Verifying LAWS Regulation - Opportunities and Challenges

iPRAW Working Paper – August 2019

The International Panel on the Regulation of Autonomous Weapons (iPRAW) is an independent, interdisciplinary group of scientists working on the issue of lethal autonomous weapon systems (LAWS). It aims at supporting the current debate within the UN Convention on Certain Conventional Weapons (CCW) with scientifically grounded information and recommendations – looking at a potential regulation of LAWS from different angles.

This iPRAW publication is a brief overview of the matter of verification for a legally binding, potential regulation of LAWS, e.g. a new protocol to the CCW. It does not aim to present a verification regime but strives to inform policy makers about potential challenges and solutions regarding this issue.

The Basis for a Verification

Verification is a topic that has hardly been considered in the CCW debate on LAWS so far. Without consensus on a legally binding regulation or even the exact subject of a regulation, it is difficult to figure out the specific details of a verification regime. Nevertheless, we will look into **guiding elements** that could characterize such a verification mechanism. To that end, we will pick up the legally binding options discussed in iPRAW's fifth report[1], focusing on human control as the subject of regulation. iPRAW defines human control, in this context, as follows: situational understanding and options for intervention by design and in the use of the weapon system.

Verification as an Arms Control Tool

Verification in arms control is supposed to enhance compliance with a specific regulation, usually by detecting non-compliance.[2] Its objective is to **enhance confidence and trust** between the States Parties. In that regard, verification plays an important role in arms control and disarmament. It is usually a technical issue derived from legal requirements backed by the political will of the States Parties to a treaty. Verification of arms control regulations has to strike a balance between (military) secrecy on the one hand and transparency on the other hand.

Even transparency measures below a verification mechanism can have a stabilizing effect because states know what to expect from certain actions and what not, e.g. regarding the range or payload of certain systems. In humanitarian arms control treaties such as the CCW Protocols, hardly any verification instruments can be found. If at all, the Protocols call for transparency measures, like voluntary reports. With regard to weapons of mass destruction,

---

1 International Panel on the Regulation of Autonomous Weapons (December 2018), *Concluding Report*, <https://www.ipraw.org/wp-content/uploads/2018/12/2018-12-14_iPRAW_Concluding-Report.pdf>. Please note that the report does not recommend either a legally binding regulation or soft-law measures, but suggests the explicit implementation of the principle of human control over the use force.

2 The term verification (and validation) is also used in **computer sciences** to describe the testing of software. Those techniques might feed into the technical solution but are singularly insufficient. In that context those terms are defined as "**Validation:** The assurance that a product, service, or system meets the needs of the customer and other identified stakeholders. It often involves acceptance and suitability with external customers." "**Verification:** The evaluation of whether or not a product, service, or system complies with a regulation, requirement, specification, or imposed condition. It is often an internal process." (The PMBOK (Project Management Body of Knowledge) guide, a standard adopted by the Institute of Electrical and Electronics Engineers (IEEE).

the Biological Weapons Convention[3] entered into force in 1975 and developed a huge normative impact without any verification mechanism; albeit with trust building measures to enhance transparency (and substantial violations by States Parties in the 1990s).[4]

Verification measures can roughly be grouped in four categories along the subject of the regulation: existence/absence, quantity, quality, and (in some instances in parallel to those categories) change. The verification of **existence or absence** of weapons, emissions or substances can be implemented through on-site inspections of production facilities, radiation measurements, and environmental samples, for example. The **quantity** of weapons can be verified through counting said weapons in a certain area. To understand the **quality** or capabilities of certain weapons observers can use standardized tests to asses, for example, payload or range. **Change**, e.g. of production facilities or the location of troops, can be detected through satellite images or with the help of open source and social intelligence. Relevant **actors** to collect the information are intelligence services, non-governmental organizations and specific third-party organizations or treaty organizations like the Organisation for the Prohibition of Chemical Weapons (OPCW).

How to Verify Human Control?

To verify a regulation that calls for human control, one should focus on the link between human and machine. Technically, one would have to evaluate every single use of a LAWS, but one could focus on systems-of-systems like battlefield management systems, too. Verifying the situational understanding and options for intervention by design and in use would require that the operator/commander actually understood what is going on, considered to intervene, and made a deliberate decision about action or inaction. In the case of autonomous functions in weapon systems, verification measures could be applied during various stages of a weapon system's life cycle that would be closely linked to the implementation of human control by design and in use.

Specific challenges arise from software as an enabling technology of many modern weapon systems, e.g. fighter aircraft. It improves key figures like speed, altitude ceilings etc. and without software assistance, many capabilities would not be possible. Nevertheless, the software itself is usually not subject of verification – instead only platform-based characteristics like range, payload, or quantity are examined.

Design Phase and Procedures

In the design stage, the existence and sufficiency of communication links and interfaces would be a first step. The type and way data about the military operation is stored would be of relevance.

If 'control in use' cannot be verified sufficiently, 'control by design' needs to be further emphasized to the point where 'control in use' is an absolute necessary condition for the use of the system being designed. Taking this path towards regulation is risky, though, because design specification might be circumvented in operation and the design could be altered through **software updates** after inspection. Not only for regulation but also for verification, software updates can be problematic if they alter the capabilities of the machine with regard to the human role. It limits, for example, the effectiveness of on-site inspections. Of course, the design of the system must allow for human control which would require hardware for a

---

[3] Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction (short: Biological Weapons Convention (BWC) or Biological and Toxin Weapons Convention (BTWC)).

[4] See Laura H. Kahn (May 2011), *The Biological Weapons Convention: Proceeding without a verification protocol*, <https://thebulletin.org/2011/05/the-biological-weapons-convention-proceeding-without-a-verification-protocol/>.

communication link and a certain set-up of the interface, but those features would not tell much about the actual use in operation.

In the preparation of the operation, observers could assess the procedures in place to ensure human control in the specific operational context. Just as design specifications such procedures would be a necessary but not sufficient element of verification.

During Operation

During operation the relevant data on the human-machine interaction could be recorded and stored for immediate or ex post evaluation.

While 'traditional' arms control regulates military capabilities, **numbers** and/or material (e.g. fissile material) and humanitarian arms control usually focuses on the **effect** of a weapon, a regulation of autonomous weapons/human control would address the **process** of use (human-machine relation). This constitutes a **qualitative feature** that is hard to grasp. Furthermore, for the potential victim or an uninvolved observer the human role in the targeting process is **not visible from the outside**. An unmanned system could be remotely controlled, fully autonomous or everything in between. With hardly any indications perceptible during use, the human role would have to be evaluated in every use of unmanned systems by the States Parties. At least, all data about this would have to be stored for a certain amount of time and could become subject to random samples,[5] which will be discussed further below.

After Use

After use, the recorded data could be evaluated. This ex post approach to verification needs to be designed into the system. The implementation of this approach depends on the type of weapon systems and context of use. There are two options on that regard: (1) assess the data of all (known) applications of autonomous functions in target selection and engagement or (2) examine only cases of doubt.

Usually, verification measures cannot provide perfect security and proof of compliance, therefore compliance is only verified through falsification. On that regard, the first option would follow an unusual path because, theoretically, the technology behind autonomous functions might allow an actual proof since all relevant process are digitalized and therefore easy to surveil and analyze. If data about all relevant activities[6] was stored and examined one could actually see if States Parties complied with the treaty. A verification approach requiring to record certain data would call for encryption technology to ensure the authenticity of the data.[7] The fact that the analysis would require similar computational methods like those enabling autonomous functions in the weapon system adds another layer of complexity.

One could also examine a subset of all collected data by drawing random samples.

The second option, the assessment of suspicious cases, would require some indicators (not necessarily proof) for the lack of human control which could be obtained from testing, training, design, and/or operational speed.

---

5 See Mark Gubrud & Jürgen Altmann (2013), *Compliance Measures for an Autonomous Weapons Convention*, <https://www.icrac.net/wp-content/uploads/2018/04/Gubrud-Altmann_Compliance-Measures-AWC_ICRAC-WP2.pdf>.

6 It also needs to be verified that the provided set of data is complete.

7 The block chain technology would probably not be suitable to address this issue because it is based on decentralization and transparency – two characteristics not necessarily desired for military information.

Actors

As with other arms control treaties States Parties could create an organization to verify compliance. Such an institution would be essential for the verifiability of LAWS in its proper use and design. The organization would need substantial funding and well-renowned experts to become an effective, trusted third party. In preparation and support of such an institution, an interdisciplinary task force/technical committee could be helpful to assess the technical and legal possibilities.[8]

Existing Proposals

Mark Gubrud and Jürgen Altmann offer some first, highly valuable steps to a solution to the challenge of a verification for LAWS.[9] They suggest a number of technical measures that could indicate the human role in the use of force. Besides checking the hardware for enabling technologies such as communication links, they propose to install cameras to supervise the operator and – especially – look for software solutions. For example, the collected information could be stored in a way that cannot be tampered with and which could (only) be accessed by a neutral treaty implementing organization. In analogy to a hardware black box, Gubrud/Altmann call this a **glass box**. They reverse the usual logic of verification (detect non-compliance) by attempting to continuously track the human involvement in an attack, which actually calls for remotely piloted systems instead of autonomous targeting functions. This way, the verification measures would probably shape the regulation.

This concept could provide a useful basis for verification. It is confronted with a few challenges, though. For example, the fact that all uses of unmanned weapon systems would have to be tracked, only allows for relatively small samples or require an automated assessment that flags suspicious uses. Concepts like context dependent human control, distributed authority or grey areas between remotely piloted and fully autonomous functions along the targeting cycle might be difficult to grasp as well. This ex post verification is also focused on unmanned aerial vehicles, building on experiences with and procedures for remotely piloted drones. It might have to be adapted to suit weapon systems in different shapes and contexts.

Conclusion

If States Parties find a consensus on a legally binding regulation and if they want to verify compliance, this verification would be challenging but not impossible. Different challenges arise from the verification of autonomous functions: they are a qualitative feature, the human role in the target selection and engagement is not visible from the outside, and the software might be altered after inspection. Those challenges would call for a mix of instruments and could depend on the specific type of weapon system and the application of autonomous targeting functions. As with the requirements for human control, there is probably no one-size-fits-all solution to verification. The regulation could be useful even without hard verification measures but would benefit from enhanced transparency.

The International Panel on the Regulation of Autonomous Weapons (iPRAW) is coordinated by:
Stiftung Wissenschaft und Politik (SWP) – German Institute for International and Security Affairs
Ludwigkirchplatz 3-4, 10719 Berlin, Germany
This project is financially supported by the German Federal Foreign Office.
Find all reports and more information online at **www.ipraw.org**.

8 See Niklas Schörnig (March 2019), *Murmeltiertag in Genf: Probleme, Knackpunkte, mögliche Lösungen*, <https://blog.prif.org/2019/03/26/murmeltiertag-in-genf-probleme-knackpunkte-moegliche-loesungen/>.

9 See Mark Gubrud & Jürgen Altmann (May 2013), *Compliance Measures for an Autonomous Weapons Convention*, <https://www.icrac.net/wp-content/uploads/2018/04/Gubrud-Altmann_Compliance-Measures-AWC_ICRAC-WP2.pdf>.