

## The Artificial Conscience of Lethal Autonomous Weapons: Marketing Ruse or Reality?

---

Type	Journal
Journal	Law and the digital age Conference
Date	19 February 2018
Issue	1
Page	1
Editor/Author	Guido Noto La Diega ( <i>Lecturer in Law, Northumbria University</i> )
Jurisdiction	Saudi Arabia, Qatar, Israel, Russia, United States, United Kingdom

---

Document link: [https://www.lexismiddleeast.com/magazine/ConferenceOnLawDigitalAge/2018\\_1\\_1](https://www.lexismiddleeast.com/magazine/ConferenceOnLawDigitalAge/2018_1_1)

## Abstract

There are two interwoven trends in cyber-counterterrorism. On the one hand, countries such as Israel and Russia announce the deployment of lethal autonomous weapons. Such weapons constitute the third revolution in warfare, after gunpowder and nuclear arms. On the other hand, researchers try and embed ethics into the design of these weapons (so-called artificial conscience or "ethics by design"). The contention of this paper is that artificial conscience is a mere marketing ruse aimed at making the deployment of lethal autonomous weapons and other autonomous robots acceptable in society. Whereas there are strong reasons to object to this trend, some solutions to the pitfalls of ethics by design have been presented. However, they do not seem viable in a military context. In particular, the so-called customised-ethics approach is applicable only to commercial and civil machines. When deciding whether to kill 600 civilians in order to hit 14 al-Qaeda leaders, which set of values should be implemented? This is a compelling argument for banning lethal autonomous weapons altogether.

## I. Introduction

There are two interwoven trends in cyber-counterterrorism. On the one hand, countries such as Israel and Russia are developing lethal autonomous weapons (LAWs).<sup>[3 p.9]</sup> On the other hand, researchers try and embed ethics into the design of these weapons (so-called artificial conscience<sup>[4 p.9]</sup> or "ethics by design"<sup>[5 p.9]</sup>). The contention of this paper is that artificial conscience is a mere marketing ruse aimed at making the deployment of lethal autonomous weapons and other autonomous robots acceptable in society.

In recent years, many scholars have published on embodying ethics into the design of software and hardware.<sup>[6 p.9]</sup> In 2016, the British Standards Institution published the first standard for the implementation of ethics through design.<sup>[7 p.9]</sup> Being the first of its kind, it is likely that the UK standard will prompt the adoption of similar guidelines in other jurisdictions, regardless of Brexit. This trend towards "ethics by design" is spreading also in the military context, with the attempts to create an "artificial conscience"<sup>[8 p.9]</sup> for LAWs. This trend may be regarded as a form of "ethical washing".<sup>[9 p.9]</sup> It constitutes an attempt to anthropomise machines in order to make them more acceptable, both in commercial and military contexts.<sup>[10 p.9]</sup> Along the same lines, one can read the recent European proposal to consider robots as electronic persons<sup>[11 p.9]</sup> and the achievement of citizenship by the robot Sophia in Saudi Arabia.<sup>[12 p.9]</sup>

This paper presents five arguments against "ethics by design" and two possible solutions. It is claimed that since it is impossible to ensure the morality of the LAWs' actions, these fully autonomous systems should be banned altogether.

On the same day of the conference "Law and the Digital Age", the second meeting of the UN Convention on Conventional Weapons<sup>[13 p.9]</sup> Group of Governmental Experts<sup>[14 p.9]</sup> will start. This author will share this paper with the High Contracting parties hoping to influence the debate, thus ensuring further impact to the Qatari conference.

A final caveat. While the dichotomy online/offline has long been challenged,<sup>[15 p.9]</sup> it does not seem already accepted that the contraposition between terrorism and cyberterrorism is obsolete. This dichotomy might have made sense at the time of computers, because a cyberterrorist hacking a computer could have perhaps stolen confidential information or transferred illegal funds, while a terrorist was a threat to the life and physical integrity of soldiers and civilians. However, in the age of the Internet of Things (IoT),<sup>[16 p.9]</sup> when every object around us can be remotely controlled and used to attack us, this contraposition no longer holds. Therefore, the debate around LAWs becomes of primary importance for cyber-terrorism experts as well.

In terms of methodology, this paper critically reviews international humanitarian law, as well as the regulation of LAWs in the United States, United Kingdom, and Russia. Particular attention is dedicated to the reports and working papers produced in the context of the UN Convention on Conventional Weapons. Hans Kelsen's theory of separation<sup>[17 p.10]</sup> is used as a theoretical framework to illustrate the relation between law and ethics as well as ethics and design.

## II. Outlawing LAWs

The recent past has been characterised by numerous attempts to bring the human soldier from "in" the loop, to "on" the loop and, finally, "out" of it.<sup>[18 p.10]</sup> Arguably, LAWs constitute the third revolution in warfare, after gunpowder and nuclear arms.<sup>[19 p.10]</sup> The development has been incremental and, therefore, difficult to detect. The vision is that of a bloodless or clean war where robots and, more generally, weapons are fully autonomous in decisions regarding their targets. For instance, an unmanned aerial vehicle (UAV) would decide that X is a terrorist, would be able to distinguish X from Y and Z (Y, Z = civilians) and, after a positive assessment that the potential death of Y and Z, the UAV complies with the principle of proportionality and delivers force and "neutralise" the target.

In the following subsections, this paper will present the US, UK, and Russian approaches to LAWs and some legal, ethical, and technical reasons for a pre-emptive ban on LAWs.

### A. LAWs in the United States, United Kingdom, and Russia.

In 2017, the US Department of Defense confirmed its Directive on Autonomy in Weapon Systems (the "Directive"),<sup>(18)</sup> which was the first public policy document on LAWs.<sup>[20 p.10]</sup> Under this Directive, autonomous weapon systems are defined as "weapon systems that, once activated, can select and engage targets without further intervention by a human operator."<sup>[21 p.10]</sup> Interestingly, the definition expressly includes human-supervised systems where an automated decision can be overridden by human operators (humans "on" the loop). However, the reference appears merely exemplificative and it would seem a contrario that also out-of-the-loop systems fall within the scope of the Directive. Now, fully autonomous weapons cannot be

used for lethal attacks,<sup>[22 p.10]</sup> however, the Directive is based on the outdated contraposition between cyber-war and war. Indeed, autonomous or semi-autonomous cyberspace systems<sup>[23 p.10]</sup> for cyberspace operations are left outside the scope of the Directive. This might be interpreted as meaning that fully autonomous systems can be used for lethal attacks if they can be considered as cyberattacks. In an IoT world, cyberattacks can have physical consequences; therefore, a clarification on this point is very much needed.

A detailed analysis of the Directive is beyond the scope of this paper. However, it is noteworthy that the first principle<sup>[24 p.10]</sup> it enshrines is that these systems should be designed to allow appropriate levels of human judgment over the use of force. Since this principle applies also to fully autonomous systems, its scope is not entirely clear. It may be interpreted in at least two ways. On the one hand, it may prevent the deployment of LAWs not only for lethal missions, but also every time that some use of force is involved. On the other hand, the principle may be interpreted as requiring two things: first, the possibility to override the decisions of semi-autonomous systems; and second, when it comes to LAWs, the necessity of an ex-ante careful consideration of the use of force, including a clarification of the time, place and limitations of the operation.

It seems clear that the development of LAWs is not ruled out. They could be approved by three Undersecretaries of Defense and the Chairman of the Joint Chiefs. Unfortunately, as confirmed by the US Delegation to the Convention on Conventional Weapons, the Directive does not prohibit the development of fully autonomous lethal weapons.<sup>[25 p.10]</sup> Experts have pointed out that it is clear that the Directive "permits the withdrawal of human judgment, control, and responsibility from points of lethal decision."<sup>[26 p.10]</sup> It is important to keep the eye of the public open, because the United States has made it clear that they may "delegate lethal authority to a machine to make a decision"<sup>[27 p.10]</sup> since authoritarian regimes may do so.<sup>[28 p.10]</sup> Moreover, the US Air Force has predicted that, by 2030, the capacity of unmanned systems will increase to the point that "humans will become the weakest link in a wide range of systems and processes."<sup>[29 p.10]</sup> Furthermore, moving from the out-of-date dichotomy between war and cyber-war, it seems that LAWs can be deployed for cyberattacks. Indeed, the Deputy Defence Secretary has stated that they will delegate machines authority "in things that go faster than human reaction time, like cyber or electronic warfare."<sup>[30 p.10]</sup> This is a cause for concern, since in an IoT world cyberattacks can have serious physical consequences.

The above-analysed approach sits in between a range of national positions with regards to LAWs. They are distributed along a continuum, with the United Kingdom closer to the pole of the LAWs' opponents<sup>[31 p.10]</sup> and the Russian Federation closer to the proponents of a regulation of, as opposed to a ban on, LAWs.

In September 2017, the United Kingdom ruled out the possibility to deploy LAWs by stating that military weapons "will always be under control as an absolute guarantee of human oversight and authority and accountability."<sup>[32 p.10]</sup> This is more than the "human judgment" required by the relevant US Directive, but clear guidelines would be needed to ensure meaningful human control.<sup>[33 p.11]</sup> For instance, it is to be hoped that the Israeli approach will not be followed. Indeed, it is not enough that there is human judgment in the "research, development, programming, testing, review, approval, and decision to employ (LAWs)."<sup>[34 p.11]</sup> Meaningful human control must entail some form of control on the delivery of force against every human target.<sup>[35 p.11]</sup> Moreover, one may criticise the way the Minister for the Armed Forces recently answered a question by a Member of Parliament who asked whether the Minister agreed that, no matter what the advances of technology on the battlefield, "only humans can effectively hold ground, deterring enemy activity and winning the hearts and minds of local communities."<sup>[36 p.11]</sup> To this, the Minister elusively answered that the United Kingdom "will definitely always need an Army."<sup>[37 p.11]</sup> Furthermore, the fact that the Ministry of Defence and BAE Systems are investing in the development of an autonomous system, the Taranis, which can autonomously locate targets and engage with them "raises concerns over the United Kingdom's assurances that it does not and will not develop LAWs."<sup>[38 p.11]</sup> The last criticism is that, while the United Kingdom is now committed not to deploy LAWs, it still refrains from supporting an international ban on such systems, with the justification that existing internal law is sufficient to regulate the use of such systems.<sup>[39 p.11]</sup> However, it is contended that a binding instrument would be of the utmost importance, particularly with regards to accountability and the principle of necessity, whose status as customary law is contested.

In Russia, responding to the Deputy Prime Minister of Defence and Space Industry, who called on industry to create weapons that "strike on their own",<sup>[40 p.11]</sup> Kalashnikov has recently developed a fully automated neural-networks-based combat module, which "which will use artificial intelligence to identify targets and make decisions on their own."<sup>[41 p.11]</sup>



Image n. 1 Autonomous weapon developed by Kalashnikov (30 June 2017). Source <https://kalashnikov.com/en/press-service/photo/>

In November 2017, the Russian Federation<sup>[42 p.11]</sup> issued a statement on LAWs that was reported in the news as if Russia were withdrawing from the UN Convention on Conventional Weapons and ignoring any future UN ban on LAWs.<sup>[43 p.11]</sup> At a closer look, the approach is more nuanced. The first, fundamental, assertion is that international humanitarian law applies to LAWs and meaningful human control should always be retained.<sup>[44 p.11]</sup> The reason why the Russian Federation does not presently support a pre-emptive ban on all LAWs is "the lack of working samples of such weapons systems,"<sup>[45 p.11]</sup> which leads to an incomplete understanding of this class of weapons. Second, there is the concern that, since the same technologies are used both for civil and military purposes, an all-encompassing pre-emptive ban may stifle innovation in autonomous technologies altogether.<sup>[46 p.11]</sup> One can argue that these reasons do not outweigh the reasons for a ban on LAWs, but one can hope that the work of the UN Convention on Conventional Weapons' Group of Governmental Experts will bring some progress in terms of definition of LAWs and, accordingly, scope of the ban. Moreover, it would seem that the combat module developed by Kalashnikov may overcome the objection regarding the lack of working samples.<sup>[47 p.11]</sup> At the same time, accepting that international humanitarian law is applicable to LAWs should be a reassurance that the new Kalashnikov's autonomous systems will not be deployed unless an international consensus is reached on the matter.<sup>[48 p.11]</sup>

## B. LAWs cannot comply with international legal, ethical and technical standards.

The deployment of LAWs raises a number of legal and ethical concerns. Human Rights Watch and the Harvard Law School's International Human Rights Clinic (IHRC) conducted a multidisciplinary study on LAWs. They concluded that LAWs are not "consistent with international humanitarian law and would increase the risk of death or injury to civilians during armed conflict."<sup>[49 p.12]</sup> This was the basis of the "Campaign to Stop Killer Robots", an international coalition formed in 2012 by non-governmental organisations<sup>[50 p.12]</sup> to pre-emptively ban fully autonomous weapons.<sup>[51 p.12]</sup>

At the time, fully autonomous systems were not in place yet, but today, thanks to advances in AI-related technologies, full autonomy is happening. As a result of the campaign, the UN Convention on Conventional Weapons held four yearly meetings (starting in May 2014) and formed a Group of Governmental Experts. The direction of this group is not very clear yet because they met only once,<sup>[53 p.12]</sup> but they seem to focus on problems of definition of LAWs and human control. Among the national statements, many endorse meaningful human control as a guiding principle and cryptographic proof of accountable human control has been suggested as a way to verify compliance with a ban on LAWs.<sup>[54 p.12]</sup> Support grows for new international laws on LAWs,<sup>[55 p.12]</sup> with Brazil, Uganda, and Iraq joining the campaign to ban LAWs and 84 countries that publicly elaborated their views on LAWs in a multilateral forum.<sup>[56 p.12]</sup> However, the international discussion around LAWs does not seem to gain momentum. In the week that followed the first and only meeting of the Group of Governmental Experts, 91 States convened for the annual meeting of the Convention on Conventional Weapons. Unfortunately, they simply agreed to continue formal deliberations on LAWs over ten days in 2018.<sup>[57 p.12]</sup> In particular, they will focus on problems of definitions, human-machine interaction, technology review, as well as humanitarian and international security challenges.<sup>[58 p.12]</sup> It is to be hoped that in 2018, they will be able to agree on a formal mandate to negotiate, which may lead to a new Additional Protocol to the Geneva Conventions to be adopted in 2020. Given the pace at which the relevant technologies are developing, no further delay is advisable.

There are legal and not-strictly-legal reasons to ban LAWs. Under the Additional Protocol I to the Geneva Conventions,<sup>[59 p.12]</sup> States must evaluate new or modified weapons to ensure compliance with international humanitarian law and, in particular, the so-called laws of war.<sup>[60 p.12]</sup> The main reference is to the rules of distinction, proportionality, and military necessity. LAWs cannot abide by them.

A fundamental principle of the *ius in bello* is distinction, whereby one must distinguish at all times between civilians and combatants and between civilian objects and military targets.<sup>[61 p.12]</sup> Moreover, enemy combatants may not be targeted if *hors de combat* because they have surrendered or are defenceless due to wounds or sickness.<sup>[62 p.12]</sup> For a machine, a frightened mother running after her two children and yelling at them to stop playing with toy guns near a soldier may be seen simply as someone running towards two armed people and, therefore, a legitimate target.<sup>[63 p.13]</sup> It has been objected that one cannot rule out compliance with the principle of distinction based on worst-case scenarios and that "LAWS are likely to be deployed in a wide variety of combat situations, including permissive environments where LAWS may attack a tank formation in a remote area (such as a desert)."<sup>[64 p.13]</sup> The latter approach seems to be as excessive as the former, even if it goes in the opposite direction. Arguably, it is not possible to define *ex ante* what are the contexts where LAWs could comply with the principle of distinction. Moreover, a normally "permissive" environment could change status, for instance, because some soldiers got lost in the desert and are wounded and, therefore, *hors de combat*.<sup>[65 p.13]</sup>

As pointed out by Human Rights Watch,<sup>[66 p.13]</sup> machines cannot distinguish between a fearful civilian and a threatening enemy combatant because this would require the understanding of the intentions behind a human's actions. It does not seem that state-of-the-art technologies allow this kind of operation.<sup>[67 p.13]</sup> It is noteworthy that over 3,000 researchers in robotics and artificial intelligence signed an open letter in favour of a pre-emptive ban on LAWs.<sup>[68 p.13]</sup>

Another reason why LAWs are incapable of distinguishing between a civilian and a soldier is the increasingly common use of un-uniformed combatants.<sup>[69 p.13]</sup> Some scholars suggest resolving the issue through a non-engagement default. In other words, LAWs should be programmed to treat potential targets "as innocent civilians when the robot is unsure of the true nature of the target."<sup>[70 p.13]</sup> However, such a default would significantly decrease the efficacy and, accordingly, the utility of such systems.

The principle of proportionality is another pillar of the *ius in bello*. Under Protocol I to the Geneva Conventions, an attack is indiscriminate and, therefore, prohibited if it

- "may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated."<sup>[71 p.13]</sup>

The assessment of whether the "collateral" damage to civilians is excessive compared to the military advantage is very delicate indeed. Proponents of LAWS claim that, in order to prove that these systems cannot comply with the principle of proportionality, their opponents should "must prove that every use of LAWS would result in civilian casualties excessive to the military advantage gained by the attack."<sup>[72 p.13]</sup> Following this line of reasoning, however, one should infer that it is legal to carry out disproportionate attacks in most cases, as long as in some instances the principle of proportionality is respected. Obviously, this is not quite accurate. The proponents of LAWS, then, make some examples where proportionality could be ensured, but they refer to scenarios where there are no civilians; in such scenarios, proportionality does not play any role. Lastly, those who claim that LAWS can comply with the principle of proportionality<sup>[73 p.13]</sup> argue that the fact that the proportionality assessment requires human judgement<sup>[74 p.13]</sup> is not a good enough reason as long as the system can reason like a human commander or when a human commander has already conducted said assessment. The first scenario, however, would seem to require the so-called "singularity",<sup>[75 p.13]</sup> which might be near, but surely is not here yet. The second argument, in turn, ignores that the definition itself of autonomy and LAWS excludes that humans are involved in the operations, including the selection of the target. It does not seem that guidelines on hypothetical scenarios could be considered a proper way to satisfy the analysed principle.

The third principle that LAWS can hardly comply with is necessity, which infuses international humanitarian law.<sup>[76 p.13]</sup> It "dictates that military force should only be used against the enemy to the extent necessary for winning the war."<sup>[77 p.13]</sup> Now, LAWS pose a threat to this principle in a twofold way. First, they cannot assess it because of its intrinsically human nature. Indeed, the necessity assessment is a value-based judgement where the commander needs to balance the imperatives of winning with the requirement of humanity.<sup>[78 p.13]</sup> From this point of view, if the attempts to embed ethics into the design of LAWS, this issue could be successfully addressed. Second, if LAWS were to be deployed, they may lead to a rethinking and expansion of the concept itself of necessity, making the use of force "necessary" in instances where, in the past, it would have been considered unnecessary.<sup>[79 p.13]</sup>

Whereas it has become hardly deniable that the principles of distinction<sup>[80 p.13]</sup> and proportionality<sup>[81 p.13]</sup> are customary, some contend that the same does not apply to the principle of necessity. This would have practical consequences, because if the principle of necessity were not customary, it would bind only the states party to the Additional Protocol I, excluding, for instance, the United States.<sup>[82 p.13]</sup>

More generally, applying the principles of distinction, proportionality, and necessity in real-world scenarios is an entirely different matter to translating them into code based on hypothetical scenarios. Understanding the world is an intrinsically holistic and discretionary operation<sup>[84 p.14]</sup> which does not lend itself to algorithmic reductions. Interpreting the world is not the same as executing computerised commands. In real-world war theatres, interpretation requires sympathy and empathy, which are quintessentially human features.

The principles of distinction, proportionality, and necessity are assessed on a case-by-case basis;<sup>[85 p.14]</sup> they are not rules that can be coded and implemented in an algorithm running on an autonomous system. A type of rule that could be implemented into a LAW could be "if there are less than ten children around a terrorist, then deliver force". However, it is not clear who would have the legitimacy to decide once and for all that it is worth it to kill nine children to "neutralise" a terrorist.<sup>[86 p.14]</sup>

One might argue, however, that no international treaty expressly bans LAWS. Such an argument would be very weak indeed. In fact, fully autonomous weapons are likely to contravene the so-called Martens Clause, which prohibits weapons that are repugnant to the public conscience.<sup>[87 p.14]</sup> It is noteworthy that the International Committee of the Red Cross<sup>[88 p.14]</sup> explained the codification of such a rule of customary international law as the prevention of the assumption whereby anything that is not explicitly prohibited by the relevant treaties is therefore permitted. Moreover, the Martens Clause is a dynamic factor ensuring that the principles of international humanitarian law apply regardless of the type of situation or technology.<sup>[89 p.14]</sup> Thus, even though originally the Clause was designed for the francs-tireurs,<sup>[90 p.14]</sup> nothing precludes an interpretation that applies it to the ban of LAWS. However, future research should collect more data on the "public conscience", i.e., whether the public presently considers repugnant the deployment of such lethal autonomous systems.<sup>[91 p.14]</sup>

Lastly, and perhaps more importantly, there is an issue of accountability. Distinction, proportionality, and necessity can be of problematic interpretation also for human operators, but the relevant violation would be persecuted, for instance by the war crimes tribunals. However, these principles are directed to "those who plan or decide upon an attack"<sup>[92 p.14]</sup> and LAWS themselves cannot be held accountable. At the same time, when a LAW identifies a target and delivers force autonomously, it is unclear whether and under which conditions the commander who deployed it, the programmer or the manufacturer could be held liable.<sup>[93 p.14]</sup> While currently machines cannot be punished for their actions, the trend towards robot legal personality may bring to a future where people consider that imprisoning a machine is an appropriate response to the breach of the legal order.<sup>[94 p.14]</sup>

Alongside the debatable compliance with said legal standards, LAWS raise a number of ethical concerns.

First, emotions and compassion provide "an important check on the killing of civilians."<sup>[69 p.13]</sup> For instance, arguably, a dictator would meet the resistance of their soldiers in case of an attack of their own people. However, this line of reasoning would not apply to non-human soldiers.<sup>[95 p.14]</sup>

Second, the idea of a clean war would increase the likelihood of a conflict.<sup>[96 p.14]</sup> As correctly noted, some nations would "be emboldened to start wars if they believe they can achieve political objectives without the loss of their troops."<sup>[97 p.15]</sup> This would have potentially tragic consequences for human civilians caught in the "bloodless" crossfire.<sup>[98 p.15]</sup>

Third, as Edward Murrow famously said, in diplomacy, what counts are the last three feet "bridged by personal contact, one person talking to another."<sup>[99 p.15]</sup> Channelling Murrow, scholars asked whether machines could be used in "the last three feet", suggesting that one cannot win hearts and minds by having machines as the interface of an occupied population.<sup>[100 p.15]</sup> This was thought for simple military robots, but it applies all the more to LAWs.

The ethical issues of LAWs have a lot in common with those of military robots, which should not come as a surprise since most LAWs can be regarded as robots. The features that a robot must have in order to qualify as a LAW are to be autonomous and to be designed – or at least to have the potential to be used – for lethal attacks. This makes the ethics of LAWs a proper minefield.

Alongside the legal and ethical issues, the technical ones abound and they are mainly related to the LAWs' vulnerability to hacking<sup>[101 p.15]</sup> and to the problem of the so-called unintended interactions.<sup>[102 p.15]</sup> An analysis of the technical issues is beyond the scope of this paper, but suffice to say that "networks of autonomous weapons could accidentally ignite a war and, once it has started, rapidly escalate it out of control."<sup>[103 p.15]</sup> With the IoT and swarm robotics, one can predict that this may become a major issue in this new era of "bloodless" war.<sup>[104 p.15]</sup>

Even if some scholars are sceptical,<sup>[105 p.15]</sup> this paper joins the majority of the scientific community in believing that the call for a pre-emptive ban on LAWs is solidly grounded in law and ethics. A common proposal<sup>[106 p.15]</sup> is the adoption of a new ad hoc treaty with provisions on distinction, proportionality, and multiple-actors accountability, as well as the establishment of a body capable of monitoring compliance. A new ad hoc binding instrument could be useful, especially to resolve issues of accountability and to clarify the customary nature of the principle of necessity.

This said, the process of adoption and ratification is a long and perilous one. Therefore, it would be advisable that States issue a joint declaration that existing international humanitarian law applies to LAWs.<sup>[107 p.15]</sup>

### III. Against Moral Killing Machines

A way to justify the final move from semi-autonomous weapons to proper LAWs can have the colourable face of the "ethics by design", also known as artificial conscience or value-sensitive design.

Ethics by design deals with embedding ethical values "into the design of specifications of technologies, procedures, practices, or infrastructures."<sup>[108 p.15]</sup> In this paper, the main reference is to the embedment into technologies. The field is characterised by a significant amount of politicisation. So, for instance, some scholars go as far as to claim that it would be immoral not to build moral robots.<sup>[109 p.15]</sup> Resisting the temptation to reply that it would be immoral to develop "moral" killing machines, there are some strong arguments against this trend.

The research in this field is decidedly led by its civil and commercial applications, but one cannot ignore the risk of a cross-fertilisation. The risk of a new wave of research in the field has gotten recently higher, with the new EU legislation on data protection coming into form on 25 May 2018.<sup>[110 p.15]</sup> Indeed, the principle of data protection by design<sup>[111 p.15]</sup> may be regarded as the legal foundation for embedding ethics into the design of machines, including LAWs.<sup>[112 p.15]</sup>

A further input to the development of ethics-by-design approaches may come from the British Standards Institution, which published the first standard for the implementation of ethics into the design of robots. Being the first of its kind and coming from one of the main standardisation organisations in the world, this standard is likely to constitute a model for other jurisdictions and to be applied also to some robotic sub-sectors, including LAWs. Interestingly, the British Standards Institution underlines that compliance with the standard does not ensure compliance with the law and cannot be used as a liability disclaimer, thus confirming the separation between law and ethics.

Another way to describe the phenomenon goes by the name of artificial conscience, which is the military application of the ethics by design. To put it simply, it is the attempt to build moral killing machines. The most relevant research projects seem those carried out by the roboticists led by Ron Arkin at Georgia Tech. They propose to integrate a code of ethics into the machines to ensure their conformity with international humanitarian laws.<sup>[113 p.15]</sup> In particular, they propose to embed an artificial "conscience" in an intelligent autonomous robotic agent, which applies limits to its actions. These are derived from the principles of international humanitarian law, as well as rules of engagement. Allegedly, the "intent is to yield robots that can perhaps act more humanely than humans."<sup>[114 p.15]</sup>

One may argue that if LAWs can think and behave morally, this would be a compelling reason to bring the human operator out of the loop, because a LAW would have the best of both worlds: human morality and artificial accuracy. This way of reasoning would appear rather weak, for a number of reasons which could be formulated as follows. The following ones are the reasons why one should not adopt an ethics-by-design approach to LAWs and, more generally, robotics. It is beyond the scope of this paper the investigation about the technical feasibility of an operation that requires transforming values into strings of code. It would seem, however, that most ethical concepts, such as "just" or "fair", are too vague to be coded.<sup>[115 p.15]</sup> Moreover, understanding the world is a holistic discretionary process that does not lend itself to algorithmic reductions.

The first and main argument against ethics by design revolves around the separation between law and ethics. If law and ethics are separate and if design is the law of cyberspace,<sup>[116 p.15]</sup> then design and ethics also belong to different realms. To explain the relationship between law and ethics, this paper presents Hans Kelsen's views, because they are particularly illustrative of the risks of conflating law and ethics. Moreover, inasmuch Kelsen favours autonomy over paternalism, his approach is consistent with another argument against ethics by design which will be presented in this paper.<sup>[117 p.16]</sup>

According to Kelsen, the point of view of legal norms and the one of moral norms are exclusive of each other (so-called separation theory or *Trennungstheorie*).<sup>[118 p.16]</sup> In particular, the legal system and the moral one co-exist, but only one of them is valid. From the lawyer's viewpoint, there are only legal norms, from the ethicist's one only moral norms.<sup>[119 p.16]</sup> These viewpoints are exclusive and exhaustive, i.e., there is no third viewpoint from which there are both valid legal norms and valid moral norms.

Kelsen's positivism was accused of rendering "the German judiciary defenceless against laws with arbitrary or criminal content."<sup>[120 p.16]</sup> Indeed, if the legal viewpoint and the moral one are exclusive of each other, "the validity of a positive legal order does not depend on its conformity with some moral system."<sup>[121 p.16]</sup> These critiques fail to understand the rationale underlying the *Trennungstheorie*, i.e., that if only just law is law, legal systems are all morally justified.<sup>[122 p.16]</sup> Indeed, affirming that "law must be moral" presupposes an absolute moral order valid outside time and space, as well as an uncritical justification of the national coercive order.<sup>[123 p.16]</sup> The conflation of law and ethics, therefore, serves the reactionary purpose of justifying a particular social order.<sup>[124 p.16]</sup> Decisively, one needs to keep in mind that the Nazi legal theorist's program of a unification of law and morality "served to extend the authority and power of the Nazi-regime."<sup>[125 p.16]</sup> Moreover, Kelsen distinguishes between "Geltung" (a norm exists and is valid regardless of its moral content) and "Verbindlichkeit", its binding nature: it must be left to the individual to answer the moral question of whether to comply with a moral standard.<sup>[126 p.16]</sup> Kelsen's theory, in other terms, empowers the individual by favouring autonomy over paternalism.

The argument can be put forward that selecting one particular set of values and embedding them in a machine that could be shipped anywhere in the worlds and used by customers with radically different morals is, at the very least, an imperialistic operation. Arguably, since every set of values is closely related to the relevant society, the choice architecture may not be acceptable to people with different backgrounds. For instance, will the values selected by a team of white American Christian middle-class heterosexual men be acceptable from the perspective of a team of black Arab Muslim working-class lesbian women?

If Kelsen's doctrine of unity<sup>[127 p.16]</sup> may sound alien to a non-lawyer, perhaps some practical examples of four potential clashes between what is ethical and what is legal may complete the picture. First, unethical behaviours can be legal. One need only think that cheating *per se* is legal in most jurisdiction, but it is objectionable from an ethical viewpoint. Second, illegal behaviours can be ethical. For instance, a mother leaving the car in a parking space reserved for disabled people to bring a dying child to the hospital is acting illegally, yet ethically. Third, an ethical behaviour can be illegal. The classical example of this is the Jewish trade under fascism and Nazism. Lastly, legal behaviour can qualify as unethical, as was executing Hitler's orders. Who would be responsible for a LAWs acting illegally due to an ethical constraint?

If law and ethics are separate and design is the new law of cyberspace, then design and ethics belong to different realms, which is a strong argument against their conflation in the form of ethics by design or artificial intelligence.

A second argument against the "ethics-by-design" trend is that embedding ethics into the design of autonomous systems is a form of illiberal paternalism. As pointed out by Floridi,<sup>[128 p.16]</sup> this approach relies on human inertia and biases as the pragmatic motivation for action; it shapes the system and hence disempowers agents. Human agents are, therefore, ethically desensitised, deskilled, and de-responsabilised; they are "merely herded, mindlessly and non-responsibly, towards some pre-established options chosen by the designers of the environment."<sup>[129 p.16]</sup> Moreover, alongside good reasons against paternalism, there are compelling reasons for its contrary, i.e. autonomy and self-determination. Indeed, favouring autonomy increases the capability of making critical choices and to assume responsibilities. Responsabilisation and ability to make critical choices are all the more crucial in a military context, when human operators routinely make life-or-death decisions.

Third, there is the contraposition between technocracy and democracy. In the law-making process, legislators may select a number of values that become law. However, in the vast majority of the instances, this selection is carried out in a transparent way, with a democratic legitimacy and with the guarantee of political and legal accountability.<sup>[130 p.16]</sup> Conversely, with the ethics –by-design, the selection and embedding of ethics is carried out in an opaque way, by a small programming community "many of whom are in the pay of a few oligopolistic corporations directly accountable to no external party."<sup>[131 p.16]</sup>

A fourth argument is that designing a machine means, first and foremost, programming it to carry out a number of tasks in a reliable and accurate way. Programmers are already required to implement, alongside the functional instructions, also the law. Indeed, for instance, they have to minimise the processing of personal data, pseudonymise them, ensure transparency and enable the data subject to monitor the data processing, as well as enabling the controller to create and improve security features.<sup>[132 p.16]</sup> If, on top of that, there was an obligation to design machines that are moral, then there is the risk that the machine would receive a number of conflicting commands, thus becoming unable to perform its primary tasks. A practical example is the European Parliament's recommendations on robotics.<sup>[133 p.16]</sup> According to the EU institution, indeed, designers of robots should take into account an ethical framework comprising several values including beneficence, non-maleficence, autonomy and justice, as well as the principles and values enshrined in Article 2 of the Treaty on the EU<sup>[134 p.17]</sup> and in the Charter of Fundamental Rights of the EU,<sup>[135 p.17]</sup> such as human dignity, equality, justice and equity, non-discrimination, informed consent, private and family life, and data protection. The potpourri is completed by the reference to the other underlying principles and values of the EU law, such as non-stigmatisation, transparency, individual responsibility and social responsibility, and on existing ethical practices and codes. It is not clear how one could embed all these values without contradictions<sup>[136 p.17]</sup> and without affecting the security and reliability of the machine.<sup>[137 p.17]</sup>

A fifth argument is that ethics is intrinsically human and, therefore, cannot be synthesized. Two of the most credited theories on ethics are Hume's and Mills' ones.<sup>[138 p.17]</sup> According to the father of skepticism, morality is grounded on our experience of our own feelings.<sup>[139 p.17]</sup> In a utilitarian perspective, in turn, desire is the only evidence of what is good i.e., moral and everyone desires happiness.<sup>[140 p.17]</sup> Similar ideas have been put forward by Kelsen, when he concluded that the "longing for justice is men's eternal longing for happiness (...) Justice is social happiness."<sup>[141 p.17]</sup> Claims that machines "feel" and "desire" seem untenable. To put it lightly, it is not a case that IBM Watson did not celebrate its victory at Jeopardy:<sup>[142 p.17]</sup> it did not feel happy and, therefore, it did not desire to celebrate.

## IV. Conclusions. Moral Lethal Autonomous Weapons are Pirandello's Old Grotesque Woman.

This paper proved that there is a twofold trend towards a dehumanised counter-terrorism and, correspondingly, an attempt to humanise the autonomous machines used to kill, i.e. the LAWs. However, some arguments have been presented against both trends. They revolve around the concepts of separation between law and ethics, illiberal paternalism, technocracy, ethics-reliability trade-off, and ethics as intrinsically human. The idea of an artificial conscience is being used to convince the citizens that LAWs can comply with the humanitarian principles of distinction, proportionality, and necessity. This paper has argued the LAWs non-compliance with these principles, as well as with other ethical standards and technical requirements.

Embedding values into LAWs is part of a clear strategy aimed at making citizens accept the replacement of human beings with non-human agents. This strategy comprises also the attempt to make machines look like humans (anthropomorphisation) and to give them legal personality and citizenship. These developments bring to one's mind Pirandello's old woman<sup>[143 p.17]</sup> who wears excessive make-up to look like a young girl. The first reaction of an observer might be to laugh at her, but bitterness would then prevail once one understands the reasons that led the woman to appear so grotesque.

A solution to the criticisms regarding ethics by design or artificial conscience might be to always leave the human being the possibility to make the moral decision. For instance, a driverless car in front of some children should alert the owner and let them decide whether to kill themselves or the children. However, this solution does not seem always practical or viable. It is not practical when a decision needs to be made in an extremely short period of time. It is not viable when we enter the stage of fully autonomous systems, where there is no room for human intervention. One could respond that if instead of wasting time on trying to build moral machines, one would focus on security, the moral dilemma would not exist. The car would be able to stop well before putting the owner in the position to decide whether to kill themselves or the children.<sup>[144 p.17]</sup> In September 2017, the European Union announced a new cybersecurity agency.<sup>[145 p.17]</sup> Hopefully, this will help to bring back the focus on security, rather than ethics.

A second solution may be to use a machine learning algorithm in order to make the autonomous system learn from the values of the owner, thus giving life to machines whose ethics is customised. Thus, one would overcome the arguments revolving around imperialism, and paternalism, technocracy. It would also work well with fully autonomous machines, capable of constantly learning from their user. Recently, some scholars presented the idea of an "ethical knob,"<sup>[146 p.17]</sup> a control on an autonomous vehicle enabling passengers to choose between an altruistic, impartial, or egoistic mode. The difference between the latter approach and the one suggested here is that the "ethical knob" would meet the same criticisms revolving around technocracy and imperialism presented with regards to ethics-by-design approaches. Moreover, it would be paternalistic, even though in quite a mild way. The here-proposed customised ethics would overcome most objections, with the exclusion of the one according to which ethics is intrinsically human and, therefore, could not be coded.

These solutions do not seem viable in a military context. In particular, the customised-ethics approach is applicable only to commercial and civil machines. When it comes military uses, however, one would not know how to choose whose values the machine should learn from. When deciding whether to killing 600 civilians in order to hit 14 al-Qaeda leaders,<sup>[147 p.17]</sup> it is unclear who should decide which set of values should be implemented and how.

The impossibility to ensure the LAWs' compliance with legal and ethical standards is a compelling argument for outlawing lethal autonomous systems altogether. Hopefully, this will be the outcome of the Convention on Conventional Weapons' Group of Governmental Experts, who thus will follow the precedent of banning the blinding laser before it was put into use.<sup>[148 p.17]</sup> Once deployed, the lack of human control may lead to an uncontrollable escalation of violence against civilians. If, however, the contracting parties stick to the guiding principle of meaningful human control, then one could accept the regulated deployment of LAWs, which would be "autonomous" only in name. An ad hoc binding instrument would be particularly important to ban LAWs and regulate accountability in human-on-the-loop systems, but international negotiations are a long and perilous process. In the meantime, it will be of the utmost importance to issue a joint statement of all States reaffirming the applicability of existing international humanitarian law to LAWs. This would have the effect of, inter alia, preventing the countries that have already developed autonomous systems from deploying them.

Another contribution of this paper is that it provides arguments to overcome the terrorism / cyber-terrorism dichotomy. The regulations on autonomous and semi-autonomous weapons tend to build on said dichotomy and, therefore, leave cyber-attacks out of their scope. However, in an IoT world, cyberattacks can have physical consequences; therefore, regulations should take account of this new reality.

## Notes

1. <sup>▲ [p.2]</sup> Authored by Guido Noto La Diega, Lecturer in Law, Northumbria University, President, Ital-IoT Centre of Multidisciplinary Research on The Internet of Things, Convenor, NINSO The Northumbria Internet & Society Research Interest Group
2. <sup>▲ [p.2]</sup> Previous versions of this paper were presented at the Annual Conference of the Society of Legal Scholars (SLS 2017, Dublin, 7-8 September) and at the European Cybersecurity Forum (CYBERSEC 2017, Krakow, 10 October 2017). The author is grateful for comments and feedback received in these venues. However, the responsibility for views and errors rests solely with the author.
3. <sup>▲ [p.2]</sup> It has been reported that LAWs are being developed in more than a dozen countries, including the United States, China, Israel, South Korea, Russia, and the United Kingdom. M. Wareham, "Banning Killer Robots in 2017", Human Rights Watch, 15 January 2017, <https://www.hrw.org/news/2017/01/15/banning-killer-robots-2017>, accessed 7 January 2018.
4. <sup>▲ [p.2]</sup> Artificial conscience is not to be confused with artificial consciousness, whose functions included awareness of self, will, instinct, and emotion (I. Aleksander, Machine consciousness, 3(2) Scholarpedia 4162 (2008)). While arguably proper conscience presupposes consciousness, one can concede that values can be implemented even in non-conscious machines. "Artificial conscience" was first used by R.C. Arkin & L. Mosnkina, Lethality and Autonomous Robots: An Ethical Stance, IEEE International Symposium on Technology and Society 1 (2007).
5. <sup>▲ [p.2]</sup> See, for example, R.C. Arkin & L. Mosnkina op. cit.; R.C Arkin, "Governing lethal behavior: embedding ethics in a hybrid deliberative/reactive robot architecture, Proceedings of the 3rd ACM/IEEE international conference on human robot interaction 121 (2008). Most of the applications are outside the military context. See, for example, D.S. Borrett et al., Research ethics by design: A collaborative research design proposal, 13 RESEARCH ETHICS 84-91 (2017); G. Baldini et. al., Ethical Design in the Internet of Things, 24 SCIENCE AND ENGINEERING ETHICS 905-925 (2018).
6. <sup>▲ [p.2]</sup> Alongside the authors cited in the precedent footnote, see B. Friedman, Value-sensitive design 3(6) INTERACTIONS 16 (1996); F. Dechesne et al., Ethical requirements for reconfigurable sensor technology: a challenge for value sensitive design, 15(3) ETHICS AND INFORMATION TECHNOLOGY 173 (2013); A. Wynsberghe, Designing Robots for Care: Care Centered Value-Sensitive Design, 19(2) SCIENCE AND ENGINEERING ETHICS 407 (2013); J. van Andel et al., What makes a good home-based nocturnal seizure detector? A value sensitive design, 10(4) PLoS ONE 1 (2015); C.D. Martin and T. Taylor Makoundou, Taking the High Road: Ethics by Design in AI, 8(4) ACM INROADS 35 (2017). cf C. Tonkinwise, Ethics by Design, or the Ethos of Things, 2(2) DESIGN PHILOSOPHY PAPERS 12 (2004).
7. <sup>▲ [p.2]</sup> British Standards Institution, *Robots and robotic devices. Guide on the ethical design and application of robots and robotic systems (BS 8611:2016)* (British Standards Institution, 2016).
8. <sup>▲ [p.2]</sup> R.C. Arkin & L. Mosnkina, op. cit.
9. <sup>▲ [p.2]</sup> cf G. Rose, "Ethical-washing: Je suis tombée dans le panneau...", *Affectueusement Votre*, 12 November 2017, <https://affectueusementvotre.com/ethical-washing/>, accessed 22 December 2017.
10. <sup>▲ [p.2]</sup> It is worth pointing out that this is not necessarily the intention of the proponents of the relevant theories. It may well be that, for instance, Arkin does not want to favour the deployment of LAWs, but his theory risks to be used to that end.
11. <sup>▲ [p.2]</sup> For more information on this proposal, see G. Noto La Diega, "The European strategy on robotics and artificial intelligence: too much ethics, too little security" 3(2) EUR'N CYBERSECURITY J. 6 (2017).
12. <sup>▲ [p.2]</sup> Z. Stone, "Everything You Need To Know About Sophia, The World's First Robot Citizen" (Forbes, 7 November 2017), <https://www.forbes.com/forbes/welcome/?toURL=https://www.forbes.com/sites/zarastone/2017/11/07/everything-you-need-to-know-about-sophia-the-worlds-first-robot-citizen/> accessed 22 December 2017.
13. <sup>▲ [p.2]</sup> United Nations 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 (hereinafter also the "Convention on Conventional Weapons").
14. <sup>▲ [p.2]</sup> In December 2016, the Fifth Review Conference of the High Contracting Parties to the 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 established an open-ended Group of Governmental Experts related to LAWs. See Decision 1 of its Final Document (CCW/CONF.V/10, Part II, Final Declaration, Section III).
15. <sup>▲ [p.2]</sup> cf L. Floridi, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality* (Oxford University Press, 2014).
16. <sup>▲ [p.2]</sup> There is no generally accepted definition of the IoT, but the International Standardization Organization (ISO) defines it as an "infrastructure of interconnected objects, people, systems, and information resources together with

intelligent services to allow them to process information of the physical and the virtual world and react" (ISO/IEC JTC 1, *Internet of Things (IoT) Preliminary Report 2014* (ISO, 2015) 4.1 [https://www.iso.org/files/live/sites/isoorg/files/developing\\_standards/docs/en/internet\\_of\\_things\\_report-jtc1.pdf](https://www.iso.org/files/live/sites/isoorg/files/developing_standards/docs/en/internet_of_things_report-jtc1.pdf)> accessed 22 December 2017). From other perspective, these "things" (popularly known as smart devices) can be seen as an inextricable mixture of software, hardware and services (G. Noto La Diega and I. Walden, "Contracting for the "Internet of Things": Looking into the Nest, 7(2) EUR. J. OF L. AND TECH. 1 (2016)).

17. <sup>▲ [p.2]</sup> The main references are H. Kelsen, *General Theory of Law and State* (A. Wedberg tr., Harvard University Press, 1945), 5-14, 363-390, 410-411; H. Kelsen, *Pure Theory of Law* (Max Knight tr., 2nd ed., University of California Press, 1967), ch. 2; H. Kelsen, *Théorie pure du droit* (Charles Eisenmann tr., Dalloz, 1962); H. Kelsen, *What is justice?* (Berkeley: University of California Press, 1957), ch. 1.
18. <sup>▲ [p.2]</sup> When the human operator is in the loop, the machine cannot select the target or deliver force without a human command. In turn, with on the loop, the machine can select targets and deliver force, with human oversight and power to override the machine's actions. Lastly, human-out-of-the-loop machines do not need any human interaction or input. See "Losing Humanity: The Case against Killer Robots", Human Rights Watch, 29 November 2012, <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>, accessed 27 January 2017. cf G. Schaub Jr. & J. Wenzel Kristoffersen, "In, On, or Out of the Loop? Denmark and Autonomous Weapon Systems", Centre for Military Studies University of Copenhagen, February 2017, [http://cms.polsci.ku.dk/publikationer/in-on-or-out-of-the-loop/In\\_On\\_or\\_Out\\_of\\_the\\_Loop.pdf](http://cms.polsci.ku.dk/publikationer/in-on-or-out-of-the-loop/In_On_or_Out_of_the_Loop.pdf), accessed 22 December 2017.
19. <sup>▲ [p.2]</sup> Future of Life Institute, "Autonomous Weapons: An Open Letter from AI & Robotics Researchers" (FLI, 28 July 2015), <https://futureoflife.org/open-letter-autonomous-weapons/>, accessed 29 December 2017.
20. <sup>▲ [p.2]</sup> M. Gubrud, "DoD Directive on Autonomy in Weapon Systems" (ICRAC, 27 November 2012), <https://icrac.net/2012/11/dod-directive-on-autonomy-in-weapon-systems/>, accessed 22 December 2017.
21. <sup>▲ [p.2]</sup> DoDD 3000.09, glossary, Part II. There is no commonly accepted definition of LAWs and the UN Convention on Certain Conventional Weapons would achieve something of the utmost importance if they could reach an agreement on the point.
22. <sup>▲ [p.2]</sup> Only semi-autonomous systems can be deployed for lethal operations. This kind of system engages with individual targets or specific target groups selected by human operators. See DoDD 3000.09, glossary, part II.
23. <sup>▲ [p.3]</sup> Further research on the meaning of semi-autonomous weapons should be carried out. Arguably, these include also lock-on-after-launch homing munitions which are de facto fully autonomous (M. Gubrud, *op. cit.*).
24. <sup>▲ [p.3]</sup> DoDD 3000.09, para 4(a).
25. <sup>▲ [p.3]</sup> US Opening Statement at the CCW Informal Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, 13 April 2015, <https://geneva.usmission.gov/2015/04/15/u-s-opening-statement-at-the-ccw-informal-meeting-of-experts-on-lethal-autonomous-weapons-systems/>, accessed 29 December 2017.
26. <sup>▲ [p.3]</sup> M. Gubrud, "Why Should We Ban Autonomous Weapons? To Survive" (IEEE Spectrum, 1 June 2016), [spectrum.ieee.org/automaton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive](http://spectrum.ieee.org/automaton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive), accessed 29 December 2017.
27. <sup>▲ [p.3]</sup> Statement of the US Deputy Defense Secretary Robert Work, in D. Lamothe, "Pentagon examining the "killer robot" threat", Washington Post, 30 March 2016. <https://www.bostonglobe.com/news/nation/2016/03/30/the-killer-robot-threat-pentagon-examining-how-enemies-could-empower-machines/sFri6ZDifwIcQR2UgyXIQI/story.html>, accessed 29 December 2017.
28. <sup>▲ [p.3]</sup> One need also mention that the Defense Advanced Research Projects Agency (DARPA, US agency responsible for the development of emerging military technologies) and the US Navy are very active in efforts to develop autonomous systems (M. Gubrud, "Why Should We Ban Autonomous Weapons? To Survive" (IEEE Spectrum, 1 June 2016), [spectrum.ieee.org/automaton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive](http://spectrum.ieee.org/automaton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive), accessed 29 December 2017).
29. <sup>▲ [p.3]</sup> US Air Force as cited in Paris Innovation Review, "Robotics Series - 2 - Can military robots adopt ethical standards?", Paris Innovation Review, 24 June 2014. [parisinnovationreview.com/articles-en/can-military-robots-adopt-ethical-standards](http://parisinnovationreview.com/articles-en/can-military-robots-adopt-ethical-standards), accessed 7 January 2018.
30. <sup>▲ [p.3]</sup> D. Lamothe, *op. cit.*
31. <sup>▲ [p.3]</sup> In 2017, Brazil, Uganda, and Iraq joined the countries that clearly positioned themselves as LAWs opponents. These include Algeria, Argentina, Bolivia, Chile, Costa Rica, Cuba, Ecuador, Egypt, Ghana, Guatemala, Holy See, Mexico, Nicaragua, Pakistan, Panama, Peru, State of Palestine, Venezuela, Zimbabwe.

32. <sup>^ [p.3]</sup> UK Ministry of Defence as cited in James Vincent, "UK government says humans will always be in charge of its robot weapon systems", *The Verge*, 12 September 2017, <https://www.theverge.com/2017/9/12/16286580/uk-government-killer-robots-drones-weapons>, accessed 29 December 2017. This is in continuity with the UK policy of 2013: HL Deb, 17 November 2014, c92W.
33. <sup>^ [p.3]</sup> On possible interpretations of this phrase, see, for instance, H.M. Roff, "Meaningful Human Control or Appropriate Human Judgement? The Necessary Limits on Autonomous Weapons. Briefing Paper for delegates at the Review Conference of the Convention on Certain Conventional Weapons, Geneva, 12-16 December 2016" (Global Security, n.d.) [globalsecurity.asu.edu/sites/default/files/files/Control-or-Judgment-Understanding-the-Scope.pdf](http://globalsecurity.asu.edu/sites/default/files/files/Control-or-Judgment-Understanding-the-Scope.pdf), accessed 7 January 2018.
34. <sup>^ [p.3]</sup> ) Israel, Statement on Lethal Autonomous Weapons Systems (LAWS) Weapon Legal Review (UN Group of Experts meeting on Lethal Autonomous Weapon Systems (LAWS), Geneva, 11-15 April 2016), [www.unog.ch/80256EDD006B8954/\(httpAssets\)/5951D4CF7936ADE3C1257F9A004B62D6/\\$file/2016\\_LAWS +MX\\_ChallengestoIHL\\_Statements\\_Israel.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/5951D4CF7936ADE3C1257F9A004B62D6/$file/2016_LAWS+%X+ChallengestoIHL+Statements+Israel.pdf), accessed 7 January 2018.
35. <sup>^ [p.3]</sup> It is advisable that States reach an agreement on "meaningful human control" as a guiding principle in the negotiations for a new ad hoc additional protocol to the Geneva Conventions, by taking into account the best existing literature in the field, such as H.M. Roff and R. Moyes, "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons. Briefing paper for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) Geneva, 11-15 April 2016" (Article 36, n.d.), [www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf](http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf) accessed 7 January 2018. They underline that meaningful human control most operate before the war, during the war, and after the war; they stress that it is of the utmost importance that human commanders have meaningful human control over direct attacks.
36. <sup>^ [p.3]</sup> Jo Swinson (East Dunbartonshire) (LD), question for oral answer n. 902555, HC Deb 27 November 2016, vol 632.
37. <sup>^ [p.3]</sup> The Minister for the Armed Forces (Mark Lancaster), *ibid.*
38. <sup>^ [p.3]</sup> Article 36, The United Kingdom and lethal autonomous weapons systems (Article 36, April 2016) [www.article36.org/wp-content/uploads/2016/04/UK-and-LAWS.pdf](http://www.article36.org/wp-content/uploads/2016/04/UK-and-LAWS.pdf), accessed 7 January 2018.
39. <sup>^ [p.3]</sup> House of Commons Hansard Debates, 17 June 2013, vol. 564, cols. 734-735.
40. <sup>^ [p.3]</sup> D. Rogozin as cited by M. Gubrud, "Why Should We Ban Autonomous Weapons? To Survive" *IEEE Spectrum*, 1 June 2016, [spectrum.ieee.org/autoton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive](http://spectrum.ieee.org/autoton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive), accessed 29 December 2017.
41. <sup>^ [p.3]</sup> D. Gilbert, "Russian weapons maker Kalashnikov developing killer AI robots", *Vice*, 13 July 2017, [news.vice.com/en\\_us/article/vbzq8y/russian-weapons-maker-kalashnikov-developing-killer-ai-robots](http://news.vice.com/en_us/article/vbzq8y/russian-weapons-maker-kalashnikov-developing-killer-ai-robots), accessed 29 December 2017.
42. <sup>^ [p.4]</sup> Russian Federation, "Examination of various dimensions of emerging technologies in the area of lethal autonomous weapons systems, in the context of the objectives and purposes of the Convention" CCW/GGE.1/2017/WP.8 (Group of Governmental Experts of the High Contracting Parties to the 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, 10 November 2017) <https://admin.govexec.com/media/russia.pdf>, accessed 1 January 2018.
43. <sup>^ [p.4]</sup> See, for instance, H.C. Hutchinson, "Russia says it will ignore any UN ban of killer robots," *Business Insider*, 30 November 2017, <http://www.businessinsider.com/russia-will-ignore-un-killer-robot-ban-2017-11?IR=T>, accessed 1 January 2018.
44. <sup>^ [p.4]</sup> Russian Federation (n 41) paras I.4 and IV.12.
45. <sup>^ [p.4]</sup> *Ibid.*, para II.6. Similarly, in October 2017, in the context of the proceedings of the First Committee of the 72nd Session of the UN General Assembly (Disarmament and International Security), Russia found the topic of LAWS premature and observed that the Group of Governmental Experts should not "impose preventive limitations or prohibitions on this type of prospective weapons and relevant technologies." (The Russian Federation, Statement by the Representative of the Russian Federation to the First Committee of the 72nd Session of the UN General Assembly Vladimir Yermakov in the First Committee on "Conventional Weapons" Cluster (Research Critical Will, n.d.), [reachingcriticalwill.org/images/documents/Disarmament-fora/1com/1com17/statements/20Oct\\_Russia.pdf](http://reachingcriticalwill.org/images/documents/Disarmament-fora/1com/1com17/statements/20Oct_Russia.pdf) accessed 7 January 2018.
46. <sup>^ [p.4]</sup> *Ibid.*, para II.8.
47. <sup>^ [p.4]</sup> Moreover, this position is not completely isolated. Indeed, part of the literature claims that the "current state of LAWS technology provides insufficient data upon which to determine if LAWS will violate" the prohibition to cause unnecessary injury and the provision to distinguish between civilian and military targets. G. Bills, "LAWS unto Themselves: Controlling the Development and Use of Lethal Autonomous Weapons Systems" 83 *GEO. WASH. L. REV.* 176, at 177 (2014).

48. <sup>▲ [p.4]</sup> In recognising the need for a balance between human rights and defence interests, however, Russia concludes that "the need to address humanitarian concerns cannot be used as the one and only sufficient prerequisite for imposing restrictive and prohibitive regimes on certain weapons" (Russian Federation (n 41) para VII.14). This might be interpreted as meaning that in addition to humanitarian law requirements, LAWs will have to meet also other criteria. Hopefully, it will not be interpreted as meaning that customary international law can be breached for mere domestic interests.
49. <sup>▲ [p.4]</sup> See "Losing Humanity: The Case against Killer Robots", Human Rights Watch, 29 November 2012, <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>, accessed 27 January 2017.
50. <sup>▲ [p.4]</sup> In the steering committee of the Campaign to Stop Killer Robots sit Human Rights Watch, Article 36, Association for Aid and Relief Japan, International Committee for Robot Arms Control, Mines Action Canada, Nobel Women's Initiative, PAX, Pugwash Conferences on Science & World Affairs, Seguridad Humana en América Latina y el Caribe, and Women's International League for Peace and Freedom.
51. <sup>▲ [p.4]</sup> For further information on the campaign, see <http://www.stopkillerrobots.org>.<sup>[52 p.12]</sup>
52. <sup>▲ [p.12]</sup> <http://www.stopkillerrobots.org>
53. <sup>▲ [p.4]</sup> The first meeting was scheduled for April 2017, but it was postponed to August and then cancelled. It eventually took place on 13-17 November 2017.
54. <sup>▲ [p.4]</sup> See M. Gubrud, "Why Should We Ban Autonomous Weapons? To Survive" (IEEE Spectrum, 1 June 2016), [spectrum.ieee.org/autaton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive](http://spectrum.ieee.org/autaton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive), accessed 29 December 2017.
55. <sup>▲ [p.4]</sup> Most of the 86 countries that participated to the first (and only) meeting of said Group of Governmental Experts called for a legally binding instrument on fully autonomous weapons. It is noteworthy that this position was reached as a consequence of a strong statement from the Non-Aligned Movement, a group of 120 states neither for nor against any power bloc and that are united in the "struggle against imperialism, colonialism, neo-colonialism, expansionism, racism (...), and all forms and manifestations of foreign occupation, domination, and hegemony." (Non-Aligned Movement, Political Declaration (Sixth Conference of Heads of State or Government of Non-Aligned Countries, Havana, Cuba, 3-9 September 1979). See Bolivarian Republic of Venezuela on behalf of the Non-Aligned Movement (NAM) and Other States Parties, "General principles on Lethal Autonomous Weapons Systems", CCW/GGE.1/2017/WP.10 (*Group of Governmental Experts of the High Contracting Parties to the 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*, 13 November 2017), [www.unog.ch/80256EDD006B8954/\(httpAssets\)/A980151CB5E662D4C12581D80025D4F3/\\$file/2017\\_GGEonLAWS\\_WP9\\_NAM.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/A980151CB5E662D4C12581D80025D4F3/$file/2017_GGEonLAWS_WP9_NAM.pdf), accessed 4 January 2018. For more information on the meeting of the Group of Governmental Experts in 2017, see Campaign to Stop Killer Robots, "Support Grows for New International Law on Killer Robots" (Stop Killer Robots, 17 November 2017), <http://www.stopkillerrobots.org/2017/11/gge/>, accessed 27 December 2017.
56. <sup>▲ [p.4]</sup> Campaign to Stop Killer Robots, "Country Views on Killer Robots" (Campaign to Stop Killer Robots, 16 November 2017), [http://www.stopkillerrobots.org/wp-content/uploads/2013/03/KRC\\_CountryViews\\_16Nov2017.pdf](http://www.stopkillerrobots.org/wp-content/uploads/2013/03/KRC_CountryViews_16Nov2017.pdf), accessed 3 December 2018.
57. <sup>▲ [p.4]</sup> High Contracting Parties to the 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, Final Report (United Nations Office at Geneva, 24 November 2017) para 34, [www.unog.ch/80256EDD006B8954/\(httpAssets\)/8A3BE602D1E4142CC12581E70054D0F4/\\$file/CCW\\_MHCP+2017\\_FinalReport\\_Advance+Version+\(003\)\\_ES.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/8A3BE602D1E4142CC12581E70054D0F4/$file/CCW_MHCP+2017_FinalReport_Advance+Version+(003)_ES.pdf), accessed 4 January 2018. For more information on the annual meeting of the UN Convention on Conventional Weapons, see Campaign to Stop Killer Robots, "Unambitious process on killer robots to continue" (Campaign to Stop Killer Robots, 24 November 2017), [www.stopkillerrobots.org/2017/11/ccwun-3/](http://www.stopkillerrobots.org/2017/11/ccwun-3/), accessed 4 January 2018.
58. <sup>▲ [p.4]</sup> Chairperson of the Group of Governmental Experts on Lethal Autonomous Weapons Systems, "Note on the Organization of Work of the 2018 Meetings of the Group of Governmental Experts (GGE) related to emerging technologies in the area of lethal autonomous weapons systems (LAWS) in the context of the objectives and purposes of the CCW" (UN Convention on Conventional Weapons, 24 December 2017), [www.unog.ch/80256EDD006B8954/\(httpAssets\)/04934D14C57BBDF0C12581FE0064310F/\\$file/CCWChairs2018\\_Joint-letter.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/04934D14C57BBDF0C12581FE0064310F/$file/CCWChairs2018_Joint-letter.pdf), accessed 4 January 2018.
59. <sup>▲ [p.4]</sup> Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), adopted 8 June 1977, 1125 UNTS 3, entered into force 7 December 1978, art 36.
60. <sup>▲ [p.4]</sup> The following principles of international humanitarian law protect civilians: distinction, necessity and proportionality, humane treatment, and non-discrimination. W. Wallach, Toward a Ban on Lethal Autonomous Weapons: Surmounting the Obstacles, 60(5) COMMUNICATIONS OF THE ACM 28-34 fn b (2017).
61. <sup>▲ [p.4]</sup> Protocol I to the Geneva Conventions, arts 48, 51(2), and 52(2).

62. <sup>▲ [p.4]</sup> International Committee of the Red Cross, "Customary International Humanitarian Law: Rule 47. Attacks Against Persons Hors de Combat", [ihl-databases.icrc.org/customary-ihl/eng/docs/v1\\_cha\\_chapter15\\_rule47#\\_su](http://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_cha_chapter15_rule47#_su), accessed 29 December 2017.
63. <sup>▲ [p.4]</sup> "Losing Humanity: The Case against Killer Robots", Human Rights Watch, 29 November 2012, 32, <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>, accessed 27 January 2017.
64. <sup>▲ [p.4]</sup> S. Groves, The US Should Oppose the UN's Attempt to Ban Autonomous Weapons, 2996 BACKGROUNDER 1, 3 (The Heritage Foundation, 5 March 2015).
65. <sup>▲ [p.4]</sup> It is possible to disagree, therefore, with B. Kastan, "Autonomous Weapons Systems: A Coming Legal 'Singularity'?" 81 U. ILL. J. L. TECH. POL. 45, 56 (2013), according to which a proportionality analysis is unnecessary if the military target is in a deserted location.
66. <sup>▲ [p.4]</sup> "Losing Humanity: The Case against Killer Robots", Human Rights Watch, 29 November 2012), <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>, accessed 27 January 2017.
67. <sup>▲ [p.4]</sup> S. Groves, The US Should Oppose the UN's Attempt to Ban Autonomous Weapons, 2996 BACKGROUNDER 1, 3 (The Heritage Foundation, 5 March 2015) - believes that technology will evolve in a way to allow the compliance with the principle of distinction. It makes the example of the Dual Mode Brimstone anti-armour missile, which strikes only armoured vehicles that match a signature and ignores other vehicles. One can put forward at least two counter-arguments. First, this kind of technology does not work as well with people, where understanding intention is crucial and requires a proper artificial consciousness. Second, the implementation of the principle of distinction requires a holistic and contextual understanding of the circumstances; therefore, for instance, a human-controlled system will be able to understand that a tank used in an anti-war demonstration organised by an NGO is not a legitimate target. An automated system may fail to appreciate the difference.
68. <sup>▲ [p.4]</sup> Future of Life Institute, "Autonomous Weapons: An Open Letter from AI & Robotics Researchers" (FLI, 28 July 2015), <https://futureoflife.org/open-letter-autonomous-weapons/>, accessed 29 December 2017. These researchers fear that soon LAWs would "appear on the black market and in the hands of terrorists, dictators wishing to better control their populace, warlords wishing to perpetrate ethnic cleansing."
69. <sup>▲ [p.4] [p.5]</sup> "Losing Humanity: The Case against Killer Robots", Human Rights Watch, 29 November 2012, <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>, accessed 27 January 2017.
70. <sup>▲ [p.4]</sup> G. Bills, LAWS unto Themselves: Controlling the Development and Use of Lethal Autonomous Weapons Systems, 83 GEO. WASH. L. REV. 176, at 199 (2014).
71. <sup>▲ [p.4]</sup> Protocol I to the Geneva Conventions, art 51(5)(b).
72. <sup>▲ [p.5]</sup> S. Groves, "The US Should Oppose the UN's Attempt to Ban Autonomous Weapons" 2996 BACKGROUNDER 1, 4 (The Heritage Foundation, 5 March 2015), italics added.
73. <sup>▲ [p.5]</sup> G. Bills, "LAWS unto Themselves: Controlling the Development and Use of Lethal Autonomous Weapons Systems" 83 GEO. WASH. L. REV. 176, at 199 (2014).
74. <sup>▲ [p.5]</sup> W.H. Boothby, *Weapons and the Law of Armed Conflict* (Oxford University Press 2009) 79.
75. <sup>▲ [p.5]</sup> See the pioneering Ray Kurzweil, *The Singularity is Near* (Penguin 2005).
76. <sup>▲ [p.5]</sup> M. Schmitt, "Military Necessity and Humanity in International Humanitarian Law: Preserving the Delicate Balance" (2010) 50(4) VIRGINIA J. INT'L LAW 795, 835.
77. <sup>▲ [p.5]</sup> A. Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Ashgate 2009) 91.
78. <sup>▲ [p.5]</sup> See B. Kastan, "Autonomous Weapons Systems: A Coming Legal 'Singularity'?" 81 U. ILL. J. L. TECH. POL. 45, 59 (2013).
79. <sup>▲ [p.5]</sup> It has been noted that once LAWs "are widely introduced, it becomes a matter of military necessity to use them, as they could prove far superior to any other type of weapon" (A. Krishnan, *op. cit.* at 91).
80. <sup>▲ [p.5]</sup> In its advisory opinion in Legality of the Threat or Use of Nuclear Weapons [1996] ICJ 2, the International Court of Justice stated that the principle of distinction was one of the "cardinal principles" of international humanitarian law and one of the "intransgressible principles of international customary law." (§ 434).
81. <sup>▲ [p.5]</sup> International Committee of the Red Cross, "Customary International Humanitarian Law: Rule 14. Proportionality in attack" [ihl-databases.icrc.org/customary-ihl/eng/docs/v2\\_cha\\_chapter4\\_rule14](http://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_cha_chapter4_rule14), accessed 29 December 2017.
82. <sup>▲ [p.5]</sup> The United States signed Protocol I on 12 December 1977, but never ratified it. This is also the case with the Islamic Republic of Iran and with Pakistan. The list of parties and signatories can be found at [ihl-databases.icrc.org](http://ihl-databases.icrc.org)

[/applic/ihl/ihl.nsf/Treaty.xsp?action=openDocument&documentId=D9E6B6264D7723C3C12563CD002D6CE4](http://applic/ihl/ihl.nsf/Treaty.xsp?action=openDocument&documentId=D9E6B6264D7723C3C12563CD002D6CE4)<sup>[83 p.14]</sup>, accessed 7 January 2018. cf G. Bills, "LAWS unto Themselves: Controlling the Development and Use of Lethal Autonomous Weapons Systems" 83 GEO. WASH. L. REV. 176, at 192 (2014).

83. <sup>▲ [p.13]</sup> [ihl-databases.icrc.org/applic/ihl/ihl.nsf/Treaty.xsp?action=openDocument&documentId=D9E6B6264D7723C3C12563CD002D6CE4](http://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Treaty.xsp?action=openDocument&documentId=D9E6B6264D7723C3C12563CD002D6CE4)
84. <sup>▲ [p.5]</sup> See H.L. Dreyfus and S.E. Dreyfus, *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer* (Free Press, 1986).
85. <sup>▲ [p.5]</sup> Groves accepts that "(e)ach such strike would require evaluation on a case-by-case basis," but he seems to think – though it does not present any convincing argument – that an autonomous system would be able to carry out such an evaluation. S. Groves, *The US Should Oppose the UN's Attempt to Ban Autonomous Weapons*, 2996 BACKGROUNDERS 1, 4 (The Heritage Foundation, 5 March 2015).
86. <sup>▲ [p.5]</sup> In order to assess in advance the damage to civilians or civilian property that is likely to result from an attack, some militaries use the so-called collateral damage estimation methodology", a statistical analysis that could be conducted by software (G. Bills, *LAWS unto Themselves: Controlling the Development and Use of Lethal Autonomous Weapons Systems*, 83 GEO. WASH. L. REV. 176, at 195 (2014)). However, this software cannot determine who is a civilian or of the attack is necessary. Moreover, a one-size-fits-all approach (e.g., "if the damage to property is below £1000, then deliver force") would be inappropriate because there are a number of illegal targets, regardless of the amount of foreseeable damage, e.g., because of their cultural or medical significance.
87. <sup>▲ [p.5]</sup> See, for instance, Protocol I to the Geneva Conventions, art 1(2) and Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II), adopted 8 June 1977, 1125 UNTS 609, entered into force 7 December 1978, fourth paragraph of the preamble. Under these provisions, "in cases not covered by the law in force, the human person remains under the protection of the principles of humanity and the dictates of the public conscience."
88. <sup>▲ [p.5]</sup> Y. Sandoz *et al.* (eds.), "Commentary on the Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (Protocol I)", in Y. Sandoz *et al.* (eds.), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (Martinus Nijhoff, 1987).
89. <sup>▲ [p.5]</sup> *Ibid.*, at 39.
90. <sup>▲ [p.5]</sup> C. Zoli, "Humanizing irregular warfare: Framing compliance for nonstate armed groups at the intersection of security and legal analysis", in W. Banks (ed.), *New Battlefields/Old Laws: Critical Debates on Asymmetric Warfare* (Columbia University Press, 2011) 199.
91. <sup>▲ [p.5]</sup> Some data has already been collected and analysed, but there is a need for more up-to-date studies. For instance, in 2013 it was found that people in the United States are against using LAWs by 2-to-1 (C. Carpenter, "How scared are people of "killer robots" and why does it matter?", *Open Democracy*, 4 July 2013, <https://www.opendemocracy.net/charli-carpenter/how-scared-are-people-of-killer-robots-and-why-does-it-matter>, accessed 29 December 2017).
92. <sup>▲ [p.5]</sup> W.H. Boothby, *Weapons and the Law of Armed Conflict* (Oxford University Press 2009) 232.
93. <sup>▲ [p.5]</sup> With regards to military robots, it has been noted that the machine's actions are a reflection of its programmer, its manufacturer and its operator (US Air Force as cited in Paris Innovation Review, "Robotics Series - 2 - Can military robots adopt ethical standards?", Paris Innovation Review, 24 June 2014. [parisinnovationreview.com/articles-en/can-military-robots-adopt-ethical-standards](http://parisinnovationreview.com/articles-en/can-military-robots-adopt-ethical-standards), accessed 7 January 2018). As pointed out by A. Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Ashgate 2009) 104, manufacturers of military technologies are rarely considered accountable for malfunctions of their weapons, therefore product liability regimes are not fit for purpose. Bills seems to believe that commander responsibility could provide an answer in three scenarios. First, if the commander knew the LAW was capable of violating the law of war. Second, if the commander should have known that the LAW was capable of violating those laws. Third, if the LAW was used in violation of the laws of war and the commander did nothing to hold responsible parties accountable. G. Bills, "LAWS unto Themselves: Controlling the Development and Use of Lethal Autonomous Weapons Systems" 83 GEO. WASH. L. REV. 176, at 197 (2014).
94. <sup>▲ [p.5]</sup> Perhaps, this could lead to the realisation of the prediction according to which there will be a "future when people blame robots for their actions" (T. Hellström, "On the moral responsibility of military robots" 15(2) ETHICS & INF'N TECH. 99 (2013)).
95. <sup>▲ [p.5]</sup> Along similar lines, it has been observed that the strong emotional bond between the human soldier and their robot decreases the former's effectivity, thus rendering killings less likely. See J. Carpenter, "Just doesn't look right: Exploring the impact of humanoid robot integration into Explosive Ordnance Disposal Teams", in R. Luppacini (ed.), *Handbook of Research on Technoself: Identity in a Technological Society* (Information Science Publishing, 2013) 609.

96. <sup>▲ [p.5]</sup> On the ethical issues related to the decrease in disincentives to start wars, related to the deployment of military robots, see N. Sharkey, "Killing Made Easy: From Joysticks to Politics" in P. Lin et al. (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press, 2011) 111, 122.
97. <sup>▲ [p.5]</sup> W. Wallach, *Toward a Ban on Lethal Autonomous Weapons: Surmounting the Obstacles*, 60(5) COMMUNICATIONS OF THE ACM 28, at 29 (2017).
98. <sup>▲ [p.5]</sup> P. Kahn, *The paradox of riskless war*, 22(3) PHIL. & PUBLIC POLICY Q. 2 (2002) puts forward a strong argument against military robots because civilians become more at risk from terrorist attacks.
99. <sup>▲ [p.6]</sup> As cited in N. Snow, *Information War: American Propaganda, Free Speech and Opinion Control Since 9-11* (Seven Stories, 2011) 99.
100. <sup>▲ [p.6]</sup> M. Armstrong, "Unintended consequences of unmanned warfare" (Proteus Management Group Futures Workshop, Carlisle Barracks, 15 August 2007).
101. <sup>▲ [p.6]</sup> LAWs require complex software and "(c)omplex software will create security risks and will make AWS critically vulnerable to hacking." (M. Klincewicz, "Autonomous Weapons Systems, the Frame Problem and Computer Security" 14 (2) JOURNAL OF MILITARY ETHICS 162 (2015)).
102. <sup>▲ [p.6]</sup> According to C. Perrow, *Normal accidents: Living with high risk technologies* (Princeton University Press, 2011), unintended interactions and their enablers (interactive complexity and tight coupling) are the primary causes of disasters in all engineering fields.
103. <sup>▲ [p.6]</sup> M. Gubrud, "Why Should We Ban Autonomous Weapons? To Survive" (*IEEE Spectrum*, 1 June 2016), [spectrum.ieee.org/automaton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive](http://spectrum.ieee.org/automaton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive), accessed 29 December 2017.
104. <sup>▲ [p.6]</sup> The problem of unintended consequences is part of a larger problem of unpredictability in the LAWs' behaviour. For instance, it has been noted that complex adaptive systems act unpredictably, "have tipping points that lead to fundamental reorganization, and can even display emergent properties that are difficult, if not impossible, to explain." W. Wallach, "Toward a Ban on Lethal Autonomous Weapons: Surmounting the Obstacles" 60(5) COMMUNICATIONS OF THE ACM 28, at 31 (2017).
105. <sup>▲ [p.6]</sup> S. Groves, *The US Should Oppose the UN's Attempt to Ban Autonomous Weapons*, 2996 BACKGROUNDER 1 (The Heritage Foundation, 5 March 2015).
106. <sup>▲ [p.6]</sup> See, for instance, V. Kanwar, *Post-Human Humanitarian Law: The Law of War in the Age of Robotic Weapons*, 2 HARV. NAT'L SEC. J. 616, 625 (2011); G. Bills, *LAWs unto Themselves: Controlling the Development and Use of Lethal Autonomous Weapons Systems*, 83 GEO. WASH. L. REV. 176, 197 (2014).
107. <sup>▲ [p.6]</sup> This is one of the points of the clever plan laid out by W. Wallach, *Toward a Ban on Lethal Autonomous Weapons: Surmounting the Obstacles*, 60(5) COMMUNICATIONS OF THE ACM 28, 30 (2017).
108. <sup>▲ [p.6]</sup> L. Floridi, *Tolerant paternalism: Pro-ethical design as a resolution of the Dilemma of toleration*, 22(6) Science and engineering ethics 1669, 1676 (2016)
109. <sup>▲ [p.6]</sup> T. Hellström, "On the moral responsibility of military robots" 15(2) ETHICS & INF'N TECH. 99 (2013) 99.
110. <sup>▲ [p.6]</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC ("General Data Protection Regulation") [2016] OJ L 119.
111. <sup>▲ [p.6]</sup> General Data Protection Regulation, art. 25
112. <sup>▲ [p.6]</sup> Technically, unless one takes a natural law approach, ethics is separate from law and when the law incorporates a value (such as privacy), this no longer belongs to the "ethics realm" and moves to the legal one. However, from the present point of view, what matters is not the subsuming of privacy under ethics, but the process of embedding values into the design of autonomous systems.
113. <sup>▲ [p.6]</sup> ) R.C. Arkin, "Governing lethal behavior: embedding ethics in a hybrid deliberative/reactive robot architecture, *Proceedings of the 3rd ACM/IEEE international conference on human robot interaction* 121 (2008). See also R.C. Arkin, *Ethical Robots in Warfare* 28(1) IEEE TECHNOLOGY AND SOCIETY MAGAZINE 30-33 (2009) and R.C. Arkin, *Governing Lethal Behavior in Autonomous Systems* (Chapman-Hall, 2009).
114. <sup>▲ [p.6]</sup> R.C. Arkin, *Governing Lethal Behavior in Autonomous Systems* (Chapman-Hall, 2009) 7.
115. <sup>▲ [p.6]</sup> This explains, for instance, why there are so many different notions of algorithmic fairness. See S. Friedler et al., "On the (im)possibility of fairness", 23 September 2016, <https://arxiv.org/pdf/1609.07236.pdf>, accessed 7 January 2018.

116. <sup>▲ [p.6]</sup> Code is law, according to Lawrence Lessig, *Code. Version 2.0* (Basic Books, 2006) 5, who accepts, however, that code is not exactly the same thing as the law, because the latter is the "self-conscious regulation, reflecting values of the society imposing that regulation" (*ibid.* at 5). Interestingly, with the embedment of ethics into the design, this difference would become less relevant.
117. <sup>▲ [p.6]</sup> A more practical reason is that a previous draft of this paper was presented to the SLS 2017 Conference, whose theme was "The diverse unities of law." Indeed, Kelsen developed a doctrine of unity with regards to the relationship between international and domestic law and then applied a version thereof to the pair "law and ethics." Even if some critiques may be moved to Kelsen's ideas on the matter, their core seems to retain its validity. In critical terms, for instance, H.L.A. Hart, "Kelsen's Doctrine of the Unity of Law", in H.E. Kiefer and M.K. Munitz (eds.), *Ethics and Social Justice* (New York State University Press, 1970), 171, who nonetheless considers the doctrine of unity as "one of the most striking doctrines" in Kelsen's thought (*ibid.* at 171). See also H.L.A. Hart, Kelsen Visited 10 UCLA L. REV. 709 (1963).
118. <sup>▲ [p.6]</sup> H. Kelsen, *General Theory of Law and State* (A. Wedberg tr., Harvard University Press, 1945) 363, 410-411; H. Kelsen, *Pure Theory of Law* (Max Knight tr., 2nd ed., University of California Press, 1967) 328.
119. <sup>▲ [p.6]</sup> As an example of an ostensible conflict between a legal duty and a moral one, Kelsen gives the example of the mandatory military service. According to Kelsen, there is no conflict because for the lawyer there is only the duty to perform military service and no contrary duty, to the point that "morality does not exist as such (...) does not count at all as a system of valid norms (Kelsen, *General Theory of Law and State*, at 374). For the moralist, in turn, the obligation to perform military service is morally irrelevant. Neither of them asserts that both systems are valid. H.L.A. Hart, Kelsen Visited 10 UCLA L. REV. 709, 728 (1963) observes that between law and morals there can be conflict without contradictions and reports that, after the publication of his pioneering books, Kelsen was "considering afresh the (...) possibility that one norm might logically conflict with another."
120. <sup>▲ [p.7]</sup> G. Radbruch, *Gesetzliches Unrecht und übergesetzliches Recht* 1(5) *Süddeutsche Juristen-Zeitung* 105 (1946), reprinted in G. Radbruch, *Gesetzliches Unrecht und übergesetzliches Recht* (tr. Pauer-Studer, *Nomos* 2002) 10, as cited in H. Pauer-Studer, *Law and Morality under Evil Conditions: The SS Judge Konrad Morgen*, 3(2) *JURISPRUDENCE* 367, 389 (2012), fn 91.
121. <sup>▲ [p.7]</sup> H. Kelsen, *Pure Theory of Law* (Max Knight tr., 2nd ed., University of California Press, 1967) 67.
122. <sup>▲ [p.7]</sup> H. Kelsen, *General Theory of Law and State* (A. Wedberg tr., Harvard University Press, 1945) 5.
123. <sup>▲ [p.7]</sup> H. Kelsen, *Pure Theory of Law*, *op.cit.* at 69.
124. <sup>▲ [p.7]</sup> H. Kelsen, *General Theory of Law and State*, *op. cit.* at 5.
125. <sup>▲ [p.7]</sup> H. Pauer-Studer, "Kelsen's Legal Positivism and the Challenge of Nazi Law" in *Yearbook Vienna Circle Institute* 17 (Springer, 2014) 223. It is noteworthy that, in Himmler's plan, the SS and the police's jurisdiction had to be liberated from the Roman legal thought and based on "a Germanic sense of justice, in accordance with the nature of Nazism. Therefore, liberating the 'judge' from the 'inflexible framework of the law' (C. Theel, "The Moral Rigour of Immorality: The Special Criminal Courts of the SS", in W. Bialas and L. Fritze (eds), *Nazi Ideology and Ethics* (Cambridge Scholars, 2014) 343).
126. <sup>▲ [p.7]</sup> Ultimately, what Kelsen claims is that there is no scientific answer to the question "what is moral" and that whether or not a legal system *should* be moral is not a legal problem, but a political one. Kelsen goes as far as affirming that the requirement that positive law should be just "is self-evident" (Kelsen, *General Theory of Law and State*, *op. cit.* at 5).
127. <sup>▲ [p.7]</sup> It is worthy underlining that the doctrine of unity was used by Kelsen to explain the relationship between law and morals in H. Kelsen, *General Theory of Law and State* (A. Wedberg tr., Harvard University Press, 1945), but it was abandoned in H. Kelsen, *Pure Theory of Law* (Max Knight tr., 2nd ed., University of California Press, 1967). However, the core ideas on said relationship are the same.
128. <sup>▲ [p.7]</sup> L. Floridi, *Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration*, 22 *SCI. ENG. ETHICS* 1669, 1670 (2016).
129. <sup>▲ [p.7]</sup> *Ibid.* at 1681.
130. <sup>▲ [p.7]</sup> This is not to say that the law can crystallise an absolute ideal of justice (which can be regarded as the synthesis of all moral values in law). As observed by H. Kelsen, *What is justice?* (Berkeley: University of California Press, 1957) 22, justice as social happiness could be achieved only if everyone's desires and interests are fulfilled. However, this is not possible due to the scarcity of resources and to the pluralism of interests and desires. Therefore, the democratically-produced law is only the tool to achieve the minimum goal of sorting out the conflict of interests through compromise.
131. <sup>▲ [p.7]</sup> K. O'Hara, *Smart contracts – Dumb idea*, *THE DIGITAL CITIZEN* 2, 5 (March/April 2017).
132. <sup>▲ [p.7]</sup> *General Data Protection Regulation*, recital 78.

133. <sup>▲ [p.7]</sup> European Parliament, Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).
134. <sup>▲ [p.7]</sup> Consolidated Version of the Treaty on the European Union [2008] OJ C115/13.
135. <sup>▲ [p.7]</sup> Charter of Fundamental Rights of the European Union [2012] OJ C326/391.
136. <sup>▲ [p.7]</sup> For instance, justice may require paternalism which is contrary to autonomy.
137. <sup>▲ [p.7]</sup> One need only think of the so-called trade-off between transparency and accuracy. On the one hand, modelers tend to develop more accurate models "with increasingly complex, data-mining-based black-box models." (I. Kamwa *et al.*, On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based catastrophe predictors, 3(1) IEEE TRANSACTIONS ON SMART GRID 152 (2012)) On the other hand, model users tend to favour "transparent, interpretable models not only for predictive decision-making but also for after-the-fact auditing and forensic purposes." (*Ibid.* at 152).
138. <sup>▲ [p.7]</sup> These theories are empiricist and relativist. The here-presented argument, however, could not be put forward by those who take a Kantian non-empirical universalistic approach to ethics. In Kant's ethical universalism, the categorical imperative acts on all people, regardless of their interests or desires. Its universal character (and corresponding refutation of moral relativism) and its independence to human attributes might lend themselves to a translation of the categorical imperative into code or design. This said, the other arguments against ethics by design would still hold true even from a Kantian perspective. The categorical imperative was introduced by Immanuel Kant, *Grundlegung zur Metaphysik der Sitten* (Hartknoch 1786). Cf T N Pelegrinis, *Kant's conceptions of the categorical imperative and the will* (Zeno 1980).
139. <sup>▲ [p.7]</sup> D. Hume, *An Enquiry concerning the Principles of Morals* (1751, Beauchamp ed, Oxford University Press, 1998).
140. <sup>▲ [p.7]</sup> J.S. Mill, *Utilitarianism* (Parker, Son and Bourn 1863) ch IV. See Necip Fikri Alican, *Mill's Principle of Utility: A Defense of John Stuart Mill's Notorious Proof* (Rodopi 1994) 67.
141. <sup>▲ [p.7]</sup> H. Kelsen, *What is Justice?* (Berkeley: University of California Press, 1957) 2. As such, social happiness or absolute justice cannot be achieved because it "is inevitable that an individual's happiness, sooner or later, will clash with the one of another individual" (H. Kelsen, *General Theory of Law and State* (A. Wedberg tr., Harvard University Press, 1945) 6)
142. <sup>▲ [p.7]</sup> The story of IBM Watson beating human champions at the game "Jeopardy" received ample media coverage. See e.g., A. Gabbatt, "IBM computer Watson wins Jeopardy clash", *The Guardian*, 17 February 2011, [www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy](http://www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy) accessed 7 January 2018.
143. <sup>▲ [p.8]</sup> L. Pirandello, *L'umorismo* (Mondadori, 2013) uses this metaphor to explain the difference between what is comic (in Pirandello's terms this is the "perception of the contrary", the laugh at the sight of the lady) and what is humorous (the subsequent bitter reflection, the "sentiment of the contrary").
144. <sup>▲ [p.8]</sup> For more information on this argument, see G. Noto La Diega, "The European strategy on robotics and artificial intelligence: too much ethics, too little security" 3(2) EUR'N CYBERSECURITY J. 6 (2017).
145. <sup>▲ [p.8]</sup> In his State of the Union Address, the President of the European Commission announced the creation of the European Cybersecurity Agency. See European Commission, "State of the Union 2017 - Cybersecurity: Commission scales up EU's response to cyber-attacks", 19 September 2017, [europa.eu/rapid/press-release\\_IP-17-3193\\_en.htm](http://europa.eu/rapid/press-release_IP-17-3193_en.htm), accessed 7 January 2018.
146. <sup>▲ [p.8]</sup> G. Contissa *et al.*, The Ethical Knob: ethically-customisable automated vehicles and the law, 25 AI & LAW 365 (2017).
147. <sup>▲ [p.8]</sup> N. Sharkey, "March of the killer robots", *Daily Telegraph*, 15 June 2009.
148. <sup>▲ [p.8]</sup> Protocol Additional to the Convention on Prohibitions or Restrictions on the Use of Certain Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects ("Protocol IV"), adopted 13 October 1995, TIAS 09-721.2, banned the development of lasers that were specifically designed to cause permanent blindness. Positively, in its first position on the matter, China declared its support for "the development of a legally binding protocol on issues related to the use of LAWS, similar to the Protocol on Blinding Laser Weapons, to fill the legal gap in this regard." (The People's Republic of China, "The position paper submitted by the Chinese delegation to CCW 5th Review Conference" (United Nations Office at Geneva, n.d.), [www.unog.ch/80256EDD006B8954/\(httpAssets\)/DD1551E60648CEBBC125808A005954FA/\\$file/China's+Position+Paper.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/DD1551E60648CEBBC125808A005954FA/$file/China's+Position+Paper.pdf), accessed 7 January 2018). However, in November 2017, the position appears watered down in The People's Republic of China, "Statement by the Chinese Delegation at the Thematic Discussion on Conventional Weapons at the First Committee of the 72nd Session of the UNGA" (Research Critical Will, n.d.), [reachingcriticalwill.org/images/documents/Disarmament-fora/1com/1com17/statements/20Oct\\_China.pdf](http://reachingcriticalwill.org/images/documents/Disarmament-fora/1com/1com17/statements/20Oct_China.pdf), accessed 7 January 2018.