I.1

"Algorithmic Target Construction" and the Challenges by International Human Rights Law

SUMMARY**:** 1. Introduction. – 2. Algorithmic Target Construction (ATC) – 3. ATC under IHRL. – 3.1. Data Collection. – 3.2. "Categorical" Decisions. – 3.3. "Legibility". – 3.4. Subjection to Solely-Automated Decisions. – 4. Conclusion: Is A Different Approach Possible?

1. The first use of an armed drone allegedly occurred during the US hunting of Osama bin Laden, and more precisely on February 4, 2002, when the CIA spotted a «tall man», much resembling bin Laden, around whom several people were «acting with reverence», and fired an *Hellfire* missile against him.[1] The CIA operators' assumption turned however to be untrue, as the target was later discovered to be a local unfortunately – for him – the same height as bin Laden. Interestingly enough, US authorities insisted the target was «legitimate»; the Pentagon then-spokeswoman so declared: «We're convinced that it was an appropriate target … [but] we *do not know yet exactly who it was*».[2]

Here the essence of a "signature strike" is captured perfectly. The term – now one of the art – refers to a methodology for selecting actual targets for drone strikes basing solely on their observed pattern of behavior (i.e. their "signature").[3] The target's personal identity remains unknown *before* the strike and may remain so also *after* it. It differs from a "personality strike" in that in the latter the target's personal identity is known to the authorities before the strike, which as a matter of fact takes place by virtue of the target's personal identification. Signature strikes have gained momentum under the Obama Administration, albeit some initial reservations.[4] They are believed to avoid a «huge number of civilian casualties»,[5] and to work as an appropriate tool to address the challenges of the well-known Global War on Terror.[6]

1 The episode is told by R. SIFTON, *A Brief History of Drones*, in *The Nation*, 27 February 2012, available at: http://www.thenation.com/article/166124/brief-history-drones#.

2 *Ibidem*, italics mine.

3 For a definition of signature strikes, reference can be made to K. BENSON, *"Kill 'em and Sort it Out Later:" Signature Drone Strikes and International Humanitarian Law*, in *Global Business & Development Law Journal*, 2014, p. 18. See also S. HOLEWINSKI, *Just Trust Us*, in P. BERGEN, D. ROTHENBERG, *Drone Wars. Transforming Conflict, Law, and Policy*, Cambridge, 2014, p. 45-46; K. KINDERVATER, *The Emergence of Lethal Surveillance: Watching and Killing in the History of Drone Technology*, in *Security Dialogue*, 2016, p. 224 ff.; T. WALL, T. MONAHAN, *Surveillance and Violence from Afar: The Politics of Drones and Liminal Security-Scapes*, in *Theoretical Criminology*, 2011, p. 239–245.

4 See the anecdote narrated by M. ZENKO, *Targeted Killing and Signature Strikes*, in *Council On Foreign Relations*, 16 July 2012, available at: http://blogs.cfr.org/zenko/2012/07/16/targeted-killings-and-signature-strikes/ (citing one legal advisor that so described the President's unease at striking at military-age males associated with terrorist activities but whose personal identity remained unknown: «[H]e didn't like the idea of kill 'em and sort it out later»).

5 See the Report *Emerging from the Shadows: US Covert Drone Strikes in 2012*, in *Bureau Of Investigative Journalism*, 3 January 2013, available at: http://www.thebureauinvestigates.com/2013/01/03/emerging-from- the-shadows-us-covert-drone-strikes-in-2012-2/.

6 Numbers and figures regarding the first year of the Trump Administration confirm that drone strikes are the first choice in counterterrorism abroad. See https://www.thebureauinvestigates.com/stories/2018-01-19/strikes-in-somalia-and-yemen-triple-in-trumps-first-year-in-office (showing that the number of strikes conducted in Yemen and Somalia in 2017 is nearly more than triple the number relating to the precedent year).

The gist of signature strikes is that they rely essentially on a process that can be named "Algorithmic Target Construction" (ATC). Current drone technology requires a human decision-maker to be present at the act of engaging that target; the ultimate decision about the life or death of the individual is thus entrusted to a human agent. The development and deployment of Lethal Autonomous Weapons Systems (henceforth: LAWS) puts such model in discussion as not only target selection, but also target engagement will be entrusted to a non-human decision-maker.[7]

The advent of LAWS is likely to act as a veritable turning point in our understanding of using force against humans; several features of this technology, and the impact thereof on international law, have already been tackled.[8] The present contribution aims at assessing the impact ATC may have on human rights. In Section 2 a brief description of how ATC functions will be provided; most suggestions will be drawn from current uses of technology in performing signature strikes via armed drones. In Section 3 I will then turn to the framework of international human rights law (IHRL) with a view to showing that ATC is suitable to affect basic human rights. I will then draw conclusions leaving some radical questions open (Section 4).

2. ATC can be defined as a process allowing for target selection and engagement through algorithms.[9] This concept has been recently adopted in a Briefing of the Geneva Academy of International Humanitarian Law and Human Rights.[10] Proceeding in reverse order, "Construction" refers to a methodology of gathering and then re-elaborating **data** which constitute the very input of the process. The outcome is the identification of a "Target", namely an individual or group of individuals that will be made the object of force delivery, while the process through which data are re-elaborated is "Algorithmic". The notion of ATC can be "unpacked" so as to distinguish between two particularly salient temporal stages at least, namely data collection and later decision-making.

As far as data collection is concerned, it is generally considered as the first component of "processing". To be employed in an operational scenario (such as battlefield or a law-enforcement operation), LAWS will need to gather as much data as possible from the field, and therefore they will presumably be endowed with software for carrying out a preliminary screening of individuals and places.[11] Collected data will then go through a process of organization and structuring; in particular, applying the same methodology currently in use for signature strikes, LAWS will likely

7 See among others: *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns*, UN General Assembly, A/HRC/23/47 (9 April 2013); *Losing Humanity: The Case Against Killer Robots*, Human Rights Watch (November 2012). The forum that is currently hosting discussion on «emerging technologies in the area of [LAWS]» is the Assembly of the States Party to the 1980 Convention on Certain Conventional Weapons (hereinafter: CCW), which decided to convene three Meetings of Experts between 2014 and 2016; the 2016 Fifth Review Conference of the CCW then established a Group of Governmental Experts (GGE) that held its meeting on November 2017, April 2018 and August 2018. A new round of meetings will probably take place in 2019. See *amplius* https://www.unog.ch/80256EE600585943/(httpPages)/8FA3C2562A60FF81C1257CE600393DF6?OpenDocument.

8 For a recent and thorough analysis, see D. AMOROSO, *Jus in bello and jus ad bellum arguments against autonomy in weapons systems: A re-appraisal*, in *QIL Zoom-In*, 2017, p. 5-31.

9 For a better understanding of how algorithms work and how influential their use can get, see P. ZELLINI, *La dittatura del calcolo*, Milan, 2018.

10 See M. BREHM, *Defending The Boundary. Constraints And Requirements On The Use Of Autonomous Weapon Systems Under International Humanitarian And Human Rights Law*, in *Academy Briefing No. 9*, May 2017, available at: https://www.geneva-academy.ch/joomlatools-files/docman-files/Briefing9_interactif.pdf.

employ software allowing for detecting individuals or group of individuals possessing certain personal attributes that are considered as statistically correlated to certain conducts – a process commonly referred to as "profiling".[12] The purpose of such process is to *predict* an individual's action on the basis of a so-called "pattern of life analysis"; the individuals whose data are gathered and elaborated by the machine are therefore "reduced" to a profile and put into "categories". Such type of analysis and technique of re-elaboration of personal data has been used for signature strikes since the beginning,[13] so that LAWS will be programmed so is more than a mere prevision.[14] In short, «an individual's pattern of behavior – or "signature" – serves as a proxy for determining if that individual» may be a target for the use of force.[15] In terms of technological feasibility, suffice it to recall that recently IBM has declared that it was about to offer an algorithm capable of distinguishing terrorists out of refugees.[16] As has been conveniently pointed out, «one of the most acute dangers of profiling is the fact that it tends to reduce the person to the profile generated by automated processes which are *liable to be used as a basis for decision-making*».[17]

Decision-making is another crucial step of ATC. While the current practice of signature strikes leaves the final decision (i.e. whether to engage or not the selected target) to a human operator, automated decision-making (understood as a process where taking the final decision is entrusted to a software, i.e. to a non-human operator) is already in place in different contexts, such

11 See A. SPAGNOLO, *Human rights implications of autonomous weapon systems in domestic law enforcement: sci-fi reflections on a lo-fi reality,* in *QIL Zoom-in*, 2017, p. 43.

12 See *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC* (General Data Protection Regulation; hereinafter: GDPR), art. 4(4): «"profiling" means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements».

13 N. ABÈ, *Dreams in Infrared: The Woes of an American Drone Operator*, in *Spiegel Online*, 14 December 2012, available at: http://www.spiegel.de/international/world/pain-continues-after-war-for-american-drone-pilot-a-872726.html («[w]e watch people for months. We see them playing with their dogs or doing their laundry. We know their patterns like we know our neighbors' patterns. We even go to their funerals»); G. MILLER, *At CIA, a Convert to Islam Leads the Terrorism Hunt*, in *Washington Post*, 24 March 2012, available at: http://articles.washingtonpost.com/2012-03-24/world/35447818_1_cia-officials-robert-grenier-ctc.

14 See M. SCHMITT, J. S. THURNER, *"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict*, in *Harvard National Security Journal*, 2013, p. 268.

15 Paraphrasing BENSON, *Kill 'em and Sort It Out Later*, cit., p. 29.

16 P. TUCKER, *Refugee or Terrorist? IBM Thinks Its Software Has the Answer*, in *Defense One*, 27 January 2016, available at: http://www.defenseone.com/technology/2016/01/refugee-or-terrorist-ibm-thinks-its-software-has-answer/125484/.

17 See the Report of the Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, *Application of Convention 108 to the Profiling Mechanism: Some Ideas for the Future Work of the Consultative Committee*, 11 January 2008, p. 5-6, available at https://rm.coe.int/16806840b9, italics mine. The matter is of particular concern when it comes to policing algorithms that may subject minorities to greater surveillance and therefore police force; see K. K. KOSS, *Leveraging Predictive Policing Algorithms to Restore Fourth Amendment Protections in High-Crime Areas in a Post-*Wardlow *World*, in *Chicago-Kent Law Review*, 2015, p. 301-334.

as recruitment, behavioral advertisement, access to credit.[18] It has been argued that entrusting LAWS to take engagement decisions will be more efficient than having a human operator do so.[19]

As will be illustrated in the following, it is by reason of dangers associated with the algorithmic processing of such data (to name a few, poor final decisions, misjudgments, or an alarming inclination for discriminatory outcomes) that a certain degree of human control over automated decision-making has been retained of paramount importance. It is however questionable the extent to which human will be able to exert control on such processes, especially when machine-learning or self-learning algorithms are employed.[20] Brief, LAWS employing machine-learning algorithms in ATC will be able to generate their own rules and conduct basing on their initial databank and gained experience "in the field",[21] which means that their ability to select and engage targets properly will depend on previous experience in the relevant operational field (battlefield; law-enforcement area of operations; etc.).[22] This has an impact on what will be later defined as "legibility" of the system: human operators will hardly be in the condition to understand how and why a LAWS operating through self-learning algorithms acted in a certain way.

In sum, ATC will be a key feature in the development of next-generation LAWS. As a complex process, it will involve data gathering, data elaboration and more importantly automated decision-making allegedly independent of human intervention. For all these steps, human rights of the potential targets are put at stake to some extent, which implies that in building algorithms developers must take those rights in due consideration – «seriously», in the famous words of Ronald Dworkin.

3. In order to assess the impact the development of ATC can have on IHRL, it seems appropriate to enucleate at least four critical features of that process: (1) mass surveillance techniques; (2) "profiling" or "categorization" of potential targets; (3) "legibility" of the algorithms leading to the final decision; (4) absence of human control on the decision-making process. For each feature a short recapitulation of existing case-law and scholarship is provided, with a view to

18 Algorithms that rely on "profiling" of the concerned individuals to reach a "decision" are currently employed, for example, to refuse an online credit application or e-recruitment practice; see GDPR, Recital 71, and G. MALGIERI, G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, 2017, p. 253. For an interesting overview of today's pervasiveness of profiling and automated decision-making, see F. PASQUALE, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Cambridge-London, 2015, and A. CHANDER, *The Racist Algorithm?*, in *Michigan Law Review*, 2017, p. 1023-1045.

19 For instance, human operators are often exposed to "machine-bias". See T. CHENGETA, *Defining the emerging notion of "Meaningful Human Control" in Weapon Systems*, in *New York University International Law and Politics*, 2017, p. 852-853.

20 For a definition of "machine-learning", see S. SHALEV-SHWARTZ, S. BEN-DAVID, *Understand Machine Learning. From Theory to Algorithms*, Cambridge, 2014.

21 See O. ULGEN, *Kantian Ethics in the Age of Artificial Intelligence and Robotics*, in *QIL Zoom-In*, 2017, p. 73.

22 See ICRC, *Report of the ICRC Expert Meeting, Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*, 9 May 2014, available at: https://www.icrc.org/en/document/report-icrc-meeting-autonomous-weapon-systems-26-28-march-2014; id., *Report of the ICRC Expert Meeting, Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons*, 15-16 March 2016, available at: https://www.icrc.org/en/publication/4283-autonomous-weapons-systems.

exploring whether similar rules apply – *mutatis mutandis* – when it is mainly the right to life to be at stake.

3.1. Widespread practices of blanket interception of communications and mass collection of data raise growing concern among human rights bodies as far as the right to privacy is concerned.[23] While the right to privacy is notoriously subject to restrictions,[24] strict conditions apply when it comes to public authorities putting in place massive surveillance and the systematic collection and storing of data.[25] A wide spectrum of data is likely to be gathered for the purposes of ATC. Generally speaking, "data" that are relevant under the right to privacy encompass "personal",[26] "sensitive",[27] "biometric"[28] and "big" data.[29] While in the recent decades there has been a growing concern about the need to protect these data, undeniably States still enjoy a certain room for maneuver.

Possibly thanks to the pervasiveness of new technologies of data interception and re-elaboration, however, human rights bodies have raised the bar for such operations to be conducted lawfully. For instance, the ECtHR has scrutinized practices such as the interception of private

23 To catch a glimpse of the existing discourse around these practices, see UNGA Res. 68/167, 21 January 2014, available at: http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/68/167; see also OHCHR, *The Right to Privacy in the Digital Age*, Report of 30 June 2014, UN doc A/HRC/27/37, available at: http://undocs.org/A/HRC/27/37. On the right to privacy generally, see Art. 12 UDHR («No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks»); see Art. 17 ICCPR and *General Comment No. 16, Article 17*, 8 April 1988; Art. 8 ECHR. For an historical overview of the international provisions protecting the right to privacy, see G. DELLA MORTE, *Big Data e protezione internazionale dei diritti umani. Regole e conflitti*, Naples, 2018, p. 75 ff..

24 See *General Comment No. 16*, cit., § 7 («[a]s all persons live in society, the protection of privacy is *necessarily* relative», italics added); Art. 8(2) ECHR (listing three parameters for limiting the right to privacy, namely legality, necessity and proportionality).

25 Such considerations move from the assumption that massive storage of data as well as constant surveillance are dangerous for a democratic society. Yet decades ago the European Court of Human Rights (hereinafter: ECtHR) acknowledged that «an unlimited discretion to subject persons within their jurisdiction to secret surveillance» would risk «undermining or even destroying democracy on the ground of defending it»; see ECtHR, *Klass et al. v. Germany*, No. 5029/71, judgment (plenary), 6 September 1978, § 49. For a discussion on the right to privacy as at particularly at stake when counterterrorism measures are put in place, see *amplius* M. NINO, *Terrorismo internazionale, privacy e protezione dei dati personali*, Naples, 2012.

26 Defined as «any information relating to an identified or identifiable» individual (see GDPR, art. 4(1)). See also the *Modernized Convention for the Protection of Individuals with Regard to the Processing of Personal Data*, adopted by the 128th Session of the Committee of Ministers in Elsinore, Denmark, on 18 May 2018, CM/Inf(2018)15-final (CETS No. 108) (hereinafter: Modernized Convention), art. 2.

27 For a list of personal data that must be considered as «sensitive» inasmuch as revealing certain personal characteristics, see GDPR, art. 9.

28 «personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data» (GDPR, art. 4(14)).

29 It is impossible to account for the immense literature that has been developed on big data so far. See DELLA MORTE, *Big Data*, cit., p. 159-169, with references.

communications[30] and addressed the issue of technologies that allow for the monitoring of private activities by State authorities. In the *Szabó and Vissy* case the Court held that while that «governments resort to cutting-edge technologies» is «a natural consequence of the forms taken by present-day terrorism», it would «defy the purpose of government efforts to keep terrorism at bay … if the terrorist threat were paradoxically substituted for by a perceived threat of unfettered executive power intruding into citizens' private spheres by virtue of uncontrolled yet far-reaching surveillance techniques and prerogatives».[31] It follows that legal safeguards have to be put in place by States when employing such «strategic, large-scale interception» of personal data.[32] What should be taken note of is that in assessing the respondent State's safeguards the Court interpreted the requirement of necessity «in a democratic society» provided for by Art. 8(2) ECHR as requiring the more stringent «strict necessity» *by reason of* the pervasiveness of the cutting-edge surveillance technologies based on «automated and systemic data collection».[33] In short, the rationale the Court leans on seems to be the following: *the more impactful on human rights a surveillance measure is, the more stringent the conditions for the resort thereto need to be*.

In the case of LAWS, the purpose of data collection would be *inter alia* to identify a target for the use of force. In this sense, it would be the right to life – the supreme one, an absolute one and one from which no derogation is allowed – to be put primarily at stake. It follows that requirements for gathering such personal data need to be far more stringent than in other contexts that have come under the scrutiny of human rights bodies so far. A blanket, massive collection of personal data for law-enforcement purposes would virtually expose every individual, in a given geographical area, to a violation of their right to privacy.

3.2. Turning to the decision-making process, human rights bodies have so far voiced concern regarding decisions taken by public authorities on the basis of an automated processing of personal data. In particular, algorithmic processes may put individuals under "categorical" suspicion solely due to their membership – either alleged or effective – in a certain category.

For instance, arresting an individual on the basis that his/her affiliation to an organization is indicative of an higher propensity of committing illegal acts has been considered by the Court of Strasburg as a violation of Art. 5 ECHR.[34] Such practice has important implications on the right to equality and non-discrimination as well.[35] It has been repeatedly outlined how profiling tends to

30 See *Liberty et al. v. the United Kingdom*, judgment, 1 July 2008; *Kennedy v. the United Kingdom*, judgment, 18 May 2010, and more recently *Roman Zakharov v. Russia* [GC], No. 47143/06, judgment, 4 December 2015.

31 See *Szabó and Vissy v. Hungary*, No. 37138/14, judgment, 12 January 2016, §§ 67, 68.

32 *Ibidem*, § 67.

33 Ibidem, § 73 *in principio* and 67.

34 See *Shimovolos v. Russia*, No. 30194/09, judgment, 21 June 2011 (for a case regarding an individual being subject to a deprivation of his personal liberty as his name had been included in a «surveillance database» set up by Russian authorities that employed algorithmic processes to detect «potential extremists»). See also *Ostendorf v. Germany*, No. 15598/08, judgment, 7 March 2013 (for a case regarding a deprivation of personal liberty considered lawful also by virtue of the fact that public authorities had not based their determination on the individual being enlisted in a police database).

35 For a general appraisal of the principle of non-discrimination, see HRCttee, *General Comment No. 18: Non-Discrimination*, 10 November 1989, § 12; see also *Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination*, UN doc A/HRC/29/46, § 22.

expose individuals to negative forms of discrimination, especially based on racial and ethnic origin, religious or other beliefs and political opinions, just to name a few.[36]

The decisive element in assessing the lawfulness of automated decision-making appears to lie in that such process fails to consider the individual *as such*; rather s/he is considered only as belonging to an abstract category. Failure in appreciating the actual situation of an individual is problematic in human rights bodies' recent case-law. Decisions regarding passports withdrawal or imposition of travel bans have been considered as in violation of human rights because public authorities have not been able to carry out a *fresh review* of the individual situation, thus imposing restrictions that could not be characterized as «necessary in a democratic society».[37] In a Dissenting Opinion delivered by three judges of the Court of Strasburg, the categorical treatment of people who happened to be at a certain place in a certain hour – without distinguishing between demonstrators and bystanders – by police forces was deemed in contrast with the right to liberty in that individuals had been treated «*like objects*».[38] Again, the rationale that underlies these judgments is that *each and every decision affecting human rights at a certain degree need to take an individual situation into due account*.

Applying this rationale to LAWS, it has been already outlined that their unique feature lies in that not only data collection and elaboration, but also the final decision – delivering lethal force against an individual – are entrusted to a non-human agent. In order for ATC to be consistent with the prohibition on categorization, it therefore has to be developed in a way that ensures that the *specific* features of each and every situation are taken in due account by LAWS before resorting to force. On closer inspection, such conclusion is implied in the well-known requirements for the use of force that IHRL instruments establish with respect to the right to life: «absolute necessity» and «proportionality».[39] A lethal decision resulting solely from a pattern-of-life analysis is inconsistent with the right to life, as implied by the abovementioned case-law: if life can be taken only when «absolutely necessary», and if arguably decisions taken upon individual categorization are incompatible with the – less stringent – requirement of «necessity in a democratic society», then *a fortiori* a categorical killing as such would be at odds with the right to life.[40] Incidentally, what has

36 See *amplius* Brehm, *Defending the boundaries*, cit., p. 61.

37 See ECtHR, *Battista v. Italy*, No. 43789/09, judgment, 2 December 2014; *Stamose v. Bulgaria*, No. 29713/05, judgment, 27 November 2012 (for cases concerning the freedom of movement as enshrined by Art. 2 Prot. 4 ECHR); *contra*, see *Landvreugd v. the Netherlands*, No. 37331/97, judgment, 4 June 2002, particularly at § 70.

38 See ECtHR, *Austin and Others v. the United Kingdom* [GC], judgment, 15 March 2012, Joint Dissenting Opinion by Judges Tulkens, Spielmann and Garlicki, particularly at § 10.

39 See Art. 2(2) ECHR; Art. 6(1) ICCPR; Art. 4(1) ACHR; Art. 4 ACHPR. For relevant case-law, see *ex multis* ECtHR, *Giuliani and Gaggio vs. Italy*, No. 23458/02, 24 March 2011, § 176 (affirming that «a stricter and more compelling test of necessity must be employed than that normally applicable when determining State action is "necessary in a democratic society" under paragraph 2 of Articles 8 and 11 of the Convention»); HRCttee, *Suarez de Guerrero v Colombia*, Views, Comm no R.11/45, 9 April 1981, Supp No. 40 (A/37/40) at 137 (1982); IACtHR, *Nadege Dorzema et al v Dominican Republic*, judgment (Merits, Reparations and Costs), 24 October 2012.

40 See ECtHR [GC], *Streletz, Kessler and Krenz v. Germany*, Nos. 34044/96, 35532/97 and 44801/98, judgment, 22 March 2001, § 73 (in which consideration was given to «recourse to anti-personnel mines and automatic-fire systems, in view of their automatic and indiscriminate effect, and the categorical nature of the border guards' orders to "annihilate border violators … and protect the border at all costs"»).

been said so far might work as a principled argument against signature strikes in today's drone operations.

However, what current case-law attaches great importance to the absence of a fresh review on an individual situation (an *objective* element) as such rather than to the subject tasked with carrying out such review (a *subjective* element). Restating the point, it does not stem from the foregoing that LAWS employing "categorical", but situationally appropriate, decisions would be proscribed, even if in the decision-making process human deliberation is absent.

3.3. Another issue regarding automated decision-making is that individuals affected by such process may not know how and why algorithms have led to the final decision. It does not seem necessary to underscore how pressing the issue gets when it comes to decisions particularly suitable for affecting basic human rights, such as the right to life.

To begin with, current IHRL applicable to the right to privacy acknowledges the individual's right to *understand* the functioning and the impact of algorithms concerning him/her. For instance, this right is inferable from several provisions regarding data protection adopted in the frameworks of the Council of Europe[41] and the European Union[42], in both cases in binding terms. In particular, the right to know the reasons that underlie an automated decision derives from a set of several rights such as the right to receive *ex ante* information from data controllers and the right to access to information *ex post*, that is after the decision-making process has been undertaken or concluded.[43] In the GDPR's words, the individual has a right to be informed about/be given access to «*meaningful* information about the logic involved, as well as the *significance* and the *envisaged consequences* of such processing» for the individual. Some have proposed to interpret the right as encompassing a right to *legibility*, in the sense that algorithms must be designed and built in a way that they are both transparent and comprehensible to people concerned thereby.[44] Legibility therefore finds a ground of justification in that it is intended to ensure that *automated decision-making does not turn into an unintelligible process*.

Thinking of ATC in terms of legibility turns to be particularly effective through the lens of the right to life as well, for the following reasons. First, an ex-ante knowledge of how algorithms involved in the ATC process appears to satisfy the legal requirement of legality of the use of force as enshrined in IHRL: any deprivation of life must result from the exercise of a power that is provided either in domestic law or in international law, or both.[45] Arguably the protection of the right to life necessitates «an appropriate legal and administrative framework *defining the limited*

41 See the Modernized Convention, art. 9; *Explanatory Report to the Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, CM(2018)2-addfinal (CETS No. 223) (hereinafter: Explanatory Report), §§ 71-83.

42 See GDPR, arts. 13, 14 (right to notification), 15 (right to access) and art. 22 (right not to be subject to a decision based solely on automated processing, including profiling). For the sake of clarity, references to the GDPR are intended only to draw analogies and not to suggest that it is applicable to our subject matter: Art. 2.2.d clearly excludes such chance.

43 See for instance GDPR arts. 13(2)(f), 14(2)(g) and 15.

44 See in particular MALGIERI, COMANDÈ, *Why a Right to Legibility*, cit., p 245.

45 Art. 6(1) ICCPR, Art. 2(1) ECHR, and Art. 4(1) ACHR all require expressly a legal basis, whereas the ACtHPR does not.

circumstances in which law enforcement officials may use force and firearms, in the light of the relevant international standards».[46] If such regulatory framework is obscure, or fails to provide individuals with understandable – "legible" – indications about the conditions in which LAWS resort to lethal force in law-enforcement situations (the abovementioned «limited circumstances»), the requirement of legality will be hardly met. The issue gets all the more troublesome once the scenario is taken into consideration where ATC employ self-learning algorithms, which as explained in the foregoing ensure low rates of predictability.

Second, ex-post explanation about the process that has actually led to a specific automated decision comes to the fore from the perspective of procedural obligations under IHRL. Procedural obligations are a particular form of positive obligations involving the duty to invest into (alleged) violations of a right and to prosecute those who are responsible; in this sense, it is «not an obligation of result but of means only».[47] In order for an investigation to comply with human rights standards, it allegedly has to be *capable* of leading to a determination of whether the force used was justified *in the circumstances*.[48] In the dynamic of ATC processing, it is therefore of paramount importance that public authorities provide an intelligible account of how an automated process has worked; a failure to "explain how",[49] *reddere rationem*, may lead to responsibility under IHRL.[50] With respect to the use of force, if LAWS employ an ATC technique that does not allow their users to understand how and why the machine has taken that particular decision – which may be the case, again, when self-learning algorithms are employed –, it follows that the right to life would be violated under the procedural tenet as well.[51]

To sum, the right to legibility as pushed forward by scholarship is a useful tool for testing the compatibility of ATC with IHRL also with respect the right to life. Failure to provide information and explanation about how and why algorithms work and lead to a certain decision,

46 See ECtHR, *Giuliani and Gaggio v. Italy*, No. 23458/02, judgment, 24 March 2011, § 209.

47 See ECtHR [GC], *Šilih v. Slovenia*, § 193. See also IACtHR, *Cantoral Huamaní and Garcia Santa Cruz v. Peru*, Preliminary objection, merits, reparations and costs, 10 July 2007, Series C 167 (2007), § 131, and *Pueblo Bello Massacre v Colombia*, 31 January 2006, Series C 140 (2006), § 143. Its aims are: (i) ensuring that those responsible are brought to justice; (ii) promoting accountability and preventing impunity; (iii) avoiding denial of justice; (iv) eventually drawing necessary lessons for revising practices and policies with a view to avoiding repeated violations.

48 ECtHR, *Isayeva v. Russia*, No. 57950/00, judgment, 24 February 2005, §§ 221-223 (for a case where the ineffectiveness of the investigation was predicated in that it had made «few attempts to find an explanation for … serious and credible allegations», thus placing the need for an explanation about the use of force at the center of the right to life under its procedural tenet).

49 See ECtHR, *Khodorkovskiy and Lebedev v. Russia*, No. 11082/06 and 13772/05, judgment, 25 July 2013, § 848 (discussing the algorithmic method used for distributing convicted individuals among prisons).

50 The issue has also been tackled from the standpoint of international humanitarian law. See P. MARGULIES, *Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts*, in J. OHLIN (ed), *Research Handbook on Remote Warfare*, Cheltenham-Northampton, 2017, p. 23 (underscoring that the onus is on the State to provide adequate details about decision-making processes at large).

51 Some scholars that support the development and deployment of LAWS are apparently prone to accepting the idea that human operators may not be able to understand how and why LAWS take specific lethal decisions. See for instance K. ANDERSON, D. REISNER, M. WAXMAN, *Adapting the Law of Armed Conflict to Autonomous Weapons Systems*, in *International Law Studies*, 2014, p. 394 (arguing that «as machine-learning and artificial intelligence technologies develop, it is becoming increasingly clear that human beings may not necessarily always be able to understand how (and possibly why) autonomous systems make decisions»).

both *ex ante* and *ex post*, will expose public authorities employing LAWS to responsibility under IHRL. Again, as we explained in our analysis of categorical decision-making, legibility does not imply human presence at the single deliberation of employing lethal force against a human target. LAWS ensuring a satisfying level of legibility would thus be acceptable under IHRL also absent human intervention in the decision-making process.

3.4. The last issue raised by ATC has to do with another right that is well-accepted today in the field of data protection, namely the individual right not to be subject to a decision *significantly* affecting them based *solely* on an automated processing of their personal data.[52] The need for maintaining human presence in the decision-making process – so far dispensable – would be of some relevance eventually.

The rationale and the scope of this right must be ascertained in the light of the relevant provisions. First, it does *not* qualify as absolute as it is permissible to make exceptions to it in certain cases, notably upon express authorization by the relevant public authorities.[53] However, on closer inspection, relevant provisions are clear in requiring that even in those cases a minimum of safeguard measures must be guaranteed to the affected individuals,[54] among which are: the right to obtain *human intervention*; to express one's point of view; to obtain an explanation of the decision reached after such assessment; and to challenge the decision.[55] Importantly, the notion of "human intervention" can be interpreted in a broad sense, encompassing not only cases in which human intervention is absent in the automated decision-making process, but also cases of *nominal* interventions (i.e. those in which humans exercise no real influence on the outcome of the decision).[56] In other words, human intervention must be *meaningful* for ensuring that the right in question is respected *in concreto*. It must be recalled that such interpretation of human intervention is actually contrasted by some authors, whose take is that also "nominal" human intervention may

52 See GDPR, art. 22(1): «The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her»; Modernized Convention, art. 9(1): «Every individual shall have a right: (a) not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration»; Explanatory Report, § 75: «It is essential that an individual who may be subject to a purely automated decision has the right to challenge such a decision by putting forward, in a meaningful manner, his or her point of view and arguments».

53 See for instance GDPR, Recital 71 («decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law to which the controller is subject, including for fraud and tax-evasion monitoring and prevention purposes conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller, or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent»).

54 See Explanatory Report, § 75 («However, an individual cannot exercise this right if the automated decision is authorised by a law … which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests»).

55 See GDPR, Recital 71 and 22(1). Importantly, art. 22(1) does not refer to the right to receive explanation, while Recital 71 does; it seems however more appropriate to interpret the former provision in the light of the latter, which is consistent with well-known interpretive criteria.

56 See MALGIERI, COMANDÈ, *Why a Right to Legibility*, cit., p. 251-252 (underscoring how only such interpretation is able to prevent an illegitimate discrimination between «fully automated systems» and «scoring systems» in performing the same activities and potentially producing the same outcome).

be deemed sufficient.[57] Be that as it may, the rationale of the provision is to protect individuals when affected by particular processes, for example "scoring" mechanisms employed by bank when assessing creditworthiness, in order to avoid discriminatory treatment.[58] Such safeguard essentially consists in introducing human components such as *critical sensibility* and *judgment capabilities* in the decision-making process: human discretion is considered as the strongest bastion against discrimination and unfair treatment of data.

Whether such a rationale can be extended (yet *mutatis mutandis*) to ATC is questionable, as the divergence between the former contexts (e.g. credit scoring) and those where LAWS will be operated in our scenario (e.g. law-enforcement operations) may warn against drawing rash and far-fetched analogies. An *a fortiori* reasoning here, albeit appealing, may risk obliterating that divergence. On the one hand, ATC undeniably produces legal effects concerning the individual and «*significantly* affects» him/her, given that it is the very right to life to be at stake.[59] On the other hand, the difference in operational contexts should not be underestimated: law-enforcement agents are often required to take split-second decisions based on the individual's conduct in specific circumstances – in short, where operational tempo may not allow for a human operator to review a decision made by a LAWS.[60] The temporal dimension, in short, can be regarded as a cutting-point element of distinction from other contexts where data processing normally occurs.

The right not to be subject to a solely automated decision is interestingly mirrored in the recently coined notion of Meaningful Human Control (MHC), a veritable *punctum dolens* in the current debate on LAWS.[61] Virtually all States, NGOs and representatives of the civil society consider MHC as the basic requirement for any weapons system, as it requires human deliberation to be present at critical decisions made by LAWS (such as target selection and engagement).

57 See S. WACHTER, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 2017, p. 92 («the phrase "solely" suggests even some nominal human involvement may be sufficient»).

58 See MALGIERI, COMANDÈ, *Why a Right to Legibility*, cit., p. 251.

59 See GDPR, Art. 22(1); for a commentary, see MALGIERI, COMANDÈ, *Why a Right to Legibility*, cit., at 252 (explicitly interpreting the provision as encompassing all practices having any influence on «human rights and constitutional rights of individuals» and pinpointing that such human rights at stake «cannot be considered a "numerus clausus"»).

60 For an appropriate example, see C. HEYNS, *Human Rights and the Use of Autonomous Weapons Systems During Domestic Law Enforcement*, in *Human Rights Quarterly*, 2016, p. 358 (arguing that LAWS programmed with facial recognition software whereby they can use lethal force against an hostage-taker exposing himself for a split second are lawful, as long as the human alternative is suboptimal and human control is ensured over the operation; in these particular circumstances, given the narrow-spatial boundaries of the operation, it does not seem to be at odds with the right to life).

61 For an overview of the notion, its scope and its purposes, see U.N. Institute For Disarmament Research (UNIDIR), *The Weaponization Of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move The Discussion Forward*, 2015, p. 4, http://www.unidir.org/en/publications/the-weaponization-of-increasingly-autonomous-technologies-considering-how-meaningful-human-control-might-move-the-discussion-forward;
CHENGETA, *Defining the emerging notion*, cit., p. 833 ff.. The first conceptualization of MHC has been pushed forward by the British NGO Article 36: see the Report *Key Areas For Debate On Autonomous Weapon Systems*, 2014, available at: http://www.article36.org/wp-content/uploads/2014/05/A36-CCW-May-2014.pdf. The importance of MHC in the debate is captured by C. Heyns, *Autonomous weapons in armed conflict and the right to a dignified life: an African perspective*, in *South African Journal On Human Rights*, 2017, p. 50 (arguing how MHC is perceived as «the dividing line between acceptable and not acceptable machine autonomy»).

However, there regrettably is ample disagreement about how to understand the notion: some consider it necessary to retain a human operator as a general supervisor with little chance to intervene in the single lethal decision (a "broad" understanding of MHC),[62] while others call for a more substantial role played by the human operator (a "narrow" understanding of MHC).[63]

The dichotomy of narrow and broad understandings of MHC reproduces the contrast around "nominal" human intervention in automated decision-making. Notwithstanding the respective different contexts it seems that the underlying rationales of meaningful human intervention relating to the right not to be subject solely to an automated decision and the specific notion of MHC actually overlap. It has been argued that the «major purpose» of a notion of MHC is to fill any possible «accountability gap».[64] Having a human agent in a relationship of control and dependence with LAWS is perceived as the only way to ensure accountability in cases where LAWS's conduct results in a violation of relevant law; a broad understanding of MHC (for example, limited to having a human presence at the act of pre-programming a LAWS) replicates similar drawbacks associated with "nominal" human intervention, namely higher risk of biases,[65] poorer situational understanding, lack of critical sensibility and judgment.[66]

To sum up, it is true that the right not to be subject to solely automated decision-making founds its place in the field of data protection and privacy, and expanding it may seem an improper use of the *a fortiori* argument. However, on closer inspection its rationale is close to the one that inspires the notion of MHC. In both cases, meaningful human presence ensures intervention at the outcome of an automated decision-making process, which allows for the human "component" to be part of the equation. In the case of LAWS, the requirement of MHC reflects the more specific need for accountability in cases where a particular use of force was not permitted.

4. Summarizing what has been argued so far, for ATC to be IHRL-compliant it must: (1) involve data collection only when «strictly necessary» as far as the right to privacy is concerned; (2) employ categories that are consistent with the right to life's requirements of «absolute necessity» and «proportionality»; (3) ensure a sufficient degree of legibility, i.e. an understanding of the actual functioning of the process that leads to the use of force; (4) allow for *meaningful* human intervention to the extent that accountability gaps are eliminated. All these requirements are quite stringent, especially if applied to existing technology; this is why most scholars who contrast the

62 See for instance M. SCHMITT, *Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics*, in *Harvard National Security Journal*, 2013, p. 1-38.

63 See for instance N. SHARKEY, *Staying in the loop: human supervisory control of weapons*, in N. BHUTA, S. BECK, R. GEIB, H. LIU, C. KREB (eds.), *Autonomous Weapons Systems*, Cambridge, 2016, p. 34 ff..

64 See CHENGETA, *Defining the emerging notion*, cit., p. 883 and *passim*.

65 For a brief explanation of «automation bias» and risks associated therewith, see CHENGETA, *Defining the emerging notion*, cit., p. 853-854 (concluding that «mere involvement of a human being in the loop» must be rejected as it does not ensure that a human operator retains effective control over the weapon).

66 Paraphrasing MALGIERI, COMANDÈ, *Why a Right to Legibility*, cit., p. 252; see *amplius* CHENGETA, *Defining the emerging notion*, cit., p. 872 ff. (arguing that «the analysis of facts on the battlefield and fitting them to pre-defined parameters» – which is what constitutes «the real decision-making to kill» – is not sufficient for the purposes of MHC, a notion that requires that human agents be «in control of the system for each individual attack because such control is central to establishing the responsibility of combatants»).

development of LAWS are inclined to do so on the basis of their actual feasibility (i.e. technological or pragmatic objection).[67]

There is however an alternative understanding of ATC's implications on human rights, focusing more on the uniqueness of autonomous decision-making as a process where human deliberation may be absent with regard to a *specific* use of force. It is an argument based on *human dignity* – possibly one of the most contrasted, debated notions in the contemporary discourse around human rights.[68] Some scholars argue that the lack of human presence at force delivery against an individual entails a violation of human dignity *as such*, as targets are treated as mere objects.[69] The essence of this kind of arguments lies in that non-human decision-makers «have no understanding of the importance of life, and of the implications of taking it»:[70] emphasis here is therefore not placed on accountability, but rather on the capacity of *understanding* the gravity of a decision. Appealing to concepts such as "human dignity" or "humanity" raises however a bunch of issues, in first place as the meaning of such expressions is contested not only in the legal debate, but also – and foremost – in the moral one.

67 The essence of such objection has been recently captured by HEYNS, *Autonomous weapons in armed conflict and the right to a dignified life*, cit., p. 58: «the requirement of meaningful human control is needed not as a matter of principle, but in order to secure accuracy in targeting and as a basis for legal reform. Clearly, this approach is conditional on technological developments, and may or not be trumped depending on such progress».

68 To recapitulate here the immense literature on this subject would be impossible. Suffice it to recall the following contributions: O. SCHACHTER, *Human Dignity as a Normative Concept*, in *American Journal of International Law*, 1983, p. 848 ff.; J. FROWEIN, *Human Dignity in International Law*, in D. KRETZMER, E. KLEIN (eds.), *The Concept of Human Dignity in Human Right Discourse*, The Hague, 2002, p. 121 ff.; for a ECHR-centered focus, see J. COSTA, *Human Dignity in the Jurisprudence of the European Court of Human Rights*, in C. MCCRUDDEN (ed.), *Understanding Human Dignity*, Oxford, 2014, p. 393 ff..

69 See *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, cit., particularly at § 95 («[d]eploying LARs has been depicted as treating people like "vermin", who are "exterminated." These descriptions conjure up the image of LARs as some kind of mechanized pesticide»); P. ASARO, *Jus Superveniens: robotic weapons and the Martens Clause*, in R. CALO, M. FROOMKIN, I. KERR, *Robot Law*, Cheltenham-Northampton, 2016, p. 385 («this also relates to the question of human dignity. If a combatant is to die with dignity, there must be some sense in which that death is meaningful. In the absence of an intentional and meaningful decision to use violence, the resulting deaths are meaningless and arbitrary, and the dignity of those killed is significantly diminished»); O. ULGEN, *Human Dignity in an Age of Autonomous Weapons: Are We in Danger of Losing an "Elementary Consideration of Humanity"?* , in *ESIL Conference Paper*, 2016, p. 8 («[h]uman targets are denied the status of rational agents with autonomy of will, and arbitrarily deemed irrational agents subject to extrajudicial killings or sub-humans not worthy of human face-to-face contact»). A dignity-based approach is defended by, *inter alios*, the Holy See: see for instance the Statement at the 2017 GGE, unfortunately unavailable on the CCW website: «[a] machine is only a complex set of circuits and this material system cannot in any case become a truly morally responsible agent. In fact, for a machine, a human person is only a datum, a set of numbers among others».

70 See HEYNS, *Autonomous weapons in armed conflict and the right to a dignified life*, cit., p. 58.

To address the issue in a satisfying manner is not possible here; probably the most appropriate conclusion would be at least to leave a door open for a more principled reflection that could take moral considerations into account. This is all the more imperative in a time when humanity faces an unprecedented scenario of «death by algorithm»:[71] the discourse around ATC, the (mis-)use of "big data" and compliance with IHRL can at most *put off* – and not *wipe out* – the intrinsic moral dilemma raised by this new technology and the risks associated thereto.

DIEGO MAURI

71 The quote is from HEYNS, *Autonomous weapons in armed conflict and the right to a dignified life*, cit., p. 48.