In a speech to the 2019 session of the UN General Assembly, UK Prime Minister Johnson focused on the challenges and opportunities presented by developments in science and technology; musing that "A.I." might mean either "helpful robots washing and caring for an ageing population" or "pink eyed terminators sent back from the future to cull the human race." Faced with the horns of such a dilemma, he pledged that the UK would work to set "global principles to shape the norms and standards that will guide the development of emerging technology", building on its position as a (self-identified) "global leader" in this area.

The ‘pink eyed terminators’ comedically evoked in PM Johnson’s speech sign-post concerns regarding ‘autonomy’ in weapons systems – the potential for further developments in systems that apply force on the basis of sensors and machine calculations, with a concomitant diminution of human moral and legal judgement in such action. Standard media representations of this issue draw on the imagery of *The Terminator* – though the details are both more mundane and more imminent.

Since 2014, ‘autonomy’ has been the subject of multilateral discussions at the UN Convention on Conventional Weapons (CCW) under the rubric of ‘Lethal Autonomous Weapons Systems’. As an issue of active international deliberation, providing a nexus for debates on ethics, values, law, technology, military practice as well as industry and security concerns, autonomy in weapons systems presents an important test-case for the now widespread political mantras of state leadership in response to emerging socio-technological challenges. After all, for those claiming leadership in ‘ethical and responsible technology’, issues of how we kill each other would seem to be a good place to start.

However, it is political leadership that is now most noticeably lacking from the UK’s response to the issue of autonomy in weapons systems. UK policy orientations have evolved significantly over the last six years. It has moved away from dogmatic assertions on the meaning of certain terms¹ to a more open and interactive mode of engagement. Dispersed through their recent statements on this theme at the United Nations are the building blocks of a conceptually coherent and socially important orientation to human control over systems that apply force on the basis of sensors and machine calculations; building blocks from which a genuine global standard could come to be constructed if the UK had the political courage to lead.

As yet, the UK’s very significant points of content are presented in some isolation from each other – dotted across statements under different formal agendum, and each swaddled in a thick blanket of rhetoric on how new law is neither needed nor possible. Yet if this swaddling is put to one side, and the points of content consolidated, a policy position emerges that has the potential to draw together many actors in the international debate. In the sections below we draw out that picture of positive content, highlighting the points on which there is convergence between the UK’s statements and policy orientations that Article 36 would subscribe to.

We then consider briefly some of the key points that we set aside – the assertions about new law being unnecessary or politically impossible – and urge the UK to rise above the anxieties that motivate these assertions, and to focus on building recognition and engagement of the valuable content they have already tabled.

¹ This paper was written by Richard Moyes.
Points of convergence regarding content

The following points are drawn from different UK statements to the UN CCW in 2019 and are highlighted because they present possible key points of substance around which international opinion could usefully converge.

General principles
The following points relate to a general orientation to the subject matter:

- “Weapons are tools developed and used by humans to achieve certain effects. Therefore, the way in which humans interact with and control such systems is central to this debate.”
  - This is an important initial position that works against a tendency by some to anthropomorphise technologies. It serves to assert the centrality of human control as a parameter in the debate.

- “Robotic systems are often...described as ‘autonomous’ where they have long endurance, react effectively to external stimuli and require little or no human oversight for all or part of their mission...use of the term is frequently independent of the complexity of the system’s programming: even relatively simple mechanical devices can sometimes fulfill some parts of the definitions we have seen. Contemporary weapons with advanced sensors and guidance functions such as air-to-air or air-to-surface missiles or remotely operated systems conform even more closely to this model, whilst still stopping short of what might widely be accepted as fully autonomous.”
  - Shifting gently away from the UK’s earlier definitional orientations (in line with a shift of emphasis noticeable in the MoD Joint Concept Note on Human-Machine Teaming⁵), this paragraph situates the issues within a broad scope of technologies.
  - This usefully opens space to focus on how systems function and how they are used, rather than on searching for a single definitional boundary. Whilst this does not describe the same broad boundary of scope that Article 36 would encourage (systems that apply force on the basis of sensors, at a time and place that has not been specified by a human) it provides a good basis for constructive engagement with the subject matter.

- “…UK believes that a technology-agnostic approach which focusses on the importance of human control and the regulatory framework used to guarantee compliance with legal obligations is most productive ... this approach is likely to be more robust in the face of future technological developments and the highly diverse operational environments in which weapon systems are likely to be deployed and focusses attention on human decision making and targeting cycles.”
  - Continuing from the paragraph considered above, this orientation generally accords with Article 36’s assessment of the advantages of a broad approach to technologies – though we are not convinced that current regulatory frameworks ‘guarantee compliance’, or that simply ‘compliance’ with current law is sufficient to address the problems presented.

Factors for ensuring human control
The points below relate to factors that can be managed in order to exert human control:

- “[A]ccountability can never be delegated to a machine or system...[I]legal accountability will always devolve to a human being, never a machine – increasing autonomy in weapons or weapons systems does not therefore present the risk of an accountability gap.”⁹
  - Article 36 supports the principle of this basic position – as effectively a reaffirmation of the earlier point about weapons systems as ‘tools’ (not moral or legal agents).
  - However, the final statement is an oversimplification (which conveniently supports the UK’s political assertions that nothing needs to be done). The challenge presented by autonomy is not that a human cannot be asserted to be accountable under a chain of command. Rather it is that such an assertion of accountability may not be morally reasonable if that human is also systemically approved to use a technology over which they cannot in practice exert sufficient control. The UK’s statement that this is a non-issue is dependent upon a set of assumptions about how existing management systems (including constraints on weapon systems) will prevent such a situation from arising – assumptions that may not be valid or universal.

- “Specific constraints or parameters that might be placed on a machine’s freedom of action might include limiting the target sets or profiles – defined by appropriate decision makers during the targeting cycle – which a machine can prosecute without additional human inputs...”¹⁰
  - This statement asserts that the target profiles¹² a system uses need to be defined by human decision makers, and that these profiles can provide one basis for constraining system functioning.
  - At a process level, ensuring that target profiles are defined by the appropriate people would render inadmissible systems that set, develop or change their own target profiles during the course of an attack. Arguably such a policy would render inadmissible systems that ‘set their own goals’ in so far as such goals relate to targets for the application of force.
  - Suggesting that the definition of target profiles should be undertaken by people during the targeting cycle makes it clear that the targeting cycle as a whole cannot be undertaken by machines.
  - Placing specific ‘constraints or parameters’ on those profiles opens up the potential to determine that such profiles should not represent certain types of target or should not function in certain ways. This could include, for example, ensuring that target profiles do not represent people (in general).¹² The UK is not arguing for such a policy here – but this framing of the issue provides an entry point to such a debate.

- “[continuing from the UK text above] “...or limiting the range and task duration within which the system may be allowed to operate away from direct human involvement.”¹³
  - This clause points directly to further key parameters that are fundamental to ensuring meaningful human control – the area
of space and period of time within which a system functions independent of human engagement. Article 36 has highlighted the significance of these factors extensively. It would be preferable if the ‘or’ that introduces this clause were an ‘and’ to indicate that constraints or parameters on target profiles should be coupled with constraints or parameters on the area of space and time within which those target profiles are being used independent of human involvement. However, the UK statement is more illustrative than technical in its drafting and so this conjunction is not a major point at this stage.

In these first two elements we have the fundamental building blocks for a policy approach to human control: control over the key component of the ‘process’ a system will apply (the target profiles) and control over the ‘context’ within which that process will operate (the area of time and space).

x “Similarly, the number of engagements that could be carried out by a machine before further human input could also be limited to a single target or related target array.”

- The number of iterations of force that a system can undertake has a bearing on the level of predictability presented to a human commander. In so far as target profiles are only an approximation of an intended target, and the locations in time and space where force will actually occur are not specifically known (in line with the factors discussed above), each such iteration of force presents some unpredictability. Further iterations of force amplify the overall unpredictability of the attack being prosecuted.

x “These pre-set conditions would need to be regularly reviewed and updated in response to any changes in the operational context.”

- This point promotes a recognition that control is linked to understandings of specific operational contexts. Referring to the entry points for control noted previously – target profiles, range and duration, number of engagements – it notes that these factors would need to be calibrated to specific operational contexts.
- It also recognises that operational context can change over time, and therefore regular human engagement is needed to evaluate such contexts.
- It recognises that it is through the calibration of certain factors that sufficient control can be exerted, but that different contexts may require different levels of constraint.
- This in turn should draw us towards thinking about the principles that underpin the necessary level of constraint, rather than assuming that one metric (e.g. one fixed, maximum duration of system functioning) can be appropriate in all circumstances.

The UK government’s comments highlighted so far all accord directly with key elements that Article 36 considers central in relation to this issue. Continuing to emphasise these points has potential to greatly enhance the clarity of the current international debate.

‘Explainability’

This convergence of thinking continues in the paragraphs considered below, which discuss the transparency or ‘explainability’ of systems.

These issues are relevant to human control and also to ensuring that accountability and responsibility for a system’s use bear coherently on system users.

x “Where a human operator is required to supervise or interact directly with an autonomous system it must be designed in a way which enables them to understand the situation and the system status to an appropriate level. Again, what constitutes appropriate will depend on the operational context and environment.”

- This point usefully highlights that human understanding of system and context is fundamental to effective control. Human supervision or interaction with a system does not enable meaningful human control unless that interaction is founded on a meaningful understanding of what is actually happening.

x “Two important factors are the type of feedback available to the user before, during and after use, and the familiarity of the operators with the system – particularly regarding its capabilities and limitations.”

- Building on the point above, this line emphasises that information and user training/competence are important throughout the use of a system.

x “Specifically, a weapon or weapon system should not be able to have a lethal effect which cannot subsequently be explained by an appropriate human authority such as system operator or commander.”

- This point suggests a normative line which is of significance to the debate. It can be linked back to the UK’s point discussed earlier, regarding human definition and control of the target profiles a system uses (though this statement is more general than that).
- The term ‘lethal effect’ here should probably be read as relating to all weapon effects.
- This is an important statement as it highlights how explicable, as a technical characteristic of a system in a social context, is a component not only of control prior to outcomes, but also accountability after outcomes.

x “This feedback and familiarity are achieved through human-centred design practices and appropriate training, among other things.”

- Concluding the paragraph analysed above, this sentence usefully notes that explicable is a function both of system design and functioning (i.e. the technology itself), and the training of those that use the system (i.e. of the people).

Consideration of these factors is elaborated further in a subsequent paragraph – which constructively situates these considerations as “one practical metric for regulating developments” in this area.

x “This concept of explicable may point to one practical metric for regulating developments in military technology with lethal or disruptive capabilities. No matter what the methods taken to acquire, fix and engage a target, they must be understandable to the chain of command responsible for authorising use of the
system in question. This is unlikely to always mean complete technical transparency as in many situations this would neither be practical nor useful. Instead those with authority must be able to understand capabilities, limitations and the justification for any actions, in order to build an appropriate level of trust. If the method taken by a system to engage a target is not explainable and the outcome is outside of predictable bounds set by the operator or commander, it poses a number of legal, ethical and operational concerns."

- The suggestion that this orientation points to a possible normative line is important in itself.
- In part the paragraph serves to qualify the concept of ‘explicability’, in an operational context, to a practical level. This is wholly reasonable, as by analogy we do not expect an aircraft pilot to understand every line of computer code that works within an aircraft’s systems – rather, we expect them to have a functional understanding of the system as a whole.
- The reference to building an “appropiate level of trust” is awkward here – and slightly out of step with content around it. This has the feeling of being imported from other UK policy writings – in particular the MoD’s Joint Concept Note on Human-Machine Teaming. It is awkward here because whilst ‘trust’ might be important in terms of military operators’ relationships with equipment it is not the primary goal that is served by people being able to understand the capabilities and limitations of a system. The wider goals of control and accountability provide a more direct pathway to addressing the legal, ethical and operational concerns the UK goes on to note. This was prefigured in in other UK policy writings – and slightly out of step with content around it. This has the feeling of being imported from other UK policy writings – in particular the MoD’s Joint Concept Note on Human-Machine Teaming.

Putting comments on some of the detailed wording aside, this paragraph provides an important and constructive contribution to the debate. However, in line with some of the critical comments above, it is followed in the UK statement by a paragraph that is technically problematic. The underpinning meaning of this sentence seems positive, but there is a shift to ascribing legal decision-making capabilities to machines that goes against the UK’s own orientation to weapon systems as tools (as we have already emphasised and supported). The paragraph in question is broken down below:

x “In simple terms, if a weapon or weapons system applies a method or means of attack which is not understandable to the operator and the relevant chain of command, it cannot be compliant with the legal obligation to exercise precautions in attack, nor the requirement to consider second or third order effects of a strike which might legitimately be described as ‘forseeable’.”

- The underpinning meaning of this sentence seems positive. In so far as it is saying that if the way a weapon system functions is not understandable then it cannot be used in accordance with the law then Article 36 would agree with that orientation.
- However, by implying a weapons system might choose ‘a method or means’ of attack there is an implication that a machine will have broad scope to make determinations that are a human legal responsibility.
- Additionally, it would be better to conclude that it would not be possible ‘to use that system in compliance’ with the legal obligations, to emphasise that the obligation of compliance bears on the human users not on the system (as a mere machine).

- This is a highly problematic assertion in its implication that a machine might be ‘applying’ legal ‘concepts’ ‘such as necessity and proportionality’. Under the law, which is addressed to humans, it is humans that have the obligation to apply legal rules, which in turn draw, inter alia, upon concepts of ‘necessity’ and ‘proportionality’. The implications of the formulation adopted here are contrary to the UK’s own statements elsewhere on the proper orientation to the law.

- The final sentence of this paragraph is more constructively formulated – phrased in terms of a machine’s actions, rather than borrowing terms with a legal meaning. It would be preferable if there was some positive qualification – such as, ‘cannot be effectively anticipated or explained’ – in order to make it clear the problem here is not only systems that cannot be anticipated or explained at all.
- Of course, adding in such a positive qualification would challenge the confident but simplistic assertion that such a system cannot pass a legal weapons review. It would introduce the need to explain what sufficient predictability and explicability would look like, rather than suggesting that these characteristics present as a simple binary. This is one of the techniques the UK uses – to skip over areas of ambiguity or uncertainty - in order to maintain their political assertion that existing law is straightforward in its sufficiency.

x “It should be remembered that whilst unpredictability of weapon effect to the adversary is an advantage, unpredictability to the operator or commander in the delivery of effect – either in time, place or target – is highly undesirable.”

- This is another helpful statement from the UK. It situates the principle points of ‘unpredictability’ in relation to ‘time, place and target’ which re-emphasises the points noted earlier about these being the key building blocks for control.
- It also promotes recognition that control (or the removal of unpredictability) in relation to these factors is militarily desirable, which works against a social tendency for some to assume that military interests and moral/humanitarian interests are necessarily antagonistic to each other. This was prefigured in
one of their earlier sentences which noted ‘legal, ethical and operational concerns’.

Conclusions on the content

For the most part, across the passages highlighted above, the UK presents an orientation to the subject matter that has conceptual clarity and accords with the key points of Article 36’s thinking.

The UK’s position is characterised by,

- a strong emphasis on human control;
- an assertion that human control can be exerted through constraints on how systems function (through the target profiles that are used), and on duration and location of that functioning, and on the number of applications of force a system can undertake;
- an assertion that both system design and operator training need to contribute to systems being ‘explicable’ or understandable at a functional level;
- and these factors and considerations are framed within a recognition that different contexts may require (or enable) different constraints in order for the human control to be sufficient.

Although the use of legal language in certain places risks causing confusion (or reveals certain problems of approach, depending on interpretation) the overall framing of content here is very positive. Yet this constructive content is consistently packaged with politically targeted messaging asserting that existing law is straightforwardly sufficient and that states should resist calls to move towards the negotiation of any further regulation. That politically orientated messaging does much to obscure the content analysed here – making it difficult for other states (or other actors in the debate) to take a positive lead from the content being offered.

What to do about the unhelpful political messaging?

The main conclusion of this paper is that the UK needs to focus on communicating and promoting the key points of content highlighted above. Building a coalescence of states and other actors around these points represents the vital next step for international discussions on autonomy in weapons systems. The UK is therefore in a position to show genuine leadership – but to do so it needs to put to one side some of its other talking points.

The UK repeatedly asserts that existing law is wholly sufficient, that no new law is needed, that disagreement in the current debate also shows that no new law is possible, and if it were possible (i.e. through a format that did not have the participation of certain militarised states) then it would be unwelcome.

The problem for maintaining this line is that disagreement amongst different actors also means that there is no clarity of shared position regarding existing law in relation to autonomy in weapons systems. Thus, whilst the UK could argue that the duration and space over which a system can function must be set or defined by a human commander so that they can make judgements about that use in relation to the rules of law, other states may not agree. Similarly, the UK’s contention that a system must be practically understandable to an operator in order for that system to be acceptable under the law is not necessarily a position that all states share.

Furthermore, for some actors in the debate there are moral or ethical considerations, or considerations of what is socially prudent, that do not find expression in the current legal framework. For example, Article 36 argues that states should adopt, inter alia, a prohibition on targeting people through systems that use target profiles to apply force. Such a position is not argued on the basis that it flows automatically from existing IHL but from a number of other bases.27

The sufficiency of the existing legal framework, and its proper interpretation, in relation to the wider community of states, is clearly moot. For those, like Article 36, that see a need for lines of regulation that are not reflected in the current legal framework it is simple enough to say that the current framework is not enough. However, for states that argue that the law is clearly sufficient, the lack of agreement – the divergent views of the subject matter, the differing interpretations of existing law, and calls for new additional rules – all stand as evidence against their claim. Setting aside those issues which do not find expression in the law at present (including major issues such as the targeting of people), the best that could be contended is that existing law could possibly be sufficient if we all agreed on its meaning, and on the boundaries that it establishes, and implemented it fully.

Towards leadership?

Despite rejecting the idea of developing a legal response, the UK has, however, noted that:

- There may be merits in a ‘code of conduct’ which would ‘provide space to allow discussions to evolve towards an outcome,’ whilst ‘reducing the risks of unchecked and unregulated research and development’; and

- Further efforts should be dedicated to ‘operationalising’ the CCW’s Guiding Principles ‘in order to provide a LAWS-specific set of guidelines which could be overlaid on and integrated with existing regulatory structures’. 28

In 2020, the UK should focus on consolidating its approach to the subject matter in one place. If the CCW’s 2019 ‘Guiding Principles’ are to provide the basis for structuring further work within the CCW’s meetings then it is principle ‘c’ that should be given the greatest attention.29 This states that:

(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;

All of the positive content analysed above could be consolidated and presented for consideration as ways in which this principle could be ‘operationalised’. Bringing that content together under this principle would be a significant contribution to the multilateral debate - a contribution which would be a modest but important step towards leadership.

5
END NOTES


2 For background and formal documents see https://www.unog.ch/80256EDD006B8954/$file/20190318-5(d)_HMI_Statement.pdf


4 UK statement to the CCW GGE on LAWS on agenda item 5(d) ‘the human element...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(d)_HMI_Statement.pdf

5 UK statement to the CCW GGE on LAWS on agenda item 5(b) ‘characterisation of systems...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(b)_Characterisation_of_Systems.pdf


8 UK statement to the CCW GGE on LAWS on agenda item 5(b) ‘characterisation of systems...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(b)_Characterisation_of_Systems.pdf

9 UK statement to the CCW GGE on LAWS on agenda item 5(a) ‘challenges to international law...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(a)_Challenges_to_International_Law.pdf

10 UK statement to the CCW GGE on LAWS on agenda item 5(d) ‘consideration of the human element...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(d)_HMI_Statement.pdf

11 Article 36 considers ‘target profiles’ to be the sensor and system conditions under which force will be applied. As such they are a representation, or model, of an intended target constructed in the ‘language’ of a system’s sensors. For detail see http://www.article36.org/wp-content/uploads/2019/08/target-profiles.pdf

12 Article 36 has considered issues around the targeting of people on the basis of sensors in some detail here http://www.article36.org/wp-content/uploads/2019/11/targeting-people.pdf

13 UK statement to the CCW GGE on LAWS on agenda item 5(d) ‘consideration of the human element...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(d)_HMI_Statement.pdf


15 UK statement to the CCW GGE on LAWS on agenda item 5(d) ‘consideration of the human element...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(d)_HMI_Statement.pdf


17 UK statement to the CCW GGE on LAWS on agenda item 5(d) ‘consideration of the human element...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(d)_HMI_Statement.pdf

18 UK statement to the CCW GGE on LAWS on agenda item 5(d) ‘consideration of the human element...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(d)_HMI_Statement.pdf

19 UK statement to the CCW GGE on LAWS on agenda item 5(d) ‘consideration of the human element...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(d)_HMI_Statement.pdf

20 UK statement to the CCW GGE on LAWS on agenda item 5(d) ‘consideration of the human element...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(d)_HMI_Statement.pdf

21 UK statement to the CCW GGE on LAWS on agenda item 5(d) ‘consideration of the human element...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(d)_HMI_Statement.pdf

22 UK statement to the CCW GGE on LAWS on agenda item 5(c) ‘potential military applications...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(c)_MI_State ment.pdf


24 For example, at the CCE in 2018, the UK stated clearly that, "The UK believes existing International Humanitarian Law (IHL) is the applicable legal framework for the assessment and use of all weapons systems, including autonomous systems, in armed conflict. The fundamental principles of distinction, proportionality, necessity and humanity are at the core of compliance with IHL. It is the UK’s position view that those principles, and the requirement for evaluating the risks implicit in the use of lethal force, can only be assessed and applied by a human.” Online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(c)_MI_Statement.pdf

25 UK statement to the CCW GGE on LAWS on agenda item 5(c) ‘potential military applications...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(c)_MI_Statement.pdf

26 UK statement to the CCW GGE on LAWS on agenda item 5(c) ‘potential military applications...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(c)_MI_Statement.pdf


28 UK statement to the CCW GGE on LAWS on agenda item 5(e) ‘options for addressing...’, online at: https://www.unog.ch/80256EDD006B8954/$file/20190318-5(e)_Policy_State ment.pdf