Article36

POLICY NOTE | NOV 2019

Critical Commentary on the "Guiding Principles"

A critical commentary on the "Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System" as adopted by the 2019 Meeting of High Contracting Parties to the Convention on Conventional Weapons (CCW)

Article 36 is a UK-based not-for-profit organisation working to promote public scrutiny over the development and use of weapons.*

www.article36.org
info@article36.org
@Article36

* This paper was written by Richard Moyes.

The final report of the Meeting of High Contracting Parties to the CCW in 2019 saw the adoption, at Annex III of "*Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System*[s]".[1] These principles had been under development since 2018, and the version adopted in 2019 added only a single additional paragraph (paragraph 'c') to the 'Possible Guiding Principles' finalised at the end of 2018.

These 'Guiding Principles' have also been embraced by the so-called 'Alliance for Multilateralism' under which France and Germany framed them under 'a political declaration' referred to as '11 Principles on Lethal Autonomous Weapons Systems'.[2] Such a reframing enables those states to argue that they have fulfilled previous assertions that they would work for a political declaration on this issue, but the content of the text remains the same.

The sections below provide a paragraph-by-paragraph commentary on the content of the 2019 Guiding Principles. The paragraphs of the original document are presented in bold followed by analysis of some of the key issues that the relevant paragraphs raise.

***

It was affirmed that international law, in particular the United Nations Charter and international humanitarian law (IHL) as well as relevant ethical perspectives, should guide the continued work of the Group. Noting the potential challenges posed by emerging technologies in the area of lethal autonomous weapons systems to IHL, the following were affirmed, without prejudice to the result of future discussions:

This introductory paragraph highlights 'the continued work of the Group' and so makes it clear that the Guiding Principles are not an end in themselves. The Principles are framed here as guidance for the conduct of further work, not as an intended structure or product for the work (which they would themselves be guiding). This orientation is reiterated in principle [i] and then [j].

The paragraph highlights ethical considerations as significant in a context where such considerations can sometimes be marginalised in the debate. It also recognises that technologies in this area present potential challenges to IHL. This recognition of 'potential challenges' comes prior to the 'without prejudice' clause and frames the subsequent principles. Yet it is noticeable, here and elsewhere in the text, that international human rights law (IHRL) is not recognised despite it being a fundamental legal framework when weapons are being used other than as means or methods of warfare.

(a) International humanitarian law continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems;

This line reinforces the exclusion of IHRL considerations noted above. In itself it adds nothing – as it can be taken for granted that insofar as weapon systems are being used during an armed conflict and for the conduct of hostilities then IHL applies. Yet by omitting reference to the contextual limitations on where and when IHL applies the 'principle' suggests that it is promoting something comprehensive when in reality its effect is to limit the discourse.

That assertions like this can become a comfortable mantra for states is an example of the erosive capacity of the CCW and similar multilateral frameworks. No state actually argued in the informal discussions or the GGE that IHL did not apply to weapon technologies in this area. Quite the opposite, many noted that it did – perhaps simply relieved to have something straightforward and incontestable to write into their own statements. By contrast, articulating the more complex (and somewhat contested) boundaries of where and when different legal regimes apply would take up significantly more statement time and risk opening up disputes that are not tied to the subject matter at hand. So a 'straw man' is defeated, and what was not said recedes still further from consideration.

(b) Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system;

Given the debate in the CCW, and elsewhere on this issue, there is utility to affirming that accountability cannot be transferred to machines. It is also, arguably, something that should be taken for granted. However, various commentators on this issue, including some NGOs opposed to autonomy in weapons systems, have slipped into (and continue to slip into) implying the attribution of human responsibilities to machines (such as arguing that 'autonomous weapons systems will not be able to apply the rules on proportionality', when those rules are obligations on humans, and so the argument against 'capability' erroneously implies a transfer of responsibility.) The line regarding the 'life cycle of the weapons system' adds little without further elaboration of the potential for greater autonomy in weapons systems to strain the relationship between 'responsibility' and 'accountability' for human commanders and for the institutions (including states) under which they are operating.

(c) Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular IHL. In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system as a whole;

This principle was added to the 2018 'possible guiding principles' text based on subsequent proposals and discussions. It is a useful addition in that it ties 'human-machine interaction' to 'compliance with applicable international law'. In the second sentence it recognises that contextual factors and technical factors both bear upon how human-machine interaction is modulated, with an implication that this bears upon achieving compliance with the law.

The paragraph still provides little guidance on how 'human-machine interaction' is to be understood (beyond the bare fact that it will occur and that it 'should ensure' potential use is in compliance with the law.) This usefully ties to the preceding paragraph (b) regarding human responsibility. The second sentence implies that different forms or human-machine interaction might be needed depending on the system and the operational context. The reference to operational context suggests that, whilst human-machine interaction may be implemented at 'various stages of the life cycle' at least some of that interaction should occur in a situation where consideration of the operational context is possible. How broadly or narrowly 'operational context' is to be understood, however, also remains an open question. As such this could relate to interaction in the context of a specific operation use, or operation context could be construed very broadly (as in, the system will be used at sea).

Ultimately, the paragraph engages with central issues regarding how human control is to be ensured in the context of increasing autonomy in weapons systems. However, it does not provide significant guidance on how that is to be done, just that human interaction is linked to legal compliance, and there are contextual and technical factors that might bear upon that. As such it is a useful addition, but only as a placeholder for more substantial work elsewhere.

(d) Accountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control;

This paragraph is awkwardly worded but can be read as a general assertion that ensuring accountability is a function of the human structures within which any weapons systems are embedded.

The reference to 'in the framework of the CCW' serves here as a reformulation of the 'limiting' effect noted earlier with the respect to the prioritisation of IHL and the disregard for IHRL (in principle [a]). Here the text rightly refers to 'applicable international law' but, by contextualising the assertion in relation to the CCW (which is primarily concerned with situations of hostilities in armed conflict) the relevance of IHRL is again diminished.

Although rather convoluted, there is some utility to this assertion in so far as it deals with arguments that there would necessarily be an 'accountability gap.' Rather, the implication of this paragraph is that accountability gaps would be the product of problems in human management structures, not of a technology *per se*.

As with paragraph (b), it does not address a better formulated concern, specific to the issue at hand, that certain technologies could operate in a way that leaves a human commander accountable (under a chain of command – e.g. under the institutional authority of a state), but not reasonably responsible for the outcomes produced (due to a system being institutionally authorised to operate with excessive scope for action). This latter problem would also be a problem of the human management structure of course, for institutionally authorising an operational situation in which accountability and responsibility are not sufficiently tied together. However, a risk in the context of greater autonomy in weapons systems is that, despite the assertions of many states that their established structures will cope effectively, technological adoption progressively recalibrates expectations of the relationship between responsibility and accountability over time. That might involve accountability and responsibility becoming more distributed within the system, such that neither really fall anywhere at all. Working papers in the CCW by the UK and more recently by Australia both evoke this direction – seeking to subsume the issues raised around autonomy into a broad system of bureau-

cratic mechanisms, all of which are asserted to be effective but none of which actually answer the central questions.

(e) In accordance with States' obligations under international law, in the study, development, acquisition, or adoption of a new weapon, means or method of warfare, determination must be made whether its employment would, in some or all circumstances, be prohibited by international law;

This is a straightforward reassertion of an established treaty law obligation. By formulating the principle 'in accordance with State's obligations…' the text effectively reinforces the specific treaty law source of the obligation that is reasserted, thus avoiding an implication that this obligation, in full, has customary law status. The principle does not add anything of value.

(f) When developing or acquiring new weapons systems based on emerging technologies in the area of lethal autonomous weapons systems, physical security, appropriate non-physical safeguards (including cyber-security against hacking or data spoofing), the risk of acquisition by terrorist groups and the risk of proliferation should be considered;

This paragraph simply asserts a general requirement for states to manage access to their weapons systems, with some specific 'high-tech' issues thrown in. The same paragraph could apply to any weapons system. The introduction of a reference to 'terrorist groups' in a CCW document is a retrograde step. All parties to a conflict can commit crimes of 'terror' and bringing the political discourse and labels of terrorism into this legally-rooted architecture was unnecessary and establishes a problematic precedent for the future.

(g) Risk assessments and mitigation measures should be part of the design, development, testing and deployment cycle of emerging technologies in any weapons systems;

This paragraph explicitly relates itself to 'any weapons system' and thus it does little to engage with the specific concerns around autonomy in weapons systems. However, it arguably provides an unqualified reinforcement of principle [e], and the legal obligation of Additional Protocol I (art. 36) from which that is drawn, without being tethered to a treaty-law basis. States are accepting here a general commitment that 'risk assessments and mitigation measures should be' undertaken. It can be assumed that such assessments would include assessing whether systems would be unlawful in some or all circumstances. Yet the implication here could be that processes should also identify other sources of unpredictability, for example, within a system, such as factors that might affect the accuracy or precision of warhead delivery. More broadly, they could also include wider assessments of harms that might result from a type of system's use, not only in individual attacks (as is treated under the IHL) but cumulatively and in terms that are not captured by IHL's concept of harms. This is perhaps the only principle that, through its comparatively casual formulation, possibly opens up some productive space, not specifically in relation to autonomy in weapons systems, but more generally for the future.

(h) Consideration should be given to the use of emerging technologies in the area of lethal autonomous weapons systems in upholding compliance with IHL and other applicable international legal obligations;

This paragraph reflects a line, driven primarily by the USA but endorsed by others, that certain technological developments in weapons may be used to reduce risk to civilians, or risk to 'friendly forces', when compared with using established weapon technologies. Whilst this is technically reasonable (in our formulation of the narrative) it is important to note that our formulation of that narrative is presented in terms of reduction of risk and establishes explicitly the comparator of established weapon technologies. The 'principle', as formulated here, avoids any engagement with the problems that the technology is purportedly solving, and in doing so it tacitly endorses the idea of developments in weapons as actively extending civilian protection rather than as reducing risks.

The 'principle' here, is also problematic because it adopts a conceptualisation of legal 'compliance' that has subtle implications that are unhelpful for the stronger protection of civilians. Similar to discussion of 'accountability' above, 'upholding compliance with IHL…' is properly a function of the human management system. By contrast, 'use of…technologies…in upholding compliance with IHL…' reorders a sequence that should run: 'technologies must be used in compliance with…'

The implication here is that 'emerging technologies' would allow people some greater ability to comply with IHL. Yet, human actors always have sufficient ability to comply – after all they can choose not to use a weapon system at all – not to launch an attack, to suspend it or to change its parameters. The states that promote this 'compliance' formulation would not usually argue that at present (absent these emerging technologies) they have a problem with military commanders not complying with IHL.

The result, as noted, is an avoidance of engagement with the problems the technology purportedly addresses. States are not accepting that there is a problem of compliance with the law in current practice; that is why the term 'upholding' is chosen – it implies that current behaviour is compliant, and due to these emerging technologies, future behaviour will be compliant too…

Viewed in the context of the debate, this paragraph is primarily just a political counter to stand against suggestions that the ethical/humanitarian/legal implications of new technological developments are entirely negative. But still, in its formulation, it manages to bundle together a number of erosive formulations, regarding the role of weapons in society and the nature of legal compliance, that subtly protect weapons (in general) from criticism and states from legal responsibility.

(i) In crafting potential policy measures, emerging technologies in the area of lethal autonomous weapons systems should not be anthropomorphized;

This line returns to themes noted in relation to paragraphs (b) and (d). It probably adds little because the slippage to anthropomorphising, on the part of some, tends to happen accidentally – such as a tendency to talk about machines 'struggling to comply with the law'

due to their inability to evaluate proportionality, when it is humans who must evaluate proportionality. Although it sits rather awkwardly in the sequence of these principles, it is probably a useful reminder. Significantly, the principle also refers explicitly to 'crafting potential policy measures' – making it clear again that the Guiding Principles provide guidance in that context and are not an end in themselves.

(j) Discussions and any potential policy measures taken within the context of the CCW should not hamper progress in or access to peaceful uses of intelligent autonomous technologies;

Anxiety about a possible regulation of weapons systems curtailing socially productive technological developments has been raised by some delegations in the CCW.

It is unclear whether anxieties about 'hampering progress' are raised as a genuine (perhaps naïve) concern or are simply thrown into the conversation as an additional obstacle to progress. There is no doubt that regulation regarding autonomy in weapons systems can be drafted such that it avoids unintended constraints on non-weapon technologies. Representatives from the tech-sector express no concerns on this issue and previous instruments, such the prohibition of blinding laser weapons, have clearly had no impact on the development of civilian technologies. The idea that states will adopt a legal instrument only to find that they have accidentally prohibited sensors, computers and algorithms in general is implausible. The implication here, that simply 'discussions' in the CCW could affect technological 'progress', endorses a level of anxiety that is wildly out of step with reality.

The potential for a legal instrument to constrain 'access' to certain technologies may be a more reasonable concern. It is perhaps feared that certain technologies or capabilities might come to be considered 'dual-use' and thus subject to possibly politicised decision-making as to where they might be transferred. Again, the risks of this are wholly attendant upon how any instrument is drafted and so this aspect of the principle may provide useful guidance when states finally move on to the development of policy measures.

Rather than 'not hamper…', states parties should rather engage earnestly with questions about the social benefits expected of new technologies. They could question, in particular, the social benefits expected of 'weapon technologies' and military applications of advances in science and technology, and make concerted efforts to address social inequalities and promote sustainable security in line with the SDGs and the Universal Declaration of Human Rights. Again, the orientation from the CCW instead is defensive, limiting of itself and the discourse that it fosters.

Principle [j] is also notable in that it adopts the term 'intelligent autonomous technologies'. There is no explanation of what that term means or how it relates to other terminology used in the principles, such as 'emerging technologies in the area of lethal autonomous weapons systems'. Is the former a specific example from within the latter, or *vice versa*, or are the two terms contiguous?

As with principle [i], this principle also refers explicitly to 'potential policy measures' – further reinforcing that the Guiding Principles are not an end in themselves, and do not constitute 'policy measures' *per se*.

(k) The CCW offers an appropriate framework for dealing with the issue of emerging technologies in the area of lethal autonomous weapons systems within the context of the objectives and purposes of the Convention, which seeks to strike a balance between military necessity and humanitarian considerations.

The final principle is simply a return to the assertion that the CCW is an appropriate framework for dealing with the issue. However, it is notable that this formulation does not say that the CCW is *the* appropriate forum. Whilst its engagement with military and humanitarian considerations might be of utility, those factors are available to states in any forum – it is, after all, up to states to bring such considerations to the table themselves. Whether the CCW's mode of operation make it the appropriate forum for actually achieving a meaningful outcome on this issue is a very different question.

That a paragraph such as this is adopted as a principle by the CCW, asserting its own appropriateness, concludes the text on a flat note of insecurity. In the context of the piece as a whole, it is probably about right.

NOTES

1 See Revised Draft Final Report, CCW/MSP/2019/CRP.2/Rev.1, 15 Nov 2019, Geneva, https://www.unog.ch/80256EDD006B8954/(httpAssets)/815F8EE33B64DADDC12584B7004CF3A4/$file/CCW+MSP+2019+CRP.2+Rev+1.pdf

2 For context see: https://multilateralism.org/ The declaration text is available at: https://multilateralism.org/declaration-on-lethal-autonomous-weapons-systems-laws.pdf The principles remain the same as those analysed in this paper, under the following chapeaux:

> ***Declaration by the Alliance for Multilateralism on Lethal Autonomous Weapons Systems (LAWS)***
>
> *The Alliance for Multilateralism calls on states to actively sustain the crucialrole of arms control, disarmament and non-proliferation in securing peace and stability in our times and for future generations.*
>
> *In view of the rapid speed of technological development in the field of artificial intelligence (AI) the Alliance calls on states to give particular attention to the possible challenges associated with future weapons systems containing autonomous functions. It is of the utmost importance that the production, use and transfer of such future weapons are firmly grounded in International Law and state-control.*
>
> *We therefore call on states to contribute actively to the clarification and development of an effective and comprehensive normative and operational framework for lethal autonomous weapons systems (LAWS) by the designated group of governmental experts (GGE) within the UN Weapons Convention (CCW).*
>
> *We encourage states to promote the worldwide application of the eleven guiding principles as affirmed by the GGE and as attached to this declaration and to work on their further elaboration and expansion.*