Daniele Amoroso

Autonomous Weapons Systems and International Law

A Study on Human-Machine Interactions in Ethically and Legally Sensitive Domains

Edizioni Scientifiche Italiane

Nomos

Cultura giuridica e scambi internazionali
Legal Culture and International Flows

4

La Collana 'Cultura giuridica e scambi internazionali' nasce dall'iniziativa di un gruppo di internazionalisti, desiderosi non tanto e non solo di colmare una lacuna nel panorama editoriale italiano, ma piuttosto di dare vita ad uno strumento in grado di cogliere e valorizzare l'incidenza sempre maggiore del diritto degli scambi internazionali sulla struttura dell'ordinamento internazionale nel suo complesso così come degli ordinamenti interni, senza trascurarne l'interazione con il diritto dell'Unione europea. In questo senso, la Collana si compone di due anime. Per un verso, lo sguardo è rivolto al diritto del commercio internazionale inteso in senso ampio, e cioè comprensivo della sua duplice dimensione privatistica (ad es. contrattualistica internazionale) e pubblicistica (ad es. protezione degli investimenti e OMC). Per altro verso, il concetto di scambio, combinato indissolubilmente alla cultura giuridica, serve invece a porre la Collana in una prospettiva più generale. Una prospettiva nella quale lo scambio è riferito anche a principi e modelli normativi e si realizza nei rapporti verticali e orizzontali tra ordinamenti, finendo per influenzarne il funzionamento. La Collana è intesa ad accogliere monografie, manuali e opere collettanee.

La pubblicazione è stata sottoposta ad una procedura di valutazione ad opera di *blind referees.*

The Book Series 'Legal Culture and International Flows' mainly originates from the necessity to give account of the increasing weight that the law of international trade exercises on the overall structure of the international legal order, as well as on domestic legal systems and EU law. Due to its peculiar cross-cutting approach, the Series fills in a gap in the editorial landscape by focusing on two main areas of interest. On the one hand, it is aimed at analysing the concept of international trades, comprehensive of both its private (e.g. international contracts) and public (e.g. investment protection and WTO) dimensions. On the other hand, the Series delves into the concept of "legal flows" in its broadest meaning, i.e. as a reference to the circulation of principles and normative schemes between legal systems. From this angle, the concept is essential for giving account of the cross-fertilization taking place between the areas of interest of the Series. The Series is intended to include monographs, textbooks, and edited volumes.

This book has been subject to a blind review process.

Daniele Amoroso

Autonomous Weapons Systems and International Law

A Study on Human-Machine Interactions in Ethically and Legally Sensitive Domains

Edizioni Scientifiche Italiane

Nomos

Il volume è stato realizzato con il contributo del Dipartimento di Giurisprudenza dell'Università degli Studi di Cagliari.

This book was realized with the financial support of the Department of Law of the University of Cagliari.

Amoroso, Daniele
Autonomous Weapons Systems and International Law
A Study on Human-Machine Interactions in Ethically and Legally Sensitive Domains
Collana: Cultura giuridica e scambi internazionali, 4
Napoli: Edizioni Scientifiche Italiane, 2020
pp. XIV+290; 23,5 cm
ISBN 978-88-495-4267-7 ISBN Nomos 978-3-7489-0953-8

www.edizioniesi.it
info@edizioniesi.it

Ad Anna
e all'inizio della nostra vita «vera»

che ben fu il più crudele e il più di quanti
mai furo al mondo ingegni empi e maligni,
ch'imaginò sì abominosi ordigni

L. Ariosto, *Orlando furioso*, canto XI, ottava 27

Table of Contents

Acknowledgments XIII

Prologue 1

Chapter I

Introduction

1. A (Brief) History of the Debate on Autonomous Weapons Systems 5
   1.1. "Taking the Human Out of the Loop": The Issue of "Autonomy" in the Military Research and Policy Documents of the US Department of Defense 7
   1.2. Questioning the "Dehumanization" of Warfare: From the Campaign to Stop Killer Robots to the Group of Governmental Experts on Lethal Autonomous Weapons Systems 11
2. What Counts as "Autonomy" in Weapons Systems? The Case for a Functional Approach 15
   2.1. An Overview of Existing and Foreseeable Autonomous Weapons Systems 18
   2.2. Normative Implications of a Functional Approach to Autonomy in Weapons Systems 21
3. Mapping the Ethical and Legal Debate on Autonomous Weapons Systems 23
4. Research Goals and Structure of the Book 25
   4.1. Disclaimer: What this Book is not About 28

Chapter II

A Legality "Test" for Autonomous Weapons Systems

The (In)compatibility of Autonomous Targeting with International Humanitarian Law and International Human Rights Law

1. Introduction 31
2. Setting Up the "Parameters" of the Test 36
   2.1. The Law of Targeting 37
   2.2. The Right to Life under International Human Rights Law 40
3. The First Prong of the Test: Distinction 45
   3.1. The Principle of Distinction under International Humanitarian Law 46
      3.1.1. The Definition of Civilians and the Problem of Direct Participation in Hostilities 47

3.1.2. The Definition of Military Objectives and the Problem of Dual-Use Objects 52
3.1.3. Other Protected Persons and Objects under International Humanitarian Law 56
3.2. Issues of Distinction under International Human Rights Law 58
3.3. Distinction and Autonomous Weapons Systems: The Problem of Situation Awareness 60
3.3.1. Perception of the Elements in the Environment 62
3.3.2. Comprehension of the Current Situation and Projection of Future Status 67
3.3.3. Critique of the Arguments Denying Legal Relevance to AWS' (Lack of) Situation Awareness 70
3.4. Summary of the Discussion 75
4. The Second Prong of the Test: Proportionality 76
4.1. The Principle of Proportionality under International Humanitarian Law 77
4.1.1. Collateral damage 78
4.1.2. Military Advantage 81
4.1.3. Excessiveness 84
4.2. Issues of Proportionality under International Human Rights Law 87
4.3. Autonomous Weapons Systems, Proportionality, and Balancing: Lost in Translation? 88
4.4. Summary of the Discussion 96
5. The Third Prong of the Test: Precaution 96
5.1. The Principle of Precaution under International Humanitarian Law 97
5.2. The Requirement of Precaution and the Right to Life under International Human Rights Law 102
5.3. The Implications of the Obligation to Take Precautionary Measures for the Human-Weapon Relationship 106
5.4. Summary of the Discussion 113
6. Conclusions: Has Autonomy in Weapons Systems "Passed the Test"? 115

Chapter III

The "Accountability Gap" Problem

Who is to Blame for Autonomous Weapons Systems' Misdoings?
1. Introduction 121
2. Structural Problems with Responsibility Ascription for AWS' Misdoings 123
2.1. The Unpredictability of Autonomous Machines 123
2.2. The "Many Hands" Problem 127
3. Individual Criminal Responsibility 131
3.1. Direct Responsibility 131
3.1.1. The Participation to the Crime by Software Developers 133
3.1.2. The "Many Hands" Problem and Joint Criminal Enterprise 135
3.1.3. Limited Predictability of AWS and the Role of *Dolus Eeventualis* 138
3.2. Superior Responsibility 140
4. State Responsibility 146

5. Corporate Responsibility 150
5.1. Tort Liability for International Crimes 150
5.2. Product Liability 152
5.3. No-Fault Liability 154
6. Conclusions: The Inevitability of Accountability Gaps and Its Implications for the Legality of Autonomous Weapons Systems 156

Chapter IV

Challenging Autonomous Weapons Systems as a Malum in Se
AWS, Human Dignity and the Martens Clause between Ethics and Law
1. Introduction 161
2. Entering Ethical Concerns in the International Legal Discourse: A Précis on Human Dignity and the Martens Clause 164
2.1. Human Dignity 165
2.2. The Martens Clause 172
2.3. Wrap-up of the Discussion 178
3. The Ethical Arguments against Autonomy in Weapons Systems 179
3.1. Autonomy in Weapons Systems as a Breach of the Human Dignity (and Humanity) of Targeted People 181
3.2. Autonomy in Weapons Systems and the "Human Agency Removal" Problem 189
4. The Preservation of Human Agency in Targeting Decisions as a "Dictate of Public Conscience" 195
4.1. Investigating the "Dictates of Public Conscience" on the Issue of Human Agency and Lethal Decision-Making: a) States 196
4.2. b) International Organizations 201
4.3. c) Global Civil Society and Private Sector 204
4.4. d) Documents on AI not Directly Addressing the Issue of Autonomy in Weapons Systems 211
5. Conclusions 214

Chapter V

A Normative Model of Meaningful Human Control over Weapons Systems
1. Introduction 217
2. Human Control vs. Weapon Autonomy: Striking a Balance between Humanitarian and Military Considerations 220
3. The Debate on Meaningful Human Control (MHC) over Weapons Systems: A Sketch 224
3.1. The Quality of Human Involvement 228
3.2. Shared Control Policies 232
4. Shaping the Content of a Normative Model of Meaningful Human Control 240
4.1. Primary Obligations: Control Privileges 241
4.2. Ancillary Obligations: Training and Design 246

5. The Way Forward: What Legal Regime for Meaningful Human Control? 250
5.1. An Overview of the Regulative Options in the AWS Debate 250
5.2. Possible Elements of a Future Protocol/Treaty on Meaningful Human Control 256

Epilogue 261

Bibliography 267

Acknowledgements

Writing a book generally creates a number of inextinguishable debts of gratitude for the author, and this volume makes no exception.

I would firstly like to thank Guglielmo Tamburrini, who introduced me to the issue of Autonomous Weapons Systems in 2016 and, since then, has been an invaluable travelling companion throughout the intellectual journey that eventually brought about the present monograph; and my mentors, Massimo Iovane and Fulvio Maria Palombino, whose guidance and support proved crucial at various stages of this academic enterprise and without whom this book would hardly have seen the light of day. I also wish to express my deepest gratitude to Giovanni Zarra, for his priceless role in the publication process, to Loris Marotti and Pierfrancesco Rossi, for their analytical (and, at times, *usefully* cruel) reading of the manuscript, to Giovanni (Joe) Dini for his competent and professional language revision, as well as to Giulia Ciliberto and Benedetta Giordano for their precious comments and research assistance. Finally, I feel indebted towards all the members of the International Committee for Robot Arms Control (ICRAC) for letting me be part of their community. They have been a continuous source of precious insights and up-to-date information. I sincerely hope that this volume may offer some contribution to their (our) cause.

I wish to acknowledge that I benefited from the financial support of the Sardinia Region (*Programma "Mobilità Giovani Ricercatori" 2018/2019)* and the *Deutscher Akademischer Austauschdienst* (DAAD), which allowed me to carry out independent research at the Brunswick European Law School in Wolfenbüttel (Summer 2018) and the *Universität der Bundeswehr* in Munich (Summer 2019). In this regard, a special thank goes to Winfried Huck (and his staff) and Frank Sauer for generously hosting me at their institutions.

The merits of this book (if any) have to be shared with all the above. All mistakes and omissions are, of course, solely my own.

Cagliari, 16 April 2020 Daniele Amoroso

Prologue

Recent advances in robotics and artificial intelligence (AI)[1] have paved the way to robots autonomously performing a wide variety of tasks[2] that may significantly affect individual and collective interests, which are worthy of protection from both ethical and legal perspectives. An exemplary case is the circulation of autonomous vehicles on public roads,[3] but one may also think of increasingly autonomous surgical and care robots.[4] To this list of robotic systems one may finally add the judicial use of AI software systems,[5] notwithstanding the lack of their (direct) kinetic interaction with the physical world. Indeed, the use of AI in the Court is aimed at replacing or supporting the human judge in decision-making processes and tasks that, *by their very definition*, are supposed to have an impact on legal rights and duties.

These technological developments revamped longstanding discussions on Ethical, Legal and Socio-Economic (ELSE) implications of

[1] The term "Artificial Intelligence", coined in 1956 by John McCarthy (S.L. Andersen, *John McCarthy: Father of AI*, IEEE Intelligent Systems, 2002, p. 84 f.), generally describes the capability of a computing machine to carry out tasks that are assumed to require human intelligence. This is done, in particular, by framing these tasks into mathematical models that machines can cope with.

[2] A robotic system may be counted as "autonomous" at given tasks if, once activated, it is able to carry out those tasks without further human intervention.

[3] See e.g. N.E. Vellinga, *Automated Driving and Its Challenges to International Traffic Law: Which Way to Go?*, in *Law, Innovation and Technology*, 2019, p. 257 f.; P. Lin, *Why Ethics Matters for Autonomous Cars* in M. Maurer *et al.* (eds), *Autonomes Fahren*, Berlin, 2015, p. 69 f. See also German Federal Ministry of Transport and Digital Infrastructure, *Report by the Ethics Commission on Automated and Connected Driving*, June 2017.

[4] S. O'Sullivan *et al.*, *Legal, Regulatory, and Ethical Frameworks for Development of Standards in Artificial Intelligence (AI) and Autonomous Robotic Surgery*, in *Int. J. of Medical Robotics and Computer Assisted Surgery*, 2019; M. Decker, *Caregiving Robots and Ethical Reflection: The Perspective of Interdisciplinary Technology Assessment*, 2008, in *AI & Society*, p. 315 f.

[5] J. Nieva Fenoll, *Inteligencia Artificial y Proceso Judicial*, Madrid, 2018. This issue, however, had already been explored in the pioneering volume by G. Sartor and L. Branting (eds), *Judicial Applications of Artificial Intelligence*, Dordrecht/Boston/London, 1998.

robotics and AI, whose origins can be traced back at least to Norbert Wiener's seminal reflections on the ethics of information technologies and robotics.[6] Present debates about machine autonomy in ethically and legally sensitive domains have now gone well beyond academic and specialist circles, entering the political debate and receiving considerable media coverage.[7]

Over and above the specificities of each technological application domain, there are a few overarching issues arising in connection with most artificial systems endowed with autonomy in the execution of tasks that are ethically and legally sensitive. First, there is the *technical question* concerning whether artificial agents are inherently unable to properly carry out certain functions governed by law (e.g. replace human judges in performing tasks involving discretionary reasoning and/or equitable evaluations), insofar as they would (allegedly) require uniquely human capabilities. Second, there is the (most strictly) *legal problem* of determining how to allocate responsibility if a machine happens to cause harm (think, for instance, of damages arising from surgical robots' mishaps). Third, we have the *philosophical debate* on the moral desirability of machine autonomy between those that argue from the perspective of *deontological ethics*[8] on one side and those that do so from a *consequentialist*[9] perspective on the other. The former maintain that it would be morally unacceptable to remove human agency from decision-making processes that are likely to impinge on individual rights and duties, as well as on relationships that are ethical in character (such as that of nursing care). The latter affirm the moral and legal duty to replace human operators with autonomous machines, whenever machine performance ensure better protection of the interests at stake (e.g. by reducing the number of road accidents and fatalities).

[6] N. Wiener, *The Human Use of Human Beings*, Boston, 1950; N. Wiener, *God & Golem, Inc. - A Comment on Certain Points Where Cybernetics Impinges on Religion*, Cambridge (Massachusetts), 1964.

[7] See, among many others, Editorial, *The Guardian View on the Future of AI: Great Power, Great Irresponsibility*, in *The Guardian*, 1 January 2019, available at: *www.theguardian.com/commentisfree/2019/jan/01/the-guardian-view-on-the-future-of-ai-great-power-great-irresponsibility*; Stephen A. Schwarzman, *Can We Make Artificial Intelligence Ethical?*, in *The Washington Post*, 23 January 2019, available at: *www.washingtonpost.com/opinions/2019/01/23/can-we-make-artificial-intelligence-ethical/*; Cade Metz, *Is Ethical A.I. Even Possible?*, in *The New York Times*, 1 March 2019, available at: *www.nytimes.com/2019/03/01/business/ethics-artificial-intelligence.html*.

[8] Broadly speaking, deontological ethics identifies moral duties as guides for acting and judging the moral worth of choices.

[9] Unlike deontological ethics, consequentialism focuses on criteria to distinguish between morally good and bad consequences of choices, and prescribes to judge the moral worth of choices in the light of consequences only.

All of these issues are emblematically encompassed in discussions on the legality, *under international law*, of autonomous robots endowed with the ability to deliver destructive force without human intervention, which are generally known as Autonomous Weapons Systems (or AWS). To begin with, it is indeed questioned whether it will ever be technically possible to program autonomous robots to faithfully comply with the norms governing the use of armed force in international law. Also, the problem has arisen as to how to ascribe responsibility whenever an autonomous weapons system takes a harmful course of action in breach of international legal prescriptions. Finally, one may detect a normative tension between consequentialist reasons favoring the use of these applications (e.g. the need to avoid targeting decisions tampered by typically-human biases) and the view, ultimately based on deontological ethics, whereby decisions encroaching upon fundamental human rights (first and foremost, the right to life) should never be entrusted to artificial agents.

Discussions on AWS have been contributed to by a multitude of academics, belonging to the most disparate of disciplinary fields (which include – in addition to international law – robotics, computer science, normative ethics, and military studies), by also involving governments, military bureaucracies, think-tanks, international organizations, and NGOs. This makes the AWS debate a privileged vantage point to carry out a study on human-machine interactions in domains, as the ones set out above, that are sensitive from both an ethical and legal perspective.[10]

At the same time, the legal problems raised by autonomy in weapons systems provide a uniquely representative sample, from both a theoretical and practical perspective, of the (potentially) disruptive impact of new technologies on norms and principles of international law.[11] Critics have indeed characterized AWS as "shaking the foun-

[10] For similar considerations, see C. Heyns, *Autonomous Weapons in Armed Conflict and the Right to a Dignified Life: an African Perspective*, in *South African J. on Human Rights*, 2017, p. 46 f., p. 48 ("The way we respond to autonomous weapons is a pivotal test case for the role of science in the future. The stakes cannot be higher – they are literally life and death – and how we deal with autonomous weapons will be the tone for how we deal with computers in general"); and R. Geiss and H. Lahmann, *Autonomous Weapons Systems: A Paradigm Shift for the Law of Armed Conflict?*, in J.D. Ohlin (ed.), *Research Handbook on Remote Warfare*, Cheltenham, 2017, p. 371 f., p. 374 ("The military aspect of this debate is only the tip of the iceberg. On a fundamental level, it needs to be asked how much 'de-humanizing' of societal mechanisms humankind can, or is willing to, afford to tolerate, before the social costs outweigh the benefits").

[11] The literature on this topic is vast and relentlessly growing. Among the most valuable and recent contributions, see E. Carpanelli and N. Lazzerini (eds), *Use*

dations" upon which international human rights law and international humanitarian law rest;[12] but even those who are in favor of them are ready to admit that this technology prompts a serious reconsideration of well-established international legal regimes, including those governing State responsibility and individual criminal responsibility.[13]

Against this backdrop, this volume's primary aim is to explore the international legal implications of autonomy in weapons systems, by inquiring what existing international law has to say in this respect, to what extent the persisting validity of its principles and categories is challenged, and what could be a way forward for future international regulation on the matter. From a broader perspective, as will be seen in the *Epilogue*, the research carried out here on the issue of the legality of AWS under international law aspires to offer some more general insights on the normative aspects of the shared control relationship between human decision-makers and artificial agents.

and Misuse of New Technologies. Contemporary Challenges in International and European Law, Cham, 2019; T. Burri, *International Law and Artificial Intelligence*, in *German Yearbook of In. L. 2017*, 2019, p. 91 f.; M.K. Land and J.D. Aronson (eds), *New Technologies for Human Rights Law and Practice*, Cambridge/New York/Port Melbourne/Singapore, 2018; S.-Y. Peng, H.-W. Liu, and C.-F. Lin (eds), *Governing Science and Technology under the International Economic Order*, Cheltenham/Northampton, 2018; R. Brownsword, E. Scotford and K. Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology*, New York, 2017; P. Achilleas and W. Mikalef (eds), *TIC, innovation et droit international: Technologie de l'information et de la communication*, Paris, 2017; H.N. Scheiber, J. Kraska, and M.-S. Kwon (eds), *Science, Technology, and New Challenges to Ocean Law*, Leiden/Boston, 2015; B. Mercurio and K.-J. Ni (eds), *Science and Technology in International Economic Law: Balancing Competing Interests*, Oxon/New York, 2014; M. Viola de Azevedo Cunha *et al.* (eds), *New Technologies and Human Rights: Challenges to Regulation*, London/New York, 2013; R. Ben Achour and S. Laghmani (eds), *Le droit international face aux nouvelles technologies*, Paris, 2002; C.B. Picker, *A View from 40,000 Feet: International Law and the Invisible Hand of Technology*, in *Cardozo L. Rev.*, 2001, p. 149 f. See also, with specific regard to technological developments in the military field, W.H. Boothby (ed.), *New Technologies and the Law in War and Peace*, Cambridge, 2019; R.T.P. Alcala and E. Talbot Jensen (eds), *The Impact of Emerging Technologies on the Law of Armed Conflict*, New York, 2019; W. Heintschel von Heinegg, R. Frau, and T. Singer (eds), *Dehumanization of Warfare. Legal Implications of New Weapon Technologies*, Cham, 2018; H. Nasu and R. McLaughlin (eds), *New Technologies and the Law of Armed Conflict*, The Hague, 2014.

[12] Human Rights Watch (HRW) and the International Human Rights Clinic at Harvard Law School (IHRC), *Shaking the Foundations. The Human Rights Implications of Killer Robots*, 12 May 2014.

[13] See, for instance, R. Crootof, *War Torts: Accountability for Autonomous Weapons*, in *University of Pennsylvania L. Rev.*, 2016, p. 1347 f.