

Completely outside human control?

Article36

WWW.ARTICLE36.ORG
INFO@ARTICLE36.ORG
@ARTICLE36

MARCH 2023
DISCUSSION PAPER

*Comments on the proposed prohibition in the Working Paper submitted by Finland, France, Germany, the Netherlands, Norway, Spain, and Sweden to the 2022 GGE on LAWS, July 2022.*¹

Background

This seven-state working paper (noted above) promotes a version of the ‘two-tier approach’ – a structure of approach that is broadly supported across discussions on autonomous weapons.

The working paper contains valuable content relating to the control of systems in practice – notably around the need for “sufficient understanding of the weapons systems’ way of operating” and the need for “spatial and temporal limits” in the deployment of a system.

However, a particular weakness of the paper is in its formulation of systems to be prohibited. The formulation adopted appears in a number of slightly different forms, as presented below:

(Pre-preamble)

In the framework of the GGE, States should commit to:

- 1) outlaw fully autonomous lethal weapons systems operating completely outside human control and a responsible chain of command...

(Preamble)

- 5) Recognise therefore that lethal autonomous weapons systems operating completely outside human control and a responsible chain of command are unlawful.

¹ See online at: https://documents.unoda.org/wp-content/uploads/2022/07/WP-LAWS_DE-ES-FI-FR-NL-NO-SE.pdf

(Operationalising)

- 1) Commit not to develop, produce, acquire, deploy or use fully autonomous lethal weapons systems operating completely outside human control and a responsible chain of command

[However – note also:]

Commit only to [develop, use etc] systems where:

- b) appropriate human control is retained during the whole life-cycle of the system... by ensuring that humans will be in a position to...
- c) human responsibility and accountability is preserved...To that end, the following measures and policies should be implemented...
 - operation of the system within a responsible chain of human command, including human responsibility for decisions to deploy and for the definition and validation of the rules of operation, use and engagement;

Analysis

For the purposes of this paper we will not dwell on the language of “fully autonomous lethal weapons systems.” The main problematic we want to analyse here relates to:

systems operating completely outside human control and a responsible chain of command

There are two components to this formulation:

“Systems operating completely outside human control”

The notion of a system being “completely” outside human control establishes a fantastical and unhelpful basis for boundary definition. A purpose of formulations such as ‘meaningful human control’ has been to build recognition that the nature of human control that is needed requires further specification if it is to be approached normatively. However, no such further specification is provided in relation to this formulation – it is presented in absolute terms.

Thus, the ‘complete’ absence of human control runs into counterarguments that:

- A system would be under ‘some’ degree of human control if a person is required simply to activate it (but nothing else) – i.e. the operator controls when the system starts.
- A system would be under ‘some’ degree of human control if it had a human-set limit to its operational duration – i.e. the operator controls how long the system functions for.
- A system would be under ‘some’ degree of human control if a person sets only the geographical boundaries within which it can operate – i.e. the operator controls only where the system will function.

Various other such examples could be generated.

All of these serve to illustrate that the development, production and use of a system that lacks any of these characteristics is completely implausible. The notion that a system could be 'used' whilst also being 'completely outside human control' is arguably impossible in itself.

This formulation cannot be sustained and no serious argument has been provided to justify it either in writing or in expert discussions. However, in this working paper the formulation is in cumulative conjunction with another requirement. To be unacceptable, not only must a system be completely outside human control, it must also be "completely outside...a responsible chain of command".

"Completely outside...a responsible chain of command"

A "responsible chain of command" can be understood in the context of the paper's later positive obligation requiring operation of the system "within a responsible chain of human command, including human responsibility for decisions to deploy and for the definition and validation of the rules of operation, use and engagement."

This formulation implies that the "responsible chain of command" is simply affirming that it is people ("human responsibility for") who will undertake the normal processes of developing rules of engagement and overseeing the deployment of systems etc. For a system to be "completely outside...a responsible chain of command" therefore implies that there is no such operational structure of authorised human planning and management within which a process of use is embedded.

Not only does the system have to be completely outside of human control, it also has to be *completely outside* of any structure of prior or ongoing human management and operational specification.

This proposal is so disconnected from reality that it is hard to capture it without resorting to parody. It is a policy to prohibit the fictional Terminator – operating both outside of human control and produced and deployed by the wider "Skynet" computer system. The logic of human commitments not "to develop, produce, acquire, deploy or use" such a system breaks down as the very processes of humanly 'deploying' or 'using' the system would *in themselves* mean that the system in question no longer falls within the prohibition as described.

If a system were deployed and used by a person, those acts would constitute some form of human control and/or chain of command. *Ipsa facto* the system would not be prohibited. The paper's position is that it is prohibited for humans to deploy or use systems that if deployed or used by a human are not prohibited.

This formulation has a history, but it would be helpful to the international conversation if certain states could accept that it does not have a future. Those states that continue to support and adopt this language (particularly those that have national positions committing them to work for international outcomes on this issue) should question the intellectual rigour and effort that they are bringing to the conversation.

Some light in the tunnel – “appropriate human control”

It is notable in the positive obligations of the same joint paper that the systems will actually only be used if “appropriate human control is retained during the whole life-cycle of the system.” The paper then goes on to describe a variety of factors that are rather jumbled between the notions of ‘control over use’ and the ‘chain of command’ but that provide some valuable practical content on how a useful understanding of “appropriate human control” might be understood.

We noted at the top of this paper that there are useful sections relating to the control of systems in practice – notably around the need for “sufficient understanding of the weapons systems’ way of operating” and the need for “spatial and temporal limits” in the deployment of a system. If the prohibition formulation was linked to this notion of “appropriate human control”, and tied to some of the key boundary structures that the working paper recognises in these areas, then this could form a serious basis for further work towards a legal instrument.