry Association Event at ISC West' (Patriot One Technologies, 4 June 2017) <<https://bit.ly/2TIqTMI>>.

⁸⁴ Aircoc, 'Where Hazards Lurk' (Aircoc, January 2020) para 5 <<https://bit.ly/2Tzvcb5>>.

⁸⁵ *ibid* para 3.

⁸⁶ *ibid* para 6.

observation systems might have in the context of urban combat where the battlespace is visually more complex than the traditional open battlefield. As it observed, ‘sloshing through dark, dangerous urban environments ... while disconnected from the outside world, is risky work [as] anything might lurk around the next bend’.⁸⁷

The consequence of all the above is that robots can observe at least as well as humans and, indeed, at higher resolution and with greater rapidity and full recording capability to boot. Their abilities have been refined to the extent that they can operate independently in areas including computer gaming, mineral exploitation, security screening, airliner take-off and battlefield mapping. Indeed, there is now an expectation that this technology will soon ‘take over many of the manual inspections, services, and deliveries currently done by humans’.⁸⁸ This seems inevitable, as does the adoption of machine observation technology by defence contractors and, in turn, militaries. In sum, machines have satisfied the first component of our tripartite test: observation.

V. THE INTERSECTION OF DISTINCTION AND AUTONOMOUS WEAPONS: MACHINE RECOGNITION

A. Machine Recognition and Military Uniforms

The ability of an autonomous weapon to ‘observe’ would no doubt be critical to its compliance with the principle of distinction. However, observation alone is not enough and such machines would need to go further by *recognising* that which they see. This is crucial because, as was explained above, distinction requires one to distinguish civilians apart from combatants. At first, this recognition task might appear to be simple. The stereotypical combatant appears clad in camouflage-pattern military uniform, adorned with various emblems to denote allegiance and rank, topped-off with a helmet and completed with a weapon. Indeed, this stereotype is usually reflected in the reality of how military personnel display themselves. As Hays Parks observed, ‘in international armed conflict, the wearing of standard uniforms by conventional military forces, including special operations forces, is the normal and expected standard’.⁸⁹ Similarly, according to Grant and Huntley, ‘the display of uniforms and weapons is the main way of distinguishing oneself in combat’.⁹⁰ Even for irregular forces such as militias and volunteer corps, Geneva Convention III requires (for prisoner of war status) that they, *inter alia*, bear a ‘fixed distinctive sign recognizable at a distance’ and that

⁸⁷ *ibid* para 2.

⁸⁸ Nanalyze (n 69) para 20.

⁸⁹ W Hays Parks, ‘Special Forces’ Wear of Non-Standard Uniforms’ (2003) 4 *Chicago Journal of International Law* 493, 542.

⁹⁰ M Grant and T Huntley, ‘Legal Issues in Special Operations’ in G Corn *et al.* (eds), *US Military Operations: Law, Policy and Practice* (Oxford University Press 2016) 589.

they ‘[carry] arms openly’.⁹¹ This reflects an earlier provision concerning such groups from the Hague Regulations requiring display of a ‘fixed distinctive emblem recognizable at a distance’.⁹² While, somewhat oddly, LOAC makes no such provision explicitly for regular military personnel, Gillich is undoubtedly correct when she observes that this rule on carrying a sign or emblem also applies to them saying: ‘it follows *a majore ad minus* that the obligation to wear at least a distinctive sign applies to members of armed forces too’.⁹³ In short, those involved in combat ordinarily wear a uniform, bear a distinctive emblem and carry arms openly. One might therefore imagine that, for compliance with distinction, it would be sufficient to programme an autonomous weapon to recognise enemy uniform designs, enemy symbols and enemy weapons. The reality is much more complex.

There are myriad reasons for this complexity. One is that while an individual is obliged to distinguish himself/herself as a combatant⁹⁴—and soldiers *ordinarily* discharge this obligation by wearing uniforms—there is no blanket requirement in LOAC to wear clothing of any particular type. Even members of the armed forces are not necessarily required to wear camouflage, green garments and the like. According to Gillich, ‘as to the appearance of regular armed forces ... [LOAC] remains silent’; preferring instead to delegate the appearance of military personnel to municipal law.⁹⁵ Additional Protocol I explicitly acknowledges deference to States in this context when it states that ‘Article [44] is not intended to change the generally accepted practice of States with respect to the wearing of ... uniform by combatants assigned to ... regular, uniformed armed units’.⁹⁶ At the fringes of the rules, Hays Parks notes that States may even dispense with uniforms altogether in the contexts of ‘intelligence collection or Special Forces operations in denied areas’.⁹⁷

Another reason for complexity in this area is that, when it comes to irregular forces, the standard is even more fluid. This is in part due to the fact that Additional Protocol I was written in the 1970s against a backdrop of decolonisation struggles which had spurred a broadening of LOAC to cover ‘freedom fighters’ or ‘guerrillas’. Those individuals were previously excluded from protection under Geneva Convention III owing to their failure to wear a distinctive sign.⁹⁸ Thus, Additional Protocol I carved out an exception for the display of emblems by irregular forces when it acknowledged that ‘there are situations ... where, owing to the nature of the hostilities, an armed

⁹¹ Convention (III) Relative to the Treatment of Prisoners of War (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 135 art 4A(2).

⁹² Convention (IV) Respecting the Laws and Customs of War on Land (adopted 18 October 1907, entered into force 26 January 1910) 187 CTS 227, Annex on Regulations Concerning the Laws and Customs of War on Land art 1(2).

⁹³ I Gillich, ‘Illegally Evading Attribution? Russia’s Use of Unmarked Troops in Crimea and International Humanitarian Law’ (2015) 48 Vanderbilt Journal of Transnational Law 1215.

⁹⁴ Additional Protocol I (n 4) art 44(3).

⁹⁵ Additional Protocol I (n 4) art 44(7).

⁹⁶ Gillich (n 93) 1215.

⁹⁷ Hays Parks (n 89) 542.

⁹⁸ Geneva Convention III (n 91) art 4A(2).

combatant cannot ... distinguish himself⁹⁹ and accepted that such a person retains combatant status provided he 'carries his arms openly ... during each military engagement, and ... during such time as he is visible to the adversary ... preceding the launching of an attack'.¹⁰⁰ In other words, the requirement for a recognisable emblem is dropped altogether here owing to the exigencies of war. These areas of flex in the requirements of distinction are no doubt well founded, but they render the principle harder to apply.

The extent to which the application of distinction is complicated by the patchwork nature of the rules on uniforms, especially emblems, was writ large in the context of the Russian annexation of Crimea in 2014. This saw heavily armed Russian speaking individuals in military uniforms and military vehicles taking control of the peninsula while bearing no clear markings to display their allegiance. Observers knew they were Russian military, but the Russian government initially denied this calling them 'pro-Russian local self-defence forces'.¹⁰¹ Similarly, it was reported that the individuals identified themselves as 'Crimean self-defence forces'.¹⁰² It was only in March 2014, after the occupation of Crimea was effectively complete, that the Russian military started acting openly and, in April, that the Russian government explicitly acknowledged the allegiance of the personnel.¹⁰³ The events no doubt triggered the application of LOAC owing to the use of warning shots and the fact that force was used to blockade Ukrainian bases.¹⁰⁴ However, there was much discussion of whether Russia's unmarked military personnel—dubbed 'little green men' on account of their olive uniforms—had complied with the principle of distinction.¹⁰⁵

Following from what was said above, the consensus was that Russia's tactics were lawful, if not particularly 'sporting'. According to Grant and Huntley, 'the law does not require that the belligerent must be able to identify the nationality of the enemy belligerent, only that the enemy belligerent is distinguishable from the civilian non-combatant population (and his own forces) "at a distance"'.¹⁰⁶ For Reeves and Wallace, 'wearing a uniform with a Russian insignia is not an absolute requirement for the commandos to comply with the principle of distinction'.¹⁰⁷ For Gillich, 'neither treaty nor customary law provides for a legal obligation to disclose the nationality of the combatants (for example, by wearing nationality emblems)'.¹⁰⁸ This is because 'nothing in [LOAC] suggests that the principle of distinction is ... aimed at serving State interests (e.g., by

⁹⁹ Additional Protocol I (n 4) art 44(3).

¹⁰⁰ *ibid.*

¹⁰¹ S Walker, 'Russian Takeover of Crimea Will Not Descend into War, Says Vladimir Putin' *The Guardian* (4 March 2014) <<https://bit.ly/2UK8peT>>.

¹⁰² M Lipman, 'Putin's Crisis Spreads' *The New Yorker* (8 March 2014) <<https://bit.ly/2vsPf2u>>.

¹⁰³ R Heinsch, 'Conflict Classification in Ukraine: The Return of the "Proxy War"?' (2015) 91 *International Law Studies* 323, 328.

¹⁰⁴ Gillich (n 93) 1208.

¹⁰⁵ SR Reeves and D Wallace, 'The Combatant Status of the "Little Green Men" and Other Participants in the Ukraine Conflict' (2015) 91 *International Law Studies* 361, 394.

¹⁰⁶ Grant and Huntley (n 90) 594.

¹⁰⁷ Reeves and Wallace (n 105) 394.

¹⁰⁸ Gillich (n 93) 1213.

guaranteeing that combatants should clearly be linked to a specific party to the conflict’.¹⁰⁹ Finally, Hays Parks agrees that there is no common standard for uniforms and notes that LOAC ‘does not prohibit the wearing of a non-standard uniform [or even] the wearing of civilian clothing so long as military personnel distinguish themselves from the civilian population ... through a distinctive device, such as a hat, scarf, or armband, recognizable at a distance’.¹¹⁰ Indeed, it was thanks to these tell-tale signs that international observers knew that the little green men belonged to the Russian military prior to any official acknowledgement. For example, Human Rights Watch had observed that they used ‘Russian military vehicles and other equipment that Ukrainian forces are not known to have’.¹¹¹ No doubt this is how combatants first identify each other in practice, rather than by scrutinising for flags or other adornments, hence the use of unmarked troops is tolerated.¹¹²

In short, LOAC requires merely that combatants distinguish themselves from civilians. There is no requirement to wear emblems and distinction can be satisfied by wearing military-style uniforms (for which there are no set parameters) or even civilian clothing (provided some unspecified distinctive device is applied). For autonomous weapons, this means that ‘recognition’ in a conflict scenario would be a highly nuanced affair. One could not simply programme a machine to recognise a series of emblems, insignia, uniforms or camouflage patterns and thereafter target the wearers as presumed enemy combatants (subject to the further difficulties below on possible oscillations in status). Instead, it would be necessary to endow machines with a far more discerning palate. They must be taught that the absence of markings does not mean the absence of combatant status and trained to recognise a vast range of different apparel and military equipment. For full accuracy, it might even be necessary to upload intelligence information showing what individual enemy commanders’ and soldiers’ faces look like to be sure that the correct people are recognised as combatants. The question then becomes whether current technology could cope. It is to that issue that we now turn.

B. Machine Recognition and Facial Recognition Technology

Attempts to endow machines with the ability to recognise faces—which could be the gold standard for distinction by autonomous weapons—have a longer history than one might imagine. As Raviv explains, efforts in this field date back to just after World War II.¹¹³ He recounts the story of Woody Bledsoe, latterly a professor at the University of Texas at Austin, who pioneered

¹⁰⁹ *ibid.*

¹¹⁰ Hays Parks (n 89) 542.

¹¹¹ Human Rights Watch, ‘Questions and Answers: Russia, Ukraine, and International Humanitarian and Human Rights Law’ (Human Rights Watch, 21 March 2014) <<https://bit.ly/3bzcqJl>>.

¹¹² Gillich (n 93) 1212.

¹¹³ S Raviv, ‘The Secret History of Facial Recognition’ *Wired* (21 January 2020) <<https://bit.ly/2U4jYxa>>.

recognition technology. In the 1940s and 1950s, Bledsoe and his team developed a system that allowed machines to recognise visual images known as the ‘n-tuple’ method:

[The team] started by projecting a printed character—the letter Q, say—onto a rectangular grid of cells, resembling a sheet of graph paper. Then each cell was assigned a binary number according to whether it contained part of the character: Empty got a 0, populated got a 1. Then the cells were randomly grouped into ordered pairs, like sets of coordinates. ...With a few further mathematical manipulations, the computer was able to assign the character’s grid a unique score. When the computer encountered a new character, it simply compared that character’s grid with others in its database until it found the closest match.¹¹⁴

The breakthrough of the n-tuple method was that it allowed early computers to ‘recognise’ many variants of the same character. Thus, it represented the dawn of machine recognition. Years later, Bledsoe moved to academia and received research funding for facial recognition work—allegedly from the Central Intelligence Agency—and in 1967 a trial was devised. The system took 400 facial photographs and for each of them noted 46 coordinates: including five on each ear, seven on the nose and four on each eyebrow. The photographs themselves were then discarded and the co-ordinates manipulated to be front-facing and made to conform to scale. Then, a secondary facial photograph of one of the participants was fed in and the system was tasked with matching it with the correct principal. In terms of results, the team asked three people to cross-match subsets of 100 faces and ‘even the fastest one took six hours to finish [while] the [system] completed a similar task in about three minutes’.¹¹⁵ Clearly then, even in the late 1960s, machines were gaining the ability to recognise the three dimensional world—including the complex topography of the human face. Admittedly, the images involved were stills that required prior human manipulation, and they did not reflect the reality of a dynamic environment such as a war zone. Nonetheless, the seeds had been sown.

Facial recognition technology has attracted a lot of attention in the media recently, in large part due to its enthusiastic adoption by China as part of its population monitoring apparatus. The Uighur population of northwest China has perhaps been the group most affected. In 2017, a paper was published by scholars at Xinjiang University based on a database-orientated research project—funded by authorities including the National Science Foundation of China—which explicitly stated that its main purpose was ‘to provide the researchers a face database containing Uighur and Kazak faces to analyze the facial characteristics of the Uighur and Kazak people’.¹¹⁶ While no direct

¹¹⁴ *ibid*, para 16.

¹¹⁵ *ibid*, para 49.

¹¹⁶ H Zuo, L Wang and J Qin, ‘XJU1: A Chinese Ethnic Minorities Face Database’ (2017) IEEE <<https://bit.ly/37KhFDL>>.

connection can be made to this particular research project, we have since seen the large-scale internment of such individuals in ‘re-education’ centres within China with Byler noting that at least one million people have been affected since 2017.¹¹⁷ It is safe to say that facial recognition technology facilitated this mass collection of people. Indeed, as noted by Read and Walters, the Chinese government uses a database dubbed the digital Integrated Joint Operations Platform ‘that aggregates extreme amounts of data [from] multiple sources [including] CCTV cameras with facial recognition, existing Uighur legal records ... Wi-Fi scanning Systems [and] 31,000 convenience police stations in urban areas of Xinjiang’ for this purpose.¹¹⁸

More recently, the Chinese firm Hikvision developed a system linking cameras to artificial intelligence that has been trained on a huge database of images to categorise ‘new’ faces based on physical traits alone.¹¹⁹ The system simply identifies whether the face presented to it belongs to a person of an ‘ethnic minority’ or not. Note that, without having ever seen an individual before, this technology can categorise them. This represents an extension in recognition capability. Accompanying this, we have seen a growth in the geographical reach of the software with China beginning to export its know-how. In this regard, 2018 saw Guangzhou-based company CloudWalk (which has received around £200 million in Chinese central government sponsorship) agree to build a mass facial recognition program in Zimbabwe to monitor public spaces.¹²⁰ There are concerns it will be used there as it has been used against the Uighurs. As Byler put it, ‘the Uyghur homeland has become an incubator for China’s “terror capitalism”’.¹²¹

Today, facial recognition technology is ubiquitous in the global commercial sector. It is deployed as a convenient security feature for phones and laptops.¹²² It is even used in passports and by payment applications.¹²³ Soon, we are likely to see it rolled out to enable targeted advertising in shopping centres where characteristics such as age and gender are used to determine which advertisements are presented to which customers.¹²⁴ In terms of conflict scenarios, it is easy to see how this ability to recognise faces—either of specific individuals or of categories of people—might present an equally useful tool in the context of distinction and weapons targeting. The

¹¹⁷ D Byler, ‘Ghost World’ *Logic* (1 May 2019) para 3 <<https://bit.ly/2GsLAE7>>.

¹¹⁸ B Read and R Walters, ‘China: Do the Uighurs Represent a Serious Threat?’ (2019) James Madison University Scholarly Commons <<https://bit.ly/30ZI5Pb>>.

¹¹⁹ J Honovich, ‘Hikvision’s Minority Analytics’ *IPVM* (8 May 2018) <<https://bit.ly/2RyG6xZ>>.

¹²⁰ L Chutel, ‘China is Exporting Facial Recognition Software to Africa, Expanding its Vast Database’ *Quartz Africa* (25 May 2018) <<https://bit.ly/2RUSHL4>>.

¹²¹ Byler (n 117) para 17.

¹²² D Shang *et al.*, ‘Face and Lip-Reading Authentication System Based on Android Smart Phones’ (2019) *IEEE* <<https://bit.ly/2IC69z4>>.

¹²³ WK Zhang and MJ Kang, ‘Factors Affecting the Use of Facial-Recognition Payment: An Example of Chinese Consumers’ (2019) *IEEE* <<https://bit.ly/2TFDf09>>.

¹²⁴ T Yu *et al.*, ‘AI-based Targeted Advertising System’ (2019) 13 *Indonesian Journal of Electrical Engineering and Computer Science* (February) 787.

technology could be used to distinguish friend from foe in, for example, occupations where members of opposing sides often belong to different ethnic groups. Notably, Israel has shown keen interest in the technology in the context of its relationship with Palestine. Israeli firm AnyVision is at the forefronts of the efforts, with Holmes noting that the ‘technology is used by the Israeli military at border crossing checkpoints, where it logs the faces of Palestinians crossing into Israel’ and also that it is ‘secretly used ... throughout the West Bank ... to monitor the movement of Palestinian residents as part of ... efforts to prevent potential terror attacks’.¹²⁵ There is no suggestion that this technology has been used for the purposes of weapons targeting, least of all by an autonomous weapon. However, we can see machine recognition beginning to creep towards LOAC-governed space.

C. Machine Recognition in Security Guard Systems

Of course, machine recognition is not limited to faces which, in many ways, sit at the more complex end of the spectrum. It can be used more generally for the detection of symbols and objects in ways that, despite the legal ambiguities mentioned above, may still become useful in a military context. Patriot One’s PatScan was introduced earlier. As was explained, it operates essentially as an automated security guard using a range of sensors including video imagery taken from existing CCTV cameras. Naturally, in order to achieve any results, it is necessary for the system to ‘recognise’ the inputs it receives so that objects can be categorised and, if necessary, flagged as potentially dangerous. Indeed, as Maddox noted when quoting the CEO Martin Cronin, ‘when a weapon is present, whether overt or concealed, we can generate an alert. ... [as] we have algorithms that have been trained to recognize weapons from the signatures we get through ... video object recognition’.¹²⁶ The system is sophisticated in the sense that it allows not only for detection of weapons generally but, rather, for ‘the identification of specific weapon types’ with a promotional graphic showing that the system calculated that it had identified a semi-automatic assault rifle with 94.7 per cent certainty.¹²⁷ This nuanced information would be useful in the sort of conflict environments in which autonomous weapons would operate where, for example, it might be culturally normal for civilians to carry small arms but where bearing larger weapons might suggest clearer intent to cause harm and thus prompt a shift in the machine’s distinction analysis.

¹²⁵ A Holmes, ‘Microsoft Funded an Israeli Facial Recognition Startup Whose Tech Is Reportedly Being Used to Secretly Surveil Palestinians’ *Business Insider* (28 October 2019) <<https://bit.ly/2PZLYPB>>.

¹²⁶ T Maddox, ‘PatScan Platform Detects Hidden Weapons, Chemicals, and Bombs’ *TechRepublic* (10 January 2020) <<https://tek.io/2ldOFZB>>.

¹²⁷ Patriot One Technologies (n 75).

In order to achieve this sophisticated level of recognition, the PatScan system ‘leverages artificial intelligence machine learning technology’.¹²⁸ In other words, the system is fed thousands of images of weapons in different scenarios and told what they are so that, in time, it is able to recognise them independently. A benefit of this is that the system, once deployed, is not static but dynamic as it can continue to learn to recognise new threats and thus is able to provide ‘an ongoing ability to adapt as security threats evolve’.¹²⁹ Another string to PatScan’s bow is that it is not only able to recognise weapons, but can spot altercations too. The company demonstrates this capability in the context of a game of American football where a disturbance breaks out in the stands and is recognised by the system as a result of the increase in frequency and pace of people’s movements.¹³⁰ It is not hard to imagine this software being developed further and used by an autonomous weapon to detect the outbreak of violence in an armed conflict or occupation scenario. Patriot One is promoting its creation forcefully and argues that, because the system is run by software rather than relying on human visual acuity, it will be ‘faster, more accurate and more effective’ than humans.¹³¹ This will make it an attractive proposition from a military point of view whereupon any tactical edge is seized.

The PatScan system is just one example of machine recognition being developed in the commercial sector but with potential military applications. Indeed, machine recognition has grown to become an industry in itself with numerous companies involved in this lucrative area.¹³² For example, the Chinese firm Meiya Pico has developed a system that can detect Uighur language text and Islamic symbols embedded in images.¹³³ Visual capabilities in this context might be useful in war for the purpose of identifying an ‘enemy’ language on uniforms or military equipment where it would indicate that the person or object bearing that language is a legitimate target. Google Health has made significant advances in recognition, specifically in the context of identifying breast cancer from mammogram images.¹³⁴ This is in addition to earlier innovations in the context of, for example, skin¹³⁵ and lung¹³⁶ cancer. McKinney asserts that his team had identified a system ‘capable of surpassing human experts in breast cancer

¹²⁸ *ibid.*

¹²⁹ *ibid.*

¹³⁰ *ibid.*

¹³¹ *ibid.*

¹³² Nanalyze, ‘Watch for These 8 AI Startups Doing Computer Vision’ (Nanalyze, 13 March 2018) <<https://bit.ly/2UItO7W>>.

¹³³ P Li and C Cadell, ‘At Beijing Security Fair: An Arms Race for Surveillance Tech’ *Reuters* (28 May 2018) <<https://reut.rs/2RCJpuJ>>.

¹³⁴ N Eddy, ‘Google AI Platform Aids Oncologists in Breast Cancer Screenings’ *HealthcareITNews* (7 January 2020) <<https://bit.ly/2tZL7H1>>.

¹³⁵ A Esteva *et al.*, ‘Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks’ (2017) 542 *Nature* 115.

¹³⁶ D Ardila, ‘End-to-End Lung Cancer Screening with Three-Dimensional Deep Learning on Low-Dose Chest Computed Tomography’ (2019) 25 *Nature Medicine* 954.

prediction'.¹³⁷ More specifically, it was noted that the system generated 'an absolute reduction of 5.7 per cent and 1.2 per cent ... in false positives and 9.4 per cent and 2.7 per cent in false negatives' and that 'in an independent study of six radiologists, the AI system outperformed all of the human readers'.¹³⁸ It should be noted that the system is not perfect with for example, humans being slightly better than it at detecting '*in situ*' cancers (the system being better with 'invasive' cancers). These variances were used to justify developing 'complementary roles' in diagnosis for humans and machine recognition.¹³⁹ Still, the system showed the ability of machines to deal even with highly complex organic imagery in a way that may have military uses. Finally, Malong Technologies in Shenzhen undertook the WebVision challenge that involved classifying over two million pictures of retail products—including clothing, furniture, textiles and beverages—into one thousand categories. The company achieved performance 'on par with human beings on the same classification task'.¹⁴⁰ Specifically, it achieved 94.78 per cent accuracy where 'human performance has been measured between 94 per cent and 94.9 per cent'.¹⁴¹ Again, this sort of machine recognition could be carried over to the battlefield and used to identify items such as enemy uniforms, weapons and vehicles.

In summary, machine recognition has now advanced to a point where it has reached parity with human recognition abilities. Further, its security and military applications have not been lost on developers. Raytheon's interest in Exyn's 'mapping autonomous drone' was highlighted above and the company has even said that 'this system can help us identify the good guys and the bad guys so we can either rescue them or prevent our troops from being ambushed'.¹⁴² Despite all this, there is need for great caution in assuming that the ability to recognise faces, symbols, weapons, tumours and so on necessarily means that an artificial intelligence powered autonomous weapon could comply with the principle of distinction. This is because real battlefields are dynamic environments where a final element is required—judgement.

VI. THE INTERSECTION OF DISTINCTION AND AUTONOMOUS WEAPONS: MACHINE JUDGEMENT

A. The Need for Machine Judgement

As demonstrated above, machines have achieved capabilities in the contexts of 'observation' and 'recognition' that rival, and sometimes exceed, those of

¹³⁷ SM McKinney *et al.*, 'International Evaluation of an AI System for Breast Cancer Screening' (2020) 577 *Nature* 89, 89. ¹³⁸ *ibid.* ¹³⁹ *ibid.* 92.

¹⁴⁰ T Hu, 'China AI Startup Malong Technologies Wins WebVision Challenge' *PR Newswire* (27 July 2017) para 5 <<https://prn.to/2TCozpS>>. ¹⁴¹ *ibid.*, para 6.

¹⁴² *Airsoc* (n 84) para 11.

human beings. One might then assume that they would be equally as good as, or better, than humans at implementing the principle of distinction in LOAC. However, distinction requires one final step for its proper discharge, namely the application of 'judgement'. Judgement is necessary because the status of individuals (and indeed objects¹⁴³) in LOAC is not solely a matter of appearance: it is also a matter of context.

In terms of persons, context can be important for several reasons. It is quite possible for an individual to move from being a targetable combatant to a protected person, or *vice versa*, without any change in appearance. For example, a person might be dressed in an army combat uniform replete with camouflage, nationality and rank insignia and carrying a rifle in the middle of a warzone. From observation and recognition alone, that person would no doubt be classified as a 'combatant' and, if allegiant to the enemy, targetable. However, upon further analysis in the light of context, and with the attendant exercise of judgement, it may become clear that the person is not in fact a legitimate target. Perhaps the most obvious reason for this might be that the individual has become *hors de combat*. In the context of IAC, the position is explained in Additional Protocol I which states that a person is *hors de combat* if '(a) he is in the power of an adverse Party; (b) he clearly expresses an intention to surrender; or (c) he has been rendered unconscious or is otherwise incapacitated by wounds or sickness'.¹⁴⁴ As Reeves and Wallace summarised, combatants are a 'legitimate object of attack [only for]... as long as they are capable of fighting, willing to fight or resist capture'.¹⁴⁵ The protections in place for those who are *hors de combat* function because human participants in conflict are able to make logical judgements about contextual factors such as the raising of hands in the air (to indicate surrender) or the collapsed or disorientated appearance of a foe (indicating incapacitation). In other words, a human will recognise when a normally targetable enemy ceases to be targetable on account of circumstantial factors.

Equally, it is quite possible for an individual to move from being a protected person to a targetable combatant based on contextual factors. Civilians are defined negatively by LOAC such that anyone who is not a combatant for the purposes of Geneva Convention III is a civilian.¹⁴⁶ The regime attempts to further ensure protection of civilians by providing that, in cases of doubt, individuals are presumed to be civilians.¹⁴⁷ However, civilians may lose protection in certain circumstances. We can find an early example of this in the form of the '*levée en masse*'. Geneva Convention III made it clear that prisoner of war, and thus 'combatant', status extended to 'inhabitants of a non-occupied territory, who on the approach of the enemy spontaneously

¹⁴³ Additional Protocol I (n 4) art 52.

¹⁴⁵ Reeves and Wallace (n 105) 386.

¹⁴⁷ *ibid*, art 50(3).

¹⁴⁴ Additional Protocol I (n 4) art 41(2).

¹⁴⁶ Additional Protocol I (n 4) art 50(1).

[took] up arms to resist the invading forces, without having had time to form themselves into regular armed units'.¹⁴⁸

Today, the rather quaint notion of the *levée en masse* has been eclipsed by the modern concept of 'direct participation in hostilities' (DPH) whereby civilians get involved in conflict. As Melzer notes, the rise of DPH is a consequence of increased urban warfare and the 'physical proximity of combatants or fighters to civilians facilitate[es] the involvement of civilians in military operations from providing food, shelter, equipment, and intelligence to combatants, up to direct participation in combat'.¹⁴⁹ In terms of treaty law, the basic position is that 'civilians shall enjoy the protections afforded ... unless and for such time as they take a direct part in hostilities' (according to Additional Protocols I¹⁵⁰ and II¹⁵¹ for IAC and NIAC respectively). Thus, when civilians go so far as to directly participate in combat, they lose their protected status and become susceptible to targeting. The theory is straightforward; the practice is difficult. Consequently, the ICRC has devoted extensive effort to providing clarification in this complex area and now sets out three requirements for the assessment of DPH status. First, the putative participant must cross the relevant threshold for harm which can be done 'either by causing harm of a specifically military nature or by inflicting death, injury, or destruction on persons or objects protected against direct attack'.¹⁵² Second, the harm they cause must occur within 'one causal step' of the attack.¹⁵³ Third, there must be a sufficient 'belligerent nexus' between the action the individual has taken and the conflict with, for example, violent crimes not ordinarily amounting to DPH.¹⁵⁴ In other words, again, contextual factors—not simply the clothes a person is wearing or the symbols they do (or do not) display—have a significant bearing on status.

The International Criminal Tribunal for the Former Yugoslavia (ICTY) attempted in *Strugar* to provide an indicative list of what would qualify both as direct and indirect participation.¹⁵⁵ To an extent this list is useful but, in some ways, it opens up more questions than it answers. The full detail of the text is worthy of consideration:

Examples of active or direct participation in hostilities include: bearing, using or taking up arms, taking part in military or hostile acts, activities, conduct or operations, armed fighting or combat, participating in attacks against enemy personnel, property or equipment, transmitting military information for the immediate use of a belligerent, transporting weapons in proximity to combat operations, and serving as guards, intelligence agents, lookouts, or observers on

¹⁴⁸ Geneva Convention III (n 91) art 4A(6).

¹⁴⁹ N Melzer, 'The Principle of Distinction between Civilians and Combatants' in A Clapham and P Gaeta (eds), *The Oxford Handbook of International Law in Armed Conflict* (Oxford University Press 2014) 298.

¹⁵⁰ Additional Protocol I (n 4) art 51(3).

¹⁵¹ Additional Protocol II (n 5) art 13(3).

¹⁵² N Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law* (ICRC 2009) 47.

¹⁵³ *ibid* 53.

¹⁵⁴ *ibid* 58–64.

¹⁵⁵ *Prosecutor v Pavle Strugar* (Appeal Judgment), ICTY-01-42 (17 July 2008).

behalf of military forces. Examples of indirect participation in hostilities include: participating in activities in support of the war or military effort of one of the parties to the conflict, selling goods to one of the parties to the conflict, expressing sympathy for the cause of one of the parties to the conflict, failing to act to prevent an incursion by one of the parties to the conflict, accompanying and supplying food to one of the parties to the conflict, gathering and transmitting military information, transporting arms and munitions, and providing supplies, and providing specialist advice regarding the selection of military personnel, their training or the correct maintenance of the weapons.¹⁵⁶

Decrypting these rules and applying them in practice requires higher levels of judgement. Judgement would be needed, for example, to determine whether military information has been transmitted for 'immediate use' (which qualifies as DPH) or simply transmitted for use at some unspecified later date (which does not qualify as DPH). Similarly, judgement would be needed to decide whether weapons were transported 'in proximity to' combat operations (which qualifies as DPH) or further away from them (which does not qualify as DPH). In short, the DPH regime is now an essential component of the broader distinction of persons framework and it requires sophisticated—currently human—judgement for its proper application.

In summary, the capacity for 'observation' and 'recognition' alone is not enough for compliance with the principle of distinction—especially in the context of targeting persons on complex modern battlefields. Contextual factors have a significant, often decisive, bearing on status. While humans can generally exercise their judgement to accommodate such factors, the rise of fully autonomous weapons would necessarily leave these judgement calls to artificial intelligence. Is it up to the challenge?

B. Machine Judgement and Artificial General Intelligence

Perhaps the ideal solution to the problem of machine judgement would be the advent of so-called 'artificial general intelligence' (AGI) whereby machines are built with cognitive abilities equal to human beings. Those machines would be able to truly understand what is happening around them in the way that humans do. This intelligence would allow machines to appreciate the sort of contextual factors mentioned previously that can render a normally targetable combatant *hors de combat* or, alternatively, that can render a normally protected civilian targetable. Interestingly, the notion of AGI is beginning to pass from the realms of science fiction into earnest discourse. One adherent to AGI is Bostrom who, in 2014, predicted that artificial intelligence will eventually run the world and that this will be a matter of unparalleled consequence for humanity.¹⁵⁷ Tegmark, in his watershed 'Life 3.0', espoused a similar view in support of

¹⁵⁶ *ibid*, para 177.

¹⁵⁷ N Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press 2014).

the forthcoming and world-changing impact of AGI.¹⁵⁸ Scientists such as Kriegman are even beginning to develop organic machines that may one day possess enough ‘intelligence’ to allow them to operate independently inside the body to ‘seek out and digest toxic or waste products, or identify molecules of interest in environments physically inaccessible to robots’.¹⁵⁹

Some States even seem to be coming around to the idea of thinking robots. As we saw above, the UK has opted to set a very high bar when defining what would actually constitute an autonomous weapon by requiring that it would need to be capable of understanding ‘higher-level intent and direction’ (although, as alluded to earlier, this move may have been more about keeping the majority of autonomous weapons below the threshold of regulation).¹⁶⁰ Others are more sceptical. Sharkey, an academic and computer scientist who also leads the Campaign to Stop Killer Robots, argues against the development of autonomous weapons.¹⁶¹ However, he also believes that the discussion of AGI is over-egged and stated in a BBC interview that we are in an ‘AI autumn’ with developments in the field slowing down in the last couple of years.¹⁶² Of course, an ‘AI autumn’ is not an ‘AI winter’ and so even Sharkey acknowledges that advancements in this area have not stopped altogether.

It seems, then, that the question is increasingly becoming when, not if, AGI will come into existence. There is, of course, a multiplicity of views on this issue of timing. In an attempt to arrive at a consensus-based ‘best guess’ on this point, Muller and Bostrom surveyed hundreds of artificial intelligence experts at a series of conferences and asked, ‘by what year would you see a (10 per cent/ 50 per cent/90 per cent) probability for ... high level machine intelligence to exist?’ - The median response for 10 per cent probability was 2022, the median response for 50 per cent probability was 2040 and median response for 90 per cent probability was 2075.¹⁶³ In short, according to artificial intelligence experts taken as a whole, AGI is not likely to arrive for decades. One respected expert in the field, Walsh, was prepared to indicate a specific year by which he thinks machines will have achieved human-level cognition capability: 2062.¹⁶⁴ That number falls within the range identified in Muller and Bostrom’s survey and seem to the present author to be a sound estimate (bearing in mind the speculative nature of this exercise). So, if AGI is not going to be around for some time yet, what does that mean for the ability of

¹⁵⁸ Tegmark (n 39).

¹⁵⁹ S Kriegman *et al.*, ‘A Scalable Pipeline for Designing Reconfigurable Organisms’ (2020) *Proceedings of the National Academy of Sciences* 5 <<https://bit.ly/36VCufl>>.

¹⁶⁰ United Kingdom (n 30) para 4.

¹⁶¹ N Sharkey, ‘The Evitability of Autonomous Robot Warfare’ (2013) 94 *International Review of the Red Cross (New Technologies and Warfare)* 787.

¹⁶² S Shead ‘Researchers: Are We on the Cusp of an “AI Winter”?’ BBC News Online (12 January 2020) <<https://bbc.in/38jFFgT>>.

¹⁶³ VC Muller and N Bostrom, ‘Future Progress in Artificial Intelligence: A Survey of Expert Opinion’ in VC Muller (ed), *Fundamental Issues of Artificial Intelligence* (Springer 2016).

¹⁶⁴ T Walsh, 2062: *The World that AI Made* (La Trobe University Press 2018).

autonomous weapons to exercise the level of judgement necessary to properly implement distinction? Certainly, it does not rule it out. This is because, while full AGI might be distant, artificial intelligence can still be very capable provided it is only required to operate in a limited field. It is instructive, therefore, to consider a cross-section of existing systems that demonstrate such capabilities as these may in time be adapted for use in autonomous weapons.

C. Machine Judgement and Computer Games

DeepMind was referred to above in the context of its attempts to create a machine that could play computer games using visual input alone. It was noted that a system was created that could indeed ‘see’—albeit a simplified world. However, in addition to enabling the system to see, DeepMind claimed that it had allowed the system effectively to think by having ‘introduced a new deep learning model for reinforcement learning’.¹⁶⁵ This claim was demonstrated in 2013 when, as has been seen, the system was able to play certain computer games well enough to beat humans. This limited-scope intelligence capability was achieved by creating a ‘neural network’ to process information. In essence, the neural network functioned ‘by evaluating each image and assessing how it will change given any of the possible actions ... based on its experience of the past’.¹⁶⁶ The details of how the system works are proprietary and thus secret. Nonetheless, it demonstrated sufficient intelligence to learn not only how to play the games, but to win.

Perhaps, then, DeepMind’s neural network is indicative of the sort of intelligence that could be leveraged in the context of autonomous weapons to allow them to learn how to navigate successfully the rules of LOAC while still being able to ‘win’ in the sense of achieving their objectives. Of course, the learning would have to occur in simulated scenarios rather than on real battlefields as enemy combatants cannot be used as guinea pigs for the development of a system. While the neural network may indeed represent such a starting point, it was conceded by the DeepMind developers that their system was not able to beat humans at all games. For example, in more complex games such as Q*bert, Seaquest and Space Invaders the human players proved superior.¹⁶⁷ Furthermore, it must be noted that the ‘successes’ were against a backdrop in which ‘at any instant in time during a game, a player can choose from a *finite* set actions that the game allows: move to the left, move to the right, fire and so on [emphasis added]’.¹⁶⁸ This is very much more limited than the virtually *infinite* range of actions that may be undertaken in the real world.

¹⁶⁵ Mnih *et al.* (n 58) 8.

¹⁶⁶ The Physics arXiv Blog (n 55) para 11.

¹⁶⁷ *ibid.*, para 16.

¹⁶⁸ *ibid.*, para 6.

Another limitation of DeepMind's system is that, in the gaming context, 'the task for any player—human or otherwise—is to choose an action at each point in the game that maximises the eventual score'.¹⁶⁹ Again, this does not reflect the reality of armed conflict where there is not merely one objective but, rather, multiple competing considerations. The main examples of this are the diametrically opposed concepts of military necessity and humanity which operate as the pillars of LOAC.¹⁷⁰ In defence of DeepMind's system, its task was complicated by the fact that 'the reward from any given action [in a game] is not always immediately apparent [as], for example, taking cover from a space invader's bomb does not increase the score but does allow it to increase later'.¹⁷¹ Therefore, there was a degree of intelligence required in the sense that the machine had to attempt different tactics and recall which one ultimately, not just immediately, generated a higher score. Nonetheless, the gaming context remains drastically more straightforward than the real world with its infinite range of scenarios and numerous, often contradictory, objectives. Thus, it seems clear that limited gaming intelligence is not ready for the battlefield just yet.

The same conclusion seems to hold true even for DeepMind's more recent, and much more advanced, efforts in this area represented by AlphaGo—which has been able to beat human champions at the ancient Chinese strategy game 'Go'. Go is vastly more complex than games like Pong, or even chess, with more potential moves than there are atoms in the universe. Again though, AlphaGo was able to win by applying deep learning in neural networks. As Gibney explained, these are 'brain-inspired programs in which connections between layers of simulated neurons are strengthened through examples and experience'.¹⁷² In terms of process, Gibney proceeded to note that the system 'studied 30 million positions from expert games, gleaning abstract information on the state of play from board data ... then it played against itself across 50 computers, improving with each iteration, a technique known as reinforcement learning'.¹⁷³ The achievement of AlphaGo is staggering as, to an extent, it provides evidence that machine judgement can be trained to operate in an environment of infinite possibilities where number-crunching alone cannot be used to determine the optimal outcome. However, it is still limited in the sense of having a sole objective: surrounding a larger total area of the board with its stones than the opponent.¹⁷⁴ It was not required to consider nuanced competing objectives.

D. Machine Judgement in Agriculture

Moving away from computer games, developers have been attempting to leverage machine judgement in the context of agriculture to meet similar

¹⁶⁹ *ibid.*, para 6.

¹⁷⁰ Winter (n 7).

¹⁷¹ The Physics arXiv Blog (n 55) para 7.

¹⁷² E Gibney, 'Google AI Algorithm Masters Ancient Game of Go' (2016) 529 *Nature* 445, 445.

¹⁷³ *ibid.* 446.

¹⁷⁴ *ibid.*

challenges to those that might be faced by autonomous weapons. The Danish agricultural technology company 'Agrointelli' develops new technologies to make plant production more profitable and describes itself as 'working within the realms of navigation, automation and vision'.¹⁷⁵ One of its research projects is known as 'RoboWeedMaPS' and its aim is to combine deep learning and big data for use in autonomous farming machines that patrol cultivated fields with a view to removing weeds while, simultaneously, leaving crops undisturbed.¹⁷⁶ The system operates on broadly the same framework as that which underpins this article and so there are observation, recognition and judgement components to it. The developers report that they have made advances in this area by enabling the recognition and mapping of more than 100 different weeds.¹⁷⁷ Indeed, the company has produced an image on its website showing an aerial picture of a small patch of farmland with different types of vegetation growing on it and with red boxes drawn around various sprouts of greenery that the system has identified as weeds to be sprayed with herbicides.¹⁷⁸ Thus, despite its very different purpose, the system operates much like a rudimentary autonomous weapon.

However, again, this capability is based on algorithms resulting from big data. The project team fed thousands of images of weeds into a database and then used deep learning to teach a computer to recognise different types of weeds through the use of those images. As a result of the reliance on these techniques, rather than perhaps AGI, the same sort of 'contextual' problems we saw above emerge. Weeds grow in the real world and so they do not have a set appearance. What may appear not to be a weed one day, because of contextual factors such as colouring or partial covering, may prove to be one the next. After all, the appearance of weeds can change. As a senior researcher on the project, Jorgensen, noted: 'it only takes a small beetle to eat a leaf and the plant doesn't look like the one in the image at all [or] the stems can be so thin that ... it looks as though the leaves aren't connected [or] if it's cold in spring, some weeds turn completely purple even though they're normally green'.¹⁷⁹ The result of the difficulties posed by contextual factors is that the system, impressive as it is, operates with a margin of error. Some plants which ought to be judged as weeds are judged as crops and *vice versa*. This may be acceptable in the context of food production, but when humans are involved inaccuracy becomes intolerable. Accurate contextual sensitivity is essential to fully informed judgement.

¹⁷⁵ Agrointelli, 'Our Company' (Agrointelli) <<https://bit.ly/3arI8qJ>>.

¹⁷⁶ Agrointelli, 'RoboWeedMaPS' (Agrointelli) <<https://bit.ly/3dAkMkZ>>.

¹⁷⁷ *ibid*, para 1.

¹⁷⁸ Agrointelli (n 176).

¹⁷⁹ RN Jorgensen, 'RoboWeedMaPS: How Deep Learning Can Help Farmers Get Rid of Weeds' Aarhus University Department of Engineering (14 January 2019) <<https://bit.ly/30aI4rm>>.

E. Machine Judgement in Healthcare

From computer games to agriculture and finally to healthcare. Here too we are seeing software developers imparting machine judgement into systems—this time to allow them to take some burden away from human clinicians or to double check their work. Google Health’s mammography artificial intelligence project was discussed above and it was noted that the system is capable of surpassing human detection rates for certain cancers.¹⁸⁰ Again, underlying the technology was a deep learning system or, more properly ‘an ensemble of three deep learning models’ absorbing the prior judgements of humans on thousands of images: these were used as a bank of knowledge which the system could draw upon and compare to new images in order to make determinations.¹⁸¹ The result of this dependence on machine learning was that the usual limitations emerged. There is no novel thought involved here; rather lots of raw processing to enable comparisons. If the data fed into the system is wrong, so will be the system’s determinations. The system cannot rely on any real ‘intelligence’ to depart from what it has ‘learned’.

For this reason perhaps, while the research team considered the potential clinical applications of the technology, there was no suggestion that machine judgement in this life-or-death context would be the sole decision maker. Rather, it was suggested that the system could be used in the UK to adapt the current system. Presently, two human ‘readers’ check mammogram scans and deliver their judgements on the presence of any cancerous tissue—with a third reader becoming involved in the case of disagreement between them and issuing a casting vote. The researchers propose that the new system could be used to replace one of those initial readers—though retaining the third (human) reader in cases of disagreements—arguing that this would ‘reduce the workload in hospitals and clinics by obviating the need for double reading in 88 per cent of UK screening cases’ while at the same time ‘preserving the standard of care’.¹⁸² In other words, while machine judgement may have a role in complementing human judgement, it should not be used to replace it. The same conclusion is surely true in the equally life-or-death context of autonomous weapons operating on the battlefield.

In sum, while there is no doubt that artificial intelligence has made great strides in recent years, ‘machine judgement’ remains very limited. This is because every system of machine judgement designed to date has required the system to achieve one sole—albeit sometimes complex—objective by marshalling big data and deep learning. That objective might be winning a game, removing weeds or detecting cancer. However, machines have not yet been challenged to achieve multiple, contradictory, objectives simultaneously. This is important because, without these competing objectives, there is no need for novel judgement or the consideration of contextual factors. Nonetheless, as was shown above, context is

¹⁸⁰ McKinney *et al.* (n 137) 89.

¹⁸¹ *ibid* 96.

¹⁸² *ibid* 91.

absolutely fundamental in more complex exercises such as the implementation of distinction in an environment as dynamic as a warzone. To date, the sort of judgement and cognitive dissonance required for this exercise is reserved solely to humans. Humans will therefore need to remain involved in any autonomous systems where numerous competing objectives are at play so that contextual shifts can be accounted for in judgements. As Airbus said in the context of a project to build an autonomous take-off system, despite advances in machine observation and artificial intelligence, 'pilots will remain at the heart of operations'.¹⁸³

VII. CONCLUSION

In conclusion, it has been seen that machines have come to rival, and perhaps even surpass, humans in the context of observation and recognition. However, when it comes to judgement, they remain inferior as they rely on machine learning and big data rather than genuine understanding. They can mimic decisions that have come before, but they are not yet able to account for context by balancing contradictory objectives such as humanity and military necessity in the manner required to discharge distinction in a complex and dynamic warzone. That said, there has been an exponential growth in the capabilities of machine intelligence in areas such as computer games, security guard systems, agriculture, healthcare and beyond. It seems inevitable that continued investment in these areas will render increasingly capable systems and that these will likely become involved in the 'critical decisions' of autonomous weapons in the long run. This might be through the advent of AGI or some other technological watershed. No matter how it arises, fully fledged machine judgement seems set to arrive in the coming decades and, when it does, so too will the prospect of distinction-compliant autonomous weapons. However, until that day arrives, fully autonomous weapons could not comply with distinction and so their lethal use in combat operations would be unlawful.

¹⁸³ Airbus (n 72) para 6.

Copyright © The Author(s) 2020. Published by
Cambridge University Press for the British Institute of
International and Comparative Law