Sunlight Glinting on Clouds: Deception and Autonomous Weapons Systems

Amanda Sharkey and Noel Sharkey

“All warfare is based on deception” Sun Tzu, 500 BC.

Abstract The art of deception has played a significant role in military conflict for centuries and has been discussed extensively. Yet there has been an absence in the literature of any scrutiny of the risks posed by the deception of Autonomous Weapons Systems (AWS). After explaining the nature of AWS, we overview reasons given in their favour and arguments against them. Examples of military deceptive strategies are considered, together with reflections on the nature of deception. The core of the paper is a technical examination of some of the ways that AWS could be deceived and the potential humanitarian consequences. Since AWS have, by definition, an absence of meaningful human control, any deception could remain hidden until too late. We conclude that awareness of the vulnerability of sensing and image processing systems of AWS to deception reinforces and strengthens the case against their development and use.

1 Introduction

Sun Tzu’s influential ‘The Art of War’, written in 500 BC and still taught in today’s military academies, stated that ‘all warfare is based on deception’ [1, p. 10]. This is supported by a long history of deception in armed conflict that shows it to be a key to achieving victory. Famous examples include Hannibal’s deception of Flaminius the Roman Consul; luring his troops into ambush [2]. The Confederate army used decoy cannons made from tree trunks [3]. In World War 2, deception about the presence of a (non-existent) superior force waiting outside a harbour led to the Germans sinking their own battleship, the Admiral Graf Spee, in 1939 [4]. In 1944, the US deployed a ‘ghost army’ of 1100 men including artists, architects and set designers to successfully impersonate other Allied Army units, using inflatable tanks, cannons,

A. Sharkey (✉) · N. Sharkey
University of Sheffield, Sheffield, UK
e-mail: a.sharkey@sheffield.ac.uk

© The Author(s) 2021

A. Henschke et al. (eds.), *Counter-Terrorism, Ethics and Technology*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-030-90221-6_3

35

aeroplanes, and both sonic and radio deception. They participated in many operations, impersonating various, larger US units [5]. The D-day invasion of Normandy relied on many deceptions by the Allies about the expected landing location—double agents and other deceptive strategies strengthened the German belief that it would occur at Pas-de-Calais and in Norway [6].

Our question here is, how does deception fit into the ongoing technological transformation of warfare where ever more control of weapons is being ceded to computer systems? In particular, we explore the risks posed by deception to the deployment of Autonomous Weapons Systems. Despite the strong links between deception, deceptive strategies and military operations, there has been little or no discussion about deception and weapons that are entirely controlled by computer algorithms. We begin with an account of Autonomous Weapons Systems (AWS), some of the driving forces in their favour, and the main arguments against them. This is followed by an overview of military uses of deception and deceptive strategies and some reflection on what counts as deception. We then turn to a consideration of how AWS are likely to impact on and be affected by deception in armed conflict and counter-terrorism.

2 Autonomous Weapons Systems

Autonomous weapons systems represent a comparatively recent development. The US Department of Defence defines them as weapons that are able, ‘once activated, to select and engage targets without further intervention by a human operator’, [7, updated 2017]. Human Rights Watch [8] defines them as weapons which ‘would identify and fire on targets without meaningful human control’, and the International Committee of the Red Cross defines them as ‘weapons that can independently select and attack targets’ [9].

Since 2013 there have been vigorous attempts at the United Nations Convention for Certain Conventional Weapons (CCW) to create a new international legal instrument to prohibit the development and use of AWS [10, 11]. In 2016, the CCW agreed to establish a Group of Governmental Experts (GGE) to discuss autonomous weapons. Yet superpowers and other major military powers seem to be embarking on their development in the belief that they will create a military edge [12]. China, Russian, Israel and the US wish to use them for force multiplication with very few human controllers operating swarms of weapons in the air, on the land as well as on and under the sea [11].

Major military powers advocate the use of AWS for armed conflict and counter-terrorism for They hold that (i) AWS could complete missions in environments where communication signals could be jammed or disrupted (e.g. [13]), (ii) armed conflict is becoming too fast for human decision making [12], (iii) they could reduce risks for military personnel (e.g. [14]), (iv) they could increase target accuracy and reduce risks for civilians [15].

All four reasons are problematic. In order (i) means that there is no possibility of human oversight to disengage from erroneous targets; (ii) the speed of armed conflict

will simply get faster in a speed race if AWS are used; (iii) they will not reduce the risks to military personnel if the enemy also has them; (iv) while they may increase target accuracy, this does not solve the problems of target selection—the wrong targets could be accurately killed.

More than 140 humanitarian disarmament organisations from 60 countries, clustered under the banner of the Campaign to Stop Killer Robots, are concerned about the harm that such weapons would have on civilian populations (<https://www.stopkillerrobots.org> consulted April 15 2021). Their concerns are expressed in a plethora of arguments that we can break down into 4 major classes: (i) *Non-compliance with International Humanitarian Law (IHL)*: AWS cannot be guaranteed to comply with IHL; (ii) *Immoral delegation of kill decision*: delegating the decision to kill to a machine is immoral (iii) *Global security*: the widespread use of weapons outside of human control would destabilize global security; (iv) *Algorithmic injustice*: algorithmic bias and injustice: the widespread use of algorithms in civil society have shown decision biases against women, ethnic minorities and people of colour and are resulting in many legal challenges. We add to these arguments with an analysis of how susceptible AWS are to deceptive military strategies. Since AWS operate without meaningful human control, they could be subject to deceptions that a human might have detected. They can also, as explained, be deceived by visual manipulations undetectable to the human eye.

3 Arguments Against the Use of Autonomous Weapons Systems

International Humanitarian Law (IHL), sometimes termed the Laws of War, is intended to protect civilians. The principles of distinction, proportionality and military necessity are crucial aspects of IHL. Arguments about the inability of AWS to always comply with IHL have been made by Noel Sharkey [11, 16]—see (i) above. The principle of distinction in IHL requires weapons systems to distinguish combatants from non-combatants and other immune actors. Sharkey [16] argues that AWS lack three necessary components for this. First, their sensory and vision systems are not able to reliably discriminate between combatants, non-combatants and other immune actors such as wounded combatants. Second, there is no codified or programmable definition of what constitutes a civilian or non-combatant. And third, AWS lack the necessary situational and battlefield awareness. For instance, a human could draw on their understanding of social situations to recognise insurgents burying their dead in a way that AWS could not.

Sharkey [16] also claims that the principles of proportionality and military necessity are beyond the capabilities of present and near future weapons systems. Some proportionality problems are relatively easy to solve, and some are much harder. The easier proportionality problems involve calculations such as working out the likely collateral damage of different forms of attack and minimising such damage. For

instance, AWS software could choose the munitions to be used near a school so as to minimise the number of children killed. Hard proportionality problems are those which involve decisions about military advantage and military necessity—in the school example, deciding whether the military advantage to be gained would justify the use of any form of attack near a school. Such decisions require, ‘responsible accountable human commanders, who can weigh the options based on experience and situational awareness’ [16]. Suchman [17] has also argued that machines cannot fulfil the requirement of situational awareness. And Sharkey [18] has argued that it is not possible to program or train computational devices to develop the necessary moral competence to make such decisions.

Other writers such as former UN special rapporteur Heyns [19] and Asaro [20] have focused more on the moral argument, (see (ii) above), and argued that irrespective of what AWS and Artificial Intelligence might be able to do in the future, there are important arguments to make about what they should and should not do. For Heyns, AWS *should not* be used to target humans because their use would be an offence against the right to life. Heyns argues that errors would be made and there would be no person to be held accountable. He also claims that the lack of human deliberation would render targeting decisions as arbitrary and against the right to dignity of those targeted and of those in whose name the force was deployed (see also [21]). Asaro [20] argues that AWS *should not* be used even if they were able to meet the requirements of international humanitarian law. For him, IHL and the principles of distinction, proportionality and military necessity, imply a requirement for *human* judgement, and a duty not to delegate the capability to initiate the use of lethal force to unsupervised machines or automated processes.

As well as concerns about the extent to which AWS can conform to IHL, the third set of arguments against them, (iii above), is that they will destabilise global security. Although it is sometimes claimed that AWS could result in more accurate targeting and freedom from human self-preservation concerns, Tamburrini [22] articulates what he terms a ‘wide consequentialist view’ that AWS will threaten global security. He agrees with Sharkey [23] that by reducing the risks of a ‘body bag count’, a major disincentive for war would be removed. Tamburrini [22] also argues that swarms of AWS could weaken traditional nuclear deterrent factors by means of the threat of destructive attacks on strategic nuclear sites that could eliminate an opponent’s second-strike nuclear capabilities, thereby increasing preferences for first strike capabilities. Amoroso and Tamburrini [24] point out that even using AWS in a non-lethal manner to destroy buildings or infrastructure could have a global destabilising effect. Sharkey [16] also highlights concerns about global security due to an increase in the pace of war as a result of deploying AWS. In addition, he emphasises the likelihood of unpredictable interactions between different computational algorithms.

The fourth class of arguments against AWS, (iv above)—strongly connected to the first, and the inability to conform to IHL, are related to problems with the widespread use of algorithms in civil society and the demonstrated biases [25]. It is sometimes suggested that AWS could be used to pick out specific people, or classes of people, as legitimate targets of attack. However, it has become increasingly clear that decisions made about people using algorithms are frequently biased [26]. This is often the

result of problems with big data being used to ‘train’ machine learning systems. One problem is there has been a consistent failure to find a way to eliminate bias in the data. Moreover, machine learning algorithms are adaptive filters that smooth out the effects of outliers in the data, but these outliers could be minority groups that the trained system will consequently be less able to recognise. The resulting decision algorithms also lack transparency, since learning results in large matrices of numbers that are then used to generate the decisions. In civil society, this has proved to create bias in many domains such as juridical decisions, policing, mortgage loans, passport application and short-listing for jobs.

It is difficult to see how such algorithms could possibly be considered for a role in making decisions about who to target with automated weapons systems. This poses a particular problem for their use in counter-terrorism activities and border control. For example, face recognition algorithms are good at recognising the white males that form the majority of their training data and much worse at recognising black and female faces [27]. It is unlikely that the data used to train weapons systems to recognise particular individuals, or classes of people, will be trained with sufficiently representative sets of data to eliminate the problems of racial, gender and cultural biases.

4 Deception in Armed Conflict

In the US Army’s FM-90 manual, [28] there is clear recognition of the need for deception to achieve operational advantage. It provides an account of 10 maxims to be followed, which includes, ‘Cry-Wolf’ and ‘Magruder’s principles’. The ‘Cry-Wolf’ maxim represents the idea of desensitising the enemy to the likelihood of attack by repeated false alarms. For instance, in the week before the Pearl Harbour attack there had been 7 reports of Japanese submarines in the area, all of which turned out to be false. ‘Magruder’s principles’ refer to the exploitation of existing perceptions, and the notion that it is easier to strengthen an existing belief than to create a new one. In the D-day invasion, Hitler and his advisers were known to expect invasion in the Pas-de-Calais region, and efforts were made to strengthen this expectation.

The Joint-Publication JP3-13.4 [29] of the US forces identifies four basic techniques of military deception: (1) Feints (offensive action conducted to deceive the adversary about the location and or time of the main offensive action); (2) Demonstrations (a show of force without adversarial contact); (3) Ruses (deliberately exposing false or confusing information for collection by the adversary); and (4) Displays (simulation or disguising of capabilities which may not exist). As also discussed in that publication, military deception can involve electronic warfare. Electronic warfare has three major subdivisions: electronic attack (EA), electronic protection (EP) and electronic warfare support (ES). Camouflage and concealment are distinguished from military deception in JP3-13.4, although described as being able to support it by providing protection for activities.

Guides to deception for the military also refer to forms of deception that are against the laws of war, termed ‘acts of perfidy’, (Article 37 (1) Additional Protocol 1, [30]). Ruses of war; ‘acts which are intended to mislead an adversary or to induce him to act recklessly’, are not prohibited (Article 37 (2) Additional Protocol 1, [30]). Prohibited acts of perfidy are deceptions that lead the enemy to the false belief that they are entitled to, or are obliged to accord, protected status under the law of armed conflict. They include the use of vehicles marked with a red cross or a red crescent to carry armed combatants, weapons or ammunition, and the use in combat of false flags, insignia or uniforms. They are against the laws of war because they undermine the effectiveness of protective signals and jeopardise the safety of civilians and non-combatants.

Although some military guides distinguish between forms of deception, and camouflage and concealment, other accounts categorise camouflage as a form of deception [31]. Discussions of deception often hinge on the intentional deception of a human mind, or minds. In terms of the military, this might be the mind of the commander in charge of operations, or it might be the minds and perceptions of combatants on the ground. A slightly different account of deception arises when considering how AWS might be deceived.

The absence of human control of AWS necessitates a changed perspective on the notion of deception that has not yet made its way into military manuals. If, for instance, the sensors and programs of the autonomous weapons were subjected to deceptive strategies and, as a result, were to attack ‘friendly’ targets, or to plunge into the sea, this would not represent a deception of the human mind in a direct sense. As autonomous weapons, after they had been activated, they would have selected and attacked the targets without any human intervention. Should examples of AWS being disrupted in this way be described as deception? To answer this question, we need to re-examine what is meant by ‘deception’.

As we have already seen, the emphasis in military manuals and guides to deception of adversaries is on misleading the *mind or minds* of the enemy. But in the AWS examples above, human minds are not directly deceived. They could be said to be *indirectly* deceived, in that the operational commander’s intended target may not have been hit. But this is not the more straightforward version of deception assumed in the manuals. Is it still appropriate to use the term ‘deception’ here?

4.1 So, What Is Deception and Could a Weapon Be Deceived?

Some definitions of deception require a person to have been deceived, and also that intention is involved. For instance, Carson [32] defines deception as ‘intentionally causing someone to have false beliefs.’ For him, deception can be distinguished from lying because deception implies success and that someone has been successfully caused to have false beliefs. A person who lies may not deceive the person to whom they lie. Zuckerman et al. [33], in a psychological investigation of deceptive communication, also define deception as requiring a human to have been deceived: deception

is ‘an act that is intended to foster in another person a belief or understanding which the deceiver considers to be false’.

Intentional deception can be undertaken with the aim of benefitting the deceiver. As well as examples of military deceptive strategies, there are others such as internet scams, or phishing attempts to gain information about someone’s bank details. Of course, it is also possible that a person or persons might intentionally deceive others with the aim of helping or improving their quality of life. Bok [34] gives several examples of deceptions created with good intentions, including placebos, and white lies.

Deception can also occur without intention as we have argued elsewhere [35]. Bok [34] points out various situations in which people might deceive without having intended to do so. They might deceive others by conveying false information in the belief that it is true. Deception also arises without intention in the natural world. In such cases, it is usually to the benefit of the deceiver. Bond and Robinson [36] define deception in the natural world as ‘a false communication that tends to benefit the communicator’. Examples include camouflage, mimicry, death feigning and distraction displays. Camouflage can make creatures less visible to their predators. Mimicry is used, for instance, by the edible viceroy butterfly which has the markings of the inedible monarch butterfly, and by the brood mimicry of the cuckoo. Death feigning as an anti-predator adaptation occurs in a range of animals, and distraction displays to draw attention away from nests and young are found in birds and fish. As Gerwehr and Glenn [37] point out, deception is used in the natural world ‘both to acquire dinner and to avoid becoming dinner’.

AWS could be used intentionally by humans in deceptive strategies. And, of course, programmed (or trained) computer algorithms do not have minds and thus cannot by themselves form an intention to deceive. But although they cannot intentionally deceive, they might be disrupted by deceptive strategies. We choose to describe AWS here as being deceived, despite our uneasiness about possible anthropomorphic language. In the present context, it is useful to use the term ‘deception’ as a shorthand to describe the situations in which the operations of AWS are disrupted, by either intentional or unintentional deception, and by deceptive strategies. Moreover, by saying that AWS can be deceived, we by no means wish to imply that they can be held responsible or accountable for their operations. At all times, responsibility for the behaviour of weapons rests with the humans who develop and use them (see e.g. [38]).

5 Deception and AWS

There are various ways in which AWS could be used to create a deception in the sense of some of the examples of military strategy described earlier. AWS are autonomous once launched, but the military can still be involved in decisions about when and where to deploy them. This is the case even if the weapons are set up to automatically launch when incoming missiles are detected—a decision has still been made by

humans to set them up in this way. AWS could be launched in an area as a *feint*: mounting an attack in an area to distract an adversary from an attack being prepared elsewhere. They could be launched as a *demonstration*, or show of force, attacking buildings or locations in order to create the impression of technological superiority. Of course, terrorists, non-state actors, and insurgents could also make use of AWS in a similar manner.

The more serious humanitarian risk is that deceptive strategies could be used against AWS to disrupt the behaviour of the machines. Humans are endlessly inventive and creative, and there is little reason to expect that the human targets of AWS will passively wait to be killed. Terrorists, insurgents, non-uniformed combatants and non-state actors are going to invent ways of deceiving and derailing AWS. Johnson [39] details the many adaptations and innovations of the Taliban in Afghanistan in the asymmetric warfare conducted there. Al Qaeda are known to make use of denial and deception strategies [40], and in 2013 Al Qaeda counter-drone manuals were discovered in Mali, detailing 22 steps for avoiding drone attacks [41]. Bolton [42] states ‘People are too messy, unpredictable, clever and tricky to meet the assumptions programmed into military technology’. He gives as examples the ways that Vietnamese communist soldiers spoofed the electronic detectors dropped from US warplanes onto their paths through the jungle: ‘they sent animals down the trail, placed bags of urine next to so-called “people sniffers”, and played tapes of vehicle noises next to microphones—prompting computerized bombers to unload explosives onto phantom guerrillas’.

It is easy to underestimate the technological ingenuity of low-tech actors. A good example was the US military capture of Shia militants with laptops containing many hours of video footage taken from US drones. They had used software called Skygrabber, available on the internet for \$26 dollars, for downloading music and video [43].

Hezbollah carried out similar operations against Israeli forces as far back as 1996 when they used photographic evidence taken from an Israeli drone of an Israeli attack. Hezbollah also claimed that they had used analyses of Israeli drone footage to plan ambushes, such as the “Shayetet catastrophe” in which 12 Israeli commandos were killed (the method they used to hack the drone signals remains unknown) [44].

An important motivation, for those likely to be subject to an attack from AWS, would be to cause them to select targets that reduced the risk of harm to either combatants or civilians. For instance, if they were deceived into attacking dummy buildings instead of military installations, expensive fire power could be drawn and exhausted. Similarly, it would be advantageous to find ways of camouflaging military targets from sensors such that they were shielded from attack. A deception that caused AWS to target neutral, or protected installations such as hospitals could create an effective public relations coup for a terrorist group (although it is not clear who should or would be held responsible in such a case). Of course, some forms of deception could have unwanted humanitarian consequences. For instance, if it was known that AWS were programmed to attack vehicles with the heat signature of tanks, efforts could be made to modify the tanks’ heat signature to resemble that of buses or lorries. But the unwanted consequence of this could be a subsequent modification to the AWS

sensors so that they targeted vehicles with the signature of buses or lorries, leading to wider devastation. This would be an example of the ‘monkey’s paw’ effect discussed in military accounts of deception, whereby seemingly effective deceptions result in unintended harmful side effects.

How could AWS be subject to deception? Once launched, they are dependent on their sensors and image recognition systems to detect the targets they have been programmed to attack. These are unlike human sensing systems and can be disrupted in ways that humans cannot even sense such as by means of high frequency sounds, bright lights, 2D images or even small dots that are entirely meaningless to us (see e.g. [45]).

There is growing awareness of the limitations of image recognition systems [46] and the risks that they could be unintentionally deceived or mistaken. For instance, problems with the sensors and image recognitions systems of autonomous cars have resulted in several Tesla crashes. Known objects in unexpected positions, such as a motorcycle lying on the ground, may not be recognised [47]. Self-driving cars and their sensors and software are known to have difficulties with rainy and snowy conditions [48]. In 1983, the sensors on Soviet satellites detected sunlight glinting on clouds and the connected computer system misclassified the sensor input as the engines of intercontinental ballistic missiles. It warned Lieutenant Colonel Petrov of an incoming nuclear attack—an unintentional deception [12].

Existing limitations are likely to be magnified by intentional efforts to mislead and confuse those sensors and image detection programs. The seemingly sophisticated sensors of AWS might be able to penetrate camouflage designed to fool human sensors. But, at the same time, available knowledge about the properties and limitations of the sensors used in computer control and classification could make it easy to hide from and misdirect AWS in ways that a human would not even notice.

There is a great deal of interest in the development of adversarial images. These are images that confuse image recognition systems trained using machine learning. For example, an adversarial image that to a human eye looks like a turtle can be perturbed by adding visual noise so that it is recognised by an image classification system as a rifle instead [49]. Adversarial images have been discussed in the context of autonomous cars—in one example, stickers added to a ‘Stop’ sign led to it being classified as a 45 mile speed limit sign [50]. There is research into ways of making classification systems resilient to adversarial images, for instance by training them to recognise them as adversarial, but it is not clear how successful this would be, and constant retraining would be needed to respond to new adversarial developments.

Another form of deception that risks AWS being directed to the wrong targets is ‘spoofing’ by sending a powerful false GPS signal from a ground base. This could cause AWS to mislocate and be guided to crash into buildings. Image classification systems could also be misled through the use of perfidious markers—such as placing a red cross on a military vehicle to prevent it from being targeted. In the light of adversarial images, it might be possible to mark such vehicles in a way that was picked up by an AWS yet was undetectable by the human eye. Other forms of perfidy are possible: for instance, if it were known that AWS would not target funeral processions, military manoeuvres could be disguised as funeral processions. Or if

they were programmed to avoid targeting children, combatants could walk on their knees.

Of course, human eyes might also be deceived by such disguises, but AWS are dependent on their programming, the limitations of the sensors and on the programmers having anticipated a deception from the infinite number possible in conflict. Humans can be susceptible to deception, but they also have an understanding of human social situations and would be able to interpret a social gathering in the way that a weapon could not. Also, in the fog of war, there is the possibility that humans might intuit that something was amiss. When Lieutenant Colonel Stanislov Petrov was on duty in 1983 at the Russian nuclear detector centre, he decided not to trust the repeated warnings from the computer about an incoming nuclear attack from the US. He did not initiate a retaliatory nuclear strike and correctly reported it as a false alarm [11, 12].

This problem of the lack of human intuition and understanding by AWS, robots, and computer systems stems from the same technological (or metaphysical) shortcomings as those that result in the inability to adhere to the principles of discrimination and proportionality. It is another manifestation of the limitations of programmed or trained algorithms. There is a risk that AWS will be misled by the inputs they receive, and that they will have already attacked before any human has had the opportunity to sense that something is not right and that a mistake is about to be made. The speed at which such weapons are likely to operate (an argument sometimes used in their favour) exacerbates this risk and means that even when humans see that something is going wrong, they will be powerless to stop them.

Another problem with AWS, mentioned earlier, is the danger of unpredictable interactions between different algorithms. As explained by Sharkey [11, 35], the algorithms controlling AWS will be kept secret from the enemy. That means that it is impossible to know what will happen when two or more top-secret algorithms from opposing forces meet each other. Apart from the unknown interactions, the algorithms could be programmed for deceptive strategies such as feinting or sensor disruption.

Not only are there reasons to fear the unpredictable effects of different algorithms interacting, there is also the problem of unexpected interactions between the programming of the AWS and unanticipated environmental situations. In software testing, it is well known that bugs and errors will remain in code. As programmed entities it is impossible to test AWS for their behaviour in all of the unanticipated circumstances that can arise in conflict. And it is impossible to ensure that their behaviour will not be catastrophic in an environment of deceptive strategies.

6 Conclusions

It is clear that a major weakness of autonomous weapons systems is that their sensors and image processing systems are vulnerable to exploitation for the purposes of deception. We argue that their application in the field would be subject to large scale

deception by enemy forces. Not only could the sensors that control the movement and target selection of AWS be misled through their limitations, their incoming information could be deliberately distorted by the enemy to alter attack strategies. Deceptions of AWS could result in wasted firepower, missed targets, and ‘friendly’ casualties and mishaps. The high speed at which AWS will operate and their autonomous nature, would make it difficult, perhaps impossible, for military commanders to prevent mistaken targeting even if they were to become aware of it. There is already a well-established set of arguments against AWS. Now add the risks of deception, and the impact that this would have on civilian populations and infrastructure, and the urgency is clear for an international legally binding prohibition treaty that comprehensively bans the development, production and use of weapons that operate without meaningful human control.

References

1. Tzu S (2018) *The art of war*. Translated by Lionel Giles, Benediction Classics (original 5th Century BC)
2. Abbot J (1901) *Hannibal*. Harper and Brothers Publishers, New York and London. Accessed <https://www.heritage-history.com/index.php?c=read&author=abbott&book=hannibal&story>
3. Tucker SC (2013) Editor: *American civil war: the definitive encyclopedia and document collection*, p 1587
4. Pope D (2005) *The battle of the river plate: the hunt for the German pocket battleship Graf Spee*. McBooks Press
5. Garber M (2013) Ghost army: the inflatable tanks that fooled Hitler. *The Atlantic*, 22 May
6. Brown A (2007) *Bodyguard of lies: the extraordinary true story behind D-day*. The Lyons Press
7. US Department of Defense (2012) Directive 3000.09. In: *Autonomy in Weapons Systems*, 21 November, pp 13–14
8. Human Rights Watch (2012) *Losing humanity: the case against killer robots*. Accessed 17 May 2018. Available at: <http://www.hrw.org/reports/2012/11/19/losing-humanity-0>; Internet
9. ICRC (2014) ICRC, autonomous weapon systems: technical, military, legal and humanitarian aspects. In: *Expert meeting*, vol 1, Geneva, Switzerland, 26–28 Mar 2014
10. Amoroso D (2020) *Autonomous weapons systems and international law: a study on human-machine interactions in ethically and legally sensitive domains*. Edizioni Scientifiche Italiane, Napoli
11. Sharkey N (2020) Fully autonomous weapons post a unique dangers to human kind. *Scientific American*, February
12. Scharre P (2018) *Army of none: autonomous weapons and the future of war*. W.W. Norton and Company
13. Crootof R (2015) The killer robots are here: legal and policy implications. *Cardoza L Rev* 36:1837
14. Zacharius G (2015) US armed services committee hearing on “advancing the science and acceptance of autonomy for future defence systems”. <http://armedservices.house.gov/index.cfm/2015/11/advancing-the-science-and-acceptance-of-autonomy-for-future-defense-systems>
15. US Mission Statement (2020) <https://geneva.usmission.gov/2020/09/30/group-of-governmental-experts-on-lethal-autonomous-weapons-systems-laws-agenda-item-5d/>
16. Sharkey N (2012) The evitability of autonomous robot warfare. *Int Rev Red Cross* 94(886):787–799
17. Suchman L (2016) Situational awareness and adherence to the principle of distinction as a necessary condition for lawful autonomy. In: *Panel presentation at CCW informal meeting of experts on lethal autonomous weapons*, Geneva, 12 April 2016

18. Sharkey A (2017) Can we program or train robots to be good? *Ethics Inf Technol* (Online First). <https://doi.org/10.1007/s10676-017-9425-5>
19. Heyns C (2017) Autonomous weapons in armed conflict and the right to a dignified life: an African perspective. *South Afr J Hum Rights* 33(1):46–71
20. Asaro P (2012) On banning autonomous lethal systems: human rights, automation and the dehumanizing of lethal decision-making, special issue on new technologies and warfare. *Int Rev Red Cross* 94(886, Summer 2012):687–709
21. Sharkey A (2019) Autonomous weapons systems, killer robots and human dignity. *Ethics Inf Technol* 21(2):75–87
22. Tamburrini G (2016) On banning autonomous weapons systems. From deontological to wide consequentialist reasons. In: Bhuta N et al (eds) *Autonomous weapons systems. Law, ethics, policy*. Cambridge University Press, pp 121–141
23. Sharkey N (2008) Grounds for discrimination: autonomous robot. *RUSI Defence Syst* 11:86
24. Amoroso D, Tamburrini G (2017) The ethical and legal case against autonomy in weapons systems. *Global Jurist* 17:3
25. Sharkey N (2018) The impact of gender and race bias in A.I. *Humanitarian Law Policy Blog* August 2018
26. O'Neil C (2016) *Weapons of math destruction: how big data increases inequality and threaten democracy*. Penguin Books
27. Buolamwini J, Gebru T (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. In: *Proceedings of machine learning research, 2018 conference on fairness, accountability, and transparency*, vol 81, pp 1–15
28. Field Manual FM 90-2 (1988) *Battlefield deception*. US Army Washington DC
29. Joint Publication 3-13.4 (2006) *Military deception*. Joint Chiefs of Staff, USA
30. Article 37 (1977) Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (Protocol I), 8 June 1977
31. Field Manual FM 3-13.4 (2019) *Army support to military deception*. <https://armypubs.army.mil>
32. Carson TL (2010) *Lying and deception: theory and practice*. Oxford University Press Inc., New York
33. Zuckerman M, DePaulo BM, Rosenthal R (1981) Verbal and nonverbal communication of deception. In: Berkowitz L (ed) *Advances in experimental social psychology*, vol 14. Academic Press, New York, pp 1–59
34. Bok S (1999) *Lying: moral choice in public and private life*. Second Vintage Books Edition, New York
35. Sharkey A, Sharkey N (2020) We need to talk about deception in social robotics! *Ethics Inf Technol* (Published online 11 November)
36. Bond CF, Robinson M (1988) The evolution of deception. *J Nonverbal Behav* 12(4):295–307
37. Gerwehr S, Glenn RW (2020) The art of darkness: deception and urban operations. RAND Corporation, MR-1132-A, 2000, Santa Monica, Calif. As of July 20, 2020. https://www.rand.org/pubs/monograph_reports/MR1132.html
38. Johnson DG, Powers TM (2005) Computer systems and responsibility: a normative look at technological complexity. *Ethics Inf Technol* 7(2):99–107. <https://doi.org/10.1007/s10676-005-4585-0>
39. Johnson TH (2013) Taliban adaptations and innovations. *Small Wars Insurgencies* 24(1):3–27
40. Jessee DD (2006) Tactical means, strategic ends: Al Qaeda's use of denial and deception. *Terrorism Polit Violence* 18:367–388
41. Burgers TJ, Romaniuk SN (2017) Learning and adapting: Al Qaeda's attempts to counter drone strikes. *Terrorism Monit* 15:11
42. Bolton M (2020) New book shows catastrophic folly of automating warfare. In: *Blog on ICRAC (international committee for robot arms control) website*, posted September 20th 2020
43. MacAskill E (2009) <https://www.theguardian.com/world/2009/dec/17/skygrabber-american-drones-hacked>

44. Defensetech (2010) <https://www.military.com/defensetech/2010/08/10/hezbollah-claims-it-hacked-israeli-drone-video-feeds>
45. Papernot N, McDaniel P, Jha S, Fredrikson M, Celik ZB, Swami A (2016) The limitations of deep learning in adversarial settings. In: 2016 IEEE European symposium on security and privacy (EuroS P), pp 372–387. <https://doi.org/10.1109/EuroSP.2016.36>
46. Cummings ML (2020) Rethinking the maturity of AI in safety critical settings. In: Advances in artificial intelligence magazine
47. Alcorn MA, Li Q, Gong Z, Wang C, Mair L, Ku WS, Nguyen A (2018) Strike (with) a pose: neural networks are easily fooled by strange poses of familiar objects. arXiv: 1811.1153
48. Zang S, Ding M, Smith D, Tyler P, Rakotoarivelo T, Kaafar MA (2019) The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog and hail affect the performance of a self-driving car. IEEE Veh Technol Mag 103–111
49. Athalye A, Engstrom L, Ilyas A, Kwok K (2018) Synthesizing robust adversarial examples. In: Proceedings of the 35th international conference on machine learning (PMLR 80: 2840293)
50. Eykholt K, Evtimov I, Fernandes E, Li B, Rahmati A, Xi C, Prakash A, Kohno T, Song D (2019) Robust physical world attacks on deep learning models. arXiv: 1707.08945

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.