

Article36

Shifting definitions - the UK and Autonomous Weapons Systems July 2018

In its recent Joint Concept Note, [Human Machine Teaming \(JCN 1/8\)](#), the UK Ministry of Defence (MoD) has signalled some welcome flexibility in its approach to the definition of autonomous weapons systems. The catalyst for this progress has likely been persistent pressure from civil society and members of parliament, most notably the House of Lords Select Committee on Artificial Intelligence (AI). Beyond the narrow perspective of the MoD, such a shift is vital if the UK is going to establish credibility as a leader towards 'ethical AI'.

The MoD has previously defined autonomous weapon systems (in part) as those “capable of understanding higher level intent and direction,”¹ therefore rendering them, in the foreseeable future, technologically infeasible. Choosing to set these key definitional criteria so far away from current technological developments and concerns resulted in the UK being unable to contribute substantively to the ongoing international debate regarding lethal autonomous weapons systems, such as those being undertaken within the framework of the UN Convention on Conventional Weapons (CCW) in Geneva. With the UK's definition being situated so far in the technological 'future' (to the point of perhaps being unrealisable, by the MoD's own admission), statements such as “we have no plans to develop or acquire such weapons” could appear progressive without actually applying any constraint on the UK's ability to develop weapons systems with greater and greater autonomy.

The UK's impractical definition has invited necessary criticism from Members of Parliament, in particular the House of Lords Select Committee on AI, as well as academic experts and civil society groups. In 2017, that House of Lords committee on AI reviewed the issue, and called for a reconceptualized definition of autonomous weapons that “should be realigned to be the same, or similar, as that used by the rest of the world.”² Sitting members of parliament have several times pushed the Minister of State for the MoD, Mark Lancaster, to amend the problematic definition - but such calls have been officially rejected.³

Article 36, alongside civil society partners, has called for clarification on the UK's concept of human control, as well as explanations of how systems, currently available or slated for future

¹ Ministry of Defence. (2018) Joint Concept Note 1/18: Human-Machine Teaming. 60.

² Parliament. House of Lords Select Committee on Artificial Intelligence (2018) AI in the UK: ready, willing, able? Report of Session (HL 2017-19 (100)). Paragraph 346.

³ HC Deb, 23 April 2018, cW. Available at:

<http://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2018-04-23/137095/>

development, will remain within those boundaries of “meaningful human control”.⁴ The MoD has, until recently, offered little of substance on such questions.

Human Machine Teaming JCN 1/8 contains a short annex (A) titled “Understanding assessments of autonomy”. This annex initially refers back to the MoD’s “endorsed definition” (as referred previously in this text), yet then notes that terms in the autonomous weapons debate are used in different ways by different communities of practice and goes on to situate the UK’s endorsed (previous) definition as the product of a particular analytical approach.⁵

It notes that in “(a) Tactical and technical contexts” even simple mechanical devices might fulfil the criteria of being autonomous. However, in “(b) Ethical or legal contexts” it suggests that autonomy is used to describe elements with “agency” ...“something far beyond the ability of simple mechanical devices”. It situates the UK’s endorsed definition in this context, reiterating that “for the technologically foreseeable future ... no machine possesses ethical or legal autonomy”, then stating again that the MoD does not plan to develop any “lethal autonomous weapons systems”.

Significantly, however, the annex then goes on to consider “(c) Relative autonomy descriptors”, where it provides a basis for stepping away from the straightjacket of the “endorsed definition”. Here, in contrast to contexts (a) and (b), it sees the term autonomy as “more useful” when considered “as a relative capability to accomplish a task”. Here, it will be important to balance the roles of “artificial intelligence operation” and “direct human control” in the context of possible risks and efficiencies.⁶

This shift is subtle. It is certainly not a rejection of the previous definition or its replacement with a new definition of what an “autonomous weapons systems” may constitute. Rather, it opens up a potentially productive space for considering how AI and direct human engagement might need to be balanced in relation to specific functions.

In the context of the current international policy and legal debate on this theme it is probably prudent for the UK not to rush to a revised definition of an autonomous weapon at this stage. In international discussions there have been divergent views on how an approach to policy or legal definitions might best be crafted. However, 2018 has seen an emerging consensus that the form of necessary human control must be the central focus for further discussion. Thus, rather than proposing an alternative definition for a ‘lethal autonomous weapons system’ it would be more productive for the UK to elaborate the nature of human control that needs to be retained in processes of identifying and choosing to apply force to targets in particular places and at particular points in time.

⁴ Moyes, Richard. “Key Elements of Meaningful Human Control.” Article 36, April 2016. <http://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf>.

⁵ Ministry of Defence. (2018) Joint Concept Note 1/18: Human-Machine Teaming. 57.

⁶ *ibid*

Through the Annex in *Human Machine Teaming* JCN 1/8, the UK appears to signal that there is space for more detailed dialogue and analysis regarding the ‘balancing’ of artificial intelligence functions and human functions within weapons systems - a signalling that is in line with their more recent statements that defining the necessary form of human control in the functioning of weapons systems should be central to the next stages of this debate.

Thus, whilst ministerial responses to questions on autonomous weapons systems have been excessively dismissive of the serious concerns being raised by parliamentarians, there are indications here that officials are recognising that the UK could play a more engaged role in the current international debate. The creation of some distance from its previous definition should be recognised by those parliamentarians that have been pressing this issue as a subtle but significant response to their efforts. It was also a necessary response. As the UK seeks to position itself as a leader towards the development of “ethical AI”, a failure to engage substantively with the extent that we can delegate life or death decisions to AI functions could still have significant implications for the UK’s reputation that extend beyond the interests of the MoD.

Recommendations

The UK needs urgently to build on the space that it has opened up for a more constructive policy making dialogue. UK officials should:

- **Support the adoption of a negotiating mandate in the UN CCW.** A significant and growing group of states are calling on the UN to move to negotiations - a shift that would allow for it to create a legal instrument containing regulations and prohibitions on weapons systems with aspects of autonomy in their critical functions. If the UK seeks to claim leadership in the context of “ethical AI” it needs to show actual leadership in pushing for negotiations.
- **Develop a cross-government dialogue on the human control necessary in relation to AI systems with implications for human life.** The UK cannot develop leadership on the ethics of AI without recognising that military issues have moral implications that extend beyond the military sphere. Consideration of military applications of AI must also be integrated into the mandates of bodies being established under the Department for Culture, Media and Sport (DCMS), and the Department for Business, Energy and Industrial Strategy (BEIS), in order to ensure logical and ethical coherence around the policy orientations to the role of AI in decisions fundamental to human welfare.
- **Urgently develop a position on the necessary elements of human control that must be present in military attacks.** The UK has expressed a recognition that further analysis on the necessary human control during attacks needs to be developed. The UK should convene military and non-military practitioners, as well as cross-government partners, to develop a coherent approach to this question in order to establish a more authoritative negotiating position in international legal and policy debates regarding lethal autonomous weapons systems.