Introduction

Paola Gaeta* and Marta Bo**

1. The Backstage of this Special Issue

This special issue of the *Journal of International Criminal Justice* is part of the research output of a 4-year project, funded by the Swiss National Science Foundation, entitled *Lethal Autonomous Weapon Systems and War Crimes: Who is to Bear Criminal Responsibility?*[1]

   The main aim of the research project was to examine the extent to which the use of so-called lethal autonomous weapon systems in armed conflict generates specific difficulties in the attribution of individual criminal responsibility for war crimes, where the use of such weapons appears to be in violation of rules of international humanitarian law. In light of the close relationship between war crimes and violations of these rules by belligerents and parties to armed conflict, it was only natural to extend the research to issues concerning responsibility under international law of the state resorting to autonomous weapon systems in warfare. In addition, the research examined both the regulatory challenges in the field of autonomous weapons, as well as interdisciplinary aspects more relevant to the issue at the core of the research project (such as 'technological' aspects of autonomous weapons; 'cognitive' aspects arising from human–machine interaction; and ethical aspects). This interdisciplinary openness has contributed significantly to enriching the field of investigation by members of the research team.[2]

* Professor of International Law, the Graduate Institute of International and Development Studies, Geneva (Switzerland). [paola.gaeta@graduateinstitute.ch]
** Senior Researcher at the Asser Institute, The Hague (The Netherlands) and Associate Senior Researcher at the Stockholm International Peace Research Institute (Sweden). [m.bo@asser.nl]
1 Other results of the research project have largely been published in articles and blogs by members of the research team, composed by Paola Gaeta (as research project leader); Marta Bo and Abhimanyu G. Jain (research associates for the entire duration of the project); Alessandra Spadaro (research associate for 6 months) and Victoria Priori (research assistant for 1 year).
2 One of the tools used to broaden the field of investigation to interdisciplinary approaches was the production of a 10-episode podcast entitled 'Lethal Autonomous Weapons: Ten Things We Want To Know'. In each episode of the podcast, members of the research team interviewed an expert on a specific issue related to the use of autonomous weapons after conducting research and preparatory studies. The podcast is available online at https://laws10.simplecast.com/ and has been released on all major podcast platforms.

The emergency caused by COVID 19 obliged the research team to cancel at the last moment an expert meeting, co-organized with the *Journal of International Criminal Justice*, scheduled to take place in Geneva in 2020. As a backup plan, we invited some of the experts to submit articles to this special issue of the *Journal* on an assigned topic related to the research project. We are very grateful to them for accepting this invitation.

This special issue of the *Journal* is the outcome of a cooperative process among all contributors. We held two online authors' meetings on the various themes at the core of each paper and commented on earlier drafts. We further discussed advanced drafts of the papers in a workshop held (this time in person) in Geneva, in November 2022, in which a selected number of invited experts participated.[3] Some of them also acted as discussants after providing anonymous peer review reports on the advanced drafts. We are very grateful to all of them for providing valuable inputs and feedback, and for making the workshop an enjoyable gathering at both the intellectual and personal levels. We are also very grateful to Gloria Gaggioli, Director of the Geneva Academy of International Humanitarian Law and Human Rights, for hosting the workshop at the Villa Moynier, in the *Salle Antonio Cassese*.

2. The Question of the Definition of Autonomous Weapon Systems and the Approach Taken in this Special Issue

The definition of autonomous weapon systems has sparked a vivid debate, and there is currently no accepted standard definition in an international legal instrument.[4] However, an assessment of the impact of autonomy *of* and *in* weapon systems on the ascription of legal responsibility to the individual or the state for, respectively, war crimes and violations of international humanitarian law does not require the adoption of a rigid definition. What matters instead is whether and to what extent the autonomy *of* and *in* weapon systems may create unique challenges from the perspective of ascription of responsibility. Consequently, the contributions included in this special issue are not organized around an unambiguous and uniform definition of autonomous weapon systems, but focus on certain issues arising from autonomy in the targeting process. However, it is clear that in all these contributions, the 'lion's share' of the discussions relate to the autonomy arising from two types of means and methods of conducting hostilities. Firstly, autonomous weapon

---

3   In addition to the members of the research team and the authors of the articles included in this special issue of the *Journal*, the participants in the workshop were Andrew Clapham, Urmila Dé, Jerome de Hemptinne, Giulia Fattori, Micaela Frulli, Julia Geneuss, Kevin Heller, Florian Jeßerger, Claus Kreß, Yvonne McDermott Rees, Robert Roth, Lena Trabucco, Harmen van der Wilt and Salvatore Zappalà.

4   A recent report identified at least 12 definitions. M. Taddeo and A. Blanchard, 'A Comparative Analysis of the Definitions of Autonomous Weapons Systems', 28 *Science and Engineering Ethics* (2022) 37, at 39.

systems enabled by artificial intelligence in the critical phases of the targeting cycle (i.e. detection, tracking, and engagement with the target). Some of these systems can operate completely or partially autonomously from the operator. In addition, thanks to self-learning artificial intelligence methods, they could potentially interact with the intended operational environment in a way that the developer or the user cannot foresee ex ante. The second is the use of artificial intelligence systems to support targeting decisions. In this case, the decision to engage with a target is made by the user of a weapons system on the basis of information processed by intelligent software and its recommendations. This situation equally raises concerns about responsibiliy, particularly due to so-called automation bias and over compliance.

3. The Structure of the Special Issue

All the contributions to this special issue of the *Journal* deal with issues of responsibility arising from the aforementioned progressive autonomy *of* and *in* weapon systems. It therefore seemed natural to the guest editors to open the special issue with two essays aimed at setting the scene. Dustin Lewis's contribution is thus valuable for navigating an intricate debate, including terminology, around the risk of the so-called responsibility gap arising from the use of autonomous weapon systems. Among other issues, he clarifies lucidly the meaning to be given to the different terms used when discussing this potential gap, distinguishing among the concepts of responsibility, liability, and accountability. On the subject of responsibility, he also makes clear that one needs to make a precise distinction between the regime of criminal responsibility for war crimes and questions of international responsibility of the state (or any other party to an armed conflict). This is an important distinction for the proper framing of possible gaps of responsibility resulting from the use of autonomous weapon systems. Following on from Dustin Lewis, Guido Acquaviva's contribution is also essential for a correct framing of the issue of potential responsibility gaps, specifically within the realm of international criminal law. This paper focuses on the principle of so-called 'meaningful human control' and its various articulations. Of particular relevance is the discussion on whether and to what extent this concept can be linked to the theory of control applied by the International Criminal Court as a key concept for identifying the direct perpetrator of a crime among the participants.

After these two more wide-ranging contributions, there follow five contributions with more pin-pointed content. Each of these five contributions provides reflections on the implications on the ascription of responsibility, to the individual or the state, resulting from military attacks enabled by artificial intelligence-driven systems.

Abhimanyu G. Jain's contribution presents us with a reflection on the impact of 'reasonable' and 'honest' mistakes in the targeting process, which result in inadvertent engagement with prohibited targets. This reflection, which is of a general nature, is specifically related to potential 'mistakes' resulting from the

use of autonomous weapon systems or artificial-intelligence-driven decision-making systems in the targeting process. His analysis departs from a conventional approach, which typically leans towards reconstructing the prohibition against attacking civilians (and other persons and objects that enjoy immunity from attack) as an obligation of result (albeit of a non-absolute nature, due to the operation of specific exceptions).[5] His thesis, informed by his reading of international practice, posits instead that there would be no violation of the prohibition in the case where engagement with prohibited targets stems from reasonable and honest mistakes in the targeting process.[6]

The contribution by Paola Gaeta, who writes this short introduction with Marta Bo as guest co-editors of the special issue, instead examines whether 'unintended engagements' with prohibited targets enabled by artificial intelligence in autonomous weapons constitute acts of the user for the purposes of attributing liability. The discussion is approached from the perspective of the conduct element for the commission of war crimes relating to unlawful targeting and the notion of an act of a state for the purposes of international state responsibility.

In the following contribution, Marta Bo reflects on the applicability of the mode of 'commission by omission' to cases where an unlawful attack with autonomous weapon systems is not interrupted. Drawing connections between this mode of responsibility and the concept of human control, the article underscores the significance of guaranteeing the presence of human control throughout the targeting process. Human control manifested in the ability to supervise, intervene and stop an AWS-driven attack, is needed to ensure the widest possible application of the doctrine of 'commission by omission' and thus contribute to the criminal prosecution of crimes of unlawful attacks committed by a failure to halt them.

---

5   An obligation of result requires a state to guarantee the achievement of the prescribed result, and not to do its best effort to attain it (which is an obligation of means). This distinction is well known in domestic legal systems of Romano-Germanic tradition, while it is largely unknown in the Anglo-American tradition. See amplius C.P. Economides, 'Content of the Obligation: Obligation of Means and Obligation of Result', in J. Crawford, A. Pellet, and S. Olleson (eds), *The Law of International Responsibility* (Oxford University Press, 2010) 371, at 375–381, including for a discussion of the proposal to include the distinction in the Articles on State Responsibility of the International Law Commission of the United Nations ('ILC Articles on State Responsibility').

6   This thesis does not fully convince the editors of this special issue, who instead lean towards the traditional reading. The wording of the ban on attacking civilians (or other persons and property protected by immunity from military attack) does not include the admissibility of 'mistakes'. There are express exceptions to the prohibition of attacking civilians, such as when they directly take part in hostilities, or when the attack is directed against a military target that may cause civilian casualties. Beyond these and other express exceptions, the prohibition against attacking civilians must be interpreted broadly. This is also in the light of theological interpretations of the rules on the conduct of hostilities, which aim to allow the greatest possible protection of the so-called 'victims of war'. It should also be noted that, according to the general regime of international responsibility codified in the ILC Articles on State Responsibility, 'error' does not figure among the circumstances precluding the wrongfulness of the act of a state.

Antonio Coco's contribution presents a well-articulated picture of the impact that the use of autonomous weapon systems can have on the cognitive sphere of the user and, consequently, on his criminal liability for war crimes in the case of unlawful military attacks. The analysis addresses both the subjective element required for these crimes and the potential relevance of the defence of mistake of fact. The article concludes with some interesting considerations on the measures that could be implemented to mitigate the impact on the cognitive sphere of AWS users.

Anna Greipl's article, on the other hand, presents an interesting new profile, as it ventures into the terrain of the intentional artificial intelligence-enabled commission of war crimes and other international crimes. Using Article 30 of the Statute of the International Criminal Court as a parameter, the author focuses on the impact that data-driven learning systems can have on the ascertainment of the subjective element.

The special issue concludes with two contributions that reflect, in broad terms, on two theses put forward in the legal discourse to address obstacles that the use of autonomous weapon systems may pose to the attribution of criminal responsibility for war crimes. Alessandra Spadaro analyses the applicability of the doctrine of superior responsibility to instances involving the use of autonomous weapon systems, potentially leading to so-called 'gaps' of criminal responsibility in the event of unlawful attacks. Thomas Weigend provides a poignant reflection on the possibility and desirability of holding autonomous weapon systems themselves criminally liable, as proposed by some scholars. Both articles remind us that the solution to the challenges posed by these new technologies should involve attempting to close potential or possible responsibility gaps through old or new theories of criminal responsibility. And that the focus should be on regulating the development and use of these technologies.

4. The Anthology

The contributions included in this special issue deal with questions of responsibility arising from the use of autonomous weapon systems in the conduct of hostilities. As is well known, however, international humanitarian law also prohibits the use of weapons contrary to certain principles dating back to the Hague Convention of 1907 and reaffirmed in the First Additional Protocol of 1977. Among these principles is the prohibition of the use of weapons that are indiscriminate by nature or that cause unnecessary suffering. It is important to note that the contributions included in this special issue do not delve into issues concerning the compliance of autonomous weapon systems with these prohibitions.

For the reasons we explain in more detail in our introductory note in the Anthology segment, we republish in this special issue (by kind permission of Editoriale Scientifica) an essay by Antonio Cassese. In this way, we also take the opportunity to pay tribute to Nino — to mark the 20th anniversary of the founding of this *Journal*.