28

On 'Responsible AI' in War

Exploring Preconditions for Respecting International Law in Armed Conflict

Dustin A. Lewis

I. INTRODUCTION

In this chapter, I seek to help strengthen cross-disciplinary linkages in discourse concerning 'responsible Artificial Intelligence (AI)'. To do so, I explore certain aspects of international law pertaining to uses of AI-related tools and techniques in situations of armed conflict.

At least five factors compel increasingly urgent consideration of these issues by governments, scientists, engineers, ethicists, and lawyers, among many others. One aspect concerns the nature and the growing complexity of the socio-technical systems through which these technologies are configured. A second factor relates to the potential for more frequent – and possibly extensive – use of these technologies in armed conflicts. Those applications may span such areas as warfighting, detention, humanitarian services, maritime systems, and logistics. A third issue concerns potential challenges and opportunities concerning the application of international law to employments of AI-related tools and techniques in armed conflicts. A fourth dimension relates to debates around whether or not the existing international legal framework applicable to armed conflicts sufficiently addresses ethical concerns and normative commitments implicated by AI – and, if it does not, how the framework ought to be adjusted. A fifth element concerns a potential 'double black box' in which humans encase technical opacity in military secrecy.

One way to seek to help identify and address potential issues and concerns in this area is to go 'back to the basics' by elaborating some key elements underpinning legal compliance, responsibility, and agency in armed conflict. In this chapter, I aim to help illuminate some of the preconditions arguably necessary for respecting international law with regard to employments of AI-related tools and techniques in armed conflicts. By respecting international law, I principally mean two things: (1) applying and observing international law with regard to relevant conduct and (2) facilitating incurrence of responsibility for violations arising in connection with relevant conduct. (The latter might be seen either as an integral element or a corollary of the former.) Underlying my exploration is the argument that there may be descriptive and normative value in framing part of the discussion related to 'responsible AI' in terms of discerning and instantiating the preconditions necessary for respecting international law.

I proceed as follows. In Section II, I frame some contextual aspects of my inquiry. In Section III, I sketch a brief primer on international law applicable to armed conflict. In Section IV, I set out some of the preconditions arguably necessary to respect international law. In Section V, I briefly conclude.

https://doi.org/10.1017/9781009207898.037 Published online by Cambridge University Press

Two caveats ought to be borne in mind. The first caveat is that the bulk of the research underlying this chapter drew primarily on English-language materials. The absence of a broader examination of legal materials, scholarship, and other resources in other languages narrows the study's scope. The second caveat is that this chapter seeks to set forth, in broad-brush strokes, some of the preconditions arguably underpinning respect for international law.[1] Therefore, the analysis and the identification of potential issues and concerns are far from comprehensive. Analysis in respect of particular actors, armed conflicts, or AI-related tools and techniques may uncover (perhaps numerous) additional preconditions.

II. FRAMING

In this section, I frame some contextual aspects of my inquiry. In particular, I briefly outline some elements concerning definitions of AI. I also enumerate some existing and anticipated uses for AI in armed conflict. Next, I sketch the status of international discussions on certain military applications of possibly related technologies. And, finally, I highlight issues around technical opacity combined with military secrecy.

1. *Definitional Parameters*

Terminological inflation may give rise to characterizations of various technologies as 'AI' even where those technologies do not fall into recognized definitions of AI. Potentially complicating matters further is that there is no agreed definition of AI expressly laid down in an international legal instrument applicable to armed conflict.

For this chapter, I will assume a relatively expansive definition of AI, one drawn from my understanding – as a non-scientific-expert – of AI science broadly conceived.[2] It may be argued that AI science pertains in part to the development of computationally-based understandings of intelligent behaviour, typically through two interrelated steps. One step relates to the determination of cognitive structures and processes and the corresponding design of ways to represent and reason effectively. The other step concerns developing (a combination of) theories, models, data, equations, algorithms, or systems that 'embody' that understanding. Under this approach, AI systems are sometimes conceived as incorporating techniques or using tools that enable systems to 'reason' more or less 'intelligently' and to 'act' more or less 'autonomously.' The systems might do so by, for example, interpreting natural languages and visual scenes; 'learning' (in the sense of training); drawing inferences; or making 'decisions' and taking action on those 'decisions'. The techniques and tools might be rooted in one or more of the following

[1] My analysis in this chapter – and especially Section IV – draws heavily on, and reproduces certain text from, a DA Lewis, 'Preconditions for Applying International Law to Autonomous Cyber Capabilities', in R Liivoja and A Väljataga (eds), *Autonomous Cyber Capabilities under International Law* (NATO Cooperative Cyber Defence Centre of Excellence, 2021). Both the current chapter and that piece draw on the work of a research project at the Harvard Law School Program on International Law and Armed Conflict titled 'International Legal and Policy Dimensions of War Algorithms: Enduring and Emerging Concerns' (Harvard Law School Program on International Law and Armed Conflict, 'Project on International Legal and Policy Dimensions of War Algorithms: Enduring and Emerging Concerns' (November 2019) https://pilac.law.harvard.edu/international-legal-and-policy-dimensions-of-war-algorithms). That project seeks to strengthen international debate and inform policy-making on the ways that AI and complex computer algorithms are transforming, and have the potential to reshape, war.

[2] This paragraph draws extensively on DA Lewis, 'Legal Reviews of Weapons, Means and Methods of Warfare Involving Artificial Intelligence: 16 Elements to Consider' (*ICRC Humanitarian Law and Policy Blog*, 21 March 2019) https://blogs.icrc.org/law-and-policy/2019/03/21/legal-reviews-weapons-means-methods-warfare-artificial-intelligence-16-elements-consider/ (hereafter Lewis, 'Legal Reviews'); see also W Burgard, Chapter 1, in this volume.

https://doi.org/10.1017/9781009207898.037 Published online by Cambridge University Press

methods: those rooted in logical reasoning broadly conceived, which are sometimes also referred to as 'symbolic AI' (as a form of model-based methods); those rooted in probability (also as a form of model-based methods); or those rooted in statistical reasoning and data (as a form of data-dependent or data-driven methods).

2. *Diversity of Applications*

Certain armed forces have long used AI-related tools and techniques. For example, in relation to the Gulf War of 1990–91, the United States employed a program called the Dynamic Analysis and Replanning Tool (DART), which increased efficiencies in scheduling and making logistical arrangements for the transportation of supplies and personnel.[3]

Today, existing and contemplated applications of AI-related tools and techniques related to warfighting range widely.[4] With the caveat concerning terminological inflation noted above in mind, certain States are making efforts to (further) automate targeting-related communications support,[5] air-to-air combat,[6] anti-unmanned-aerial-vehicle countermeasures,[7] so-called loitering-attack munitions,[8] target recognition,[9] and analysis of intelligence, reconnaissance, and surveillance sources.[10] Armed forces are developing machine-learning techniques to generate

[3] See M Bienkowski, 'Demonstrating the Operational Feasibility of New Technologies: the ARPI IFDs' (1995) 10(1) *IEEE Expert* 27, 28–29.

[4] See, e.g., MAC Ekelhof and G Paoli, 'The Human Element in Decisions about the Use of Force' (UN Institute for Disarmament Research, 2020) https://unidir.org/publication/human-element-decisions-about-use-force; E Kania, '"AI Weapons" in China's Military Innovation' (Brookings Institution, April 2020) www.brookings.edu/wp-content/uploads/2020/04/FP_20200427_ai_weapons_kania_v2.pdf; MAC Ekelhof and GP Paoli, 'Swarm Robotics: Technical and Operational Overview of the Next Generation of Autonomous Systems' (2020) UN Institute for Disarmament Research https://unidir.org/sites/default/files/2020-04/UNIDIR_Swarms_SinglePages_web.pdf; MAC Ekelhof, 'The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting' (PhD Dissertation, Vrije Universiteit 2019); KM Sayler, 'Artificial Intelligence and National Security' (21 November 2019) Congressional Research Service Report No R45178 https://fas.org/sgp/crs/natsec/R45178.pdf; International Committee of the Red Cross, 'Autonomy, Artificial Intelligence and Robotics: Technical Aspects of Human Control' (ICRC Report, August 2019) www.icrc.org/en/download/file/102852/autonomy_artificial_intelligence_and_robotics.pdf; United Nations Institute for Disarmament Research, 'The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence – A Primer for CCW delegates' (2018) UNIDIR Paper No 8 https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-artificial-intelligence; MAC Ekelhof, 'Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control the Lens of Military Targeting' (2018) 73 *Nav War Coll Rev* 61; P Sharre, *Army of One* (2018) 27–56; V Boulanin and M Verbruggen, 'Mapping the Development of Autonomy in Weapons Systems' (Stockholm International Peace Research Institute, 2017) www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf.

[5] 'DOD Official Briefs Reporters on Artificial Intelligence Developments' (Transcript of Nand Mulchandani, 8 July 2020) www.defense.gov/Newsroom/Transcripts/Transcript/Article/2270329/dod-official-briefs-reporters-on-artificial-intelligence-developments/.

[6] K Reichmann, 'Can Artificial Intelligence Improve Aerial Dogfighting?' (*C4ISRNET*, 7 June 2019) www.c4isrnet.com/artificial-intelligence/2019/06/07/can-artificial-intelligence-improve-aerial-dogfighting/.

[7] See Industry News Release, 'Air Force to Deploy Citadel Defense Titan CUAS Solutions to Defeat Drone Swarms' *Defense Media Network* (17 September 2019) www.defensemedianetwork.com/stories/air-force-to-deploy-citadel-defense-titan-cuas-solutions-to-defeat-drone-swarms/.

[8] See, e.g., D Gettinger and AH Michel, 'Loitering Munitions' (Center for the Study of the Drone, 2 February 2017) https://dronecenter.bard.edu/files/2017/02/CSD-Loitering-Munitions.pdf.

[9] On legal aspects of automatic target recognition systems involving 'deep learning' methods, see JG Hughes, 'The Law of Armed Conflict Issues Created by Programming Automatic Target Recognition Systems Using Deep Learning Methods' (2018) 21 *YBIHL* 99.

[10] See, e.g., N Strout, 'Inside the Army's Futuristic Test of Its Battlefield Artificial Intelligence in the Desert' (*C4ISRNET*, 25 September 2020) www.c4isrnet.com/artificial-intelligence/2020/09/25/the-army-just-conducted-a-massive-test-of-its-battlefield-artificial-intelligence-in-the-desert/.

targeting data.[11] Prototypes of automated target-recognition heads-up displays are also under development.[12] Rationales underlying these efforts are often rooted in military doctrines and security strategies that place a premium on enhancing speed and agility in decision-making and tasks and preserving operational capabilities in restricted environments.[13]

In the naval context, recent technological developments – including those related to AI – afford uninhabited military maritime systems, whether on or below the surface, capabilities to navigate and explore with less direct ongoing human supervision and interaction than before. Reportedly, for example, China is developing a surface system called the JARI that, while remotely controlled, purports to use AI to autonomously navigate and undertake combat missions once it receives commands.[14]

The likelihood seems to be increasing that AI-related tools and techniques may be used to help make factual determinations as well as related evaluative decisions and normative judgements around detention in armed conflict.[15] Possible antecedent technologies include algorithmic filtering of data and statistically-based risk assessments initially created for domestic policing and criminal-law settings. Potential applications in armed conflict might include prioritizing military patrols, assessing levels and kinds of threats purportedly posed by individuals or groups, and determining who should be held and when someone should be released. For example, authorities in Israel have reportedly used algorithms as part of attempts to obviate anticipated attacks by Palestinians through a process that involves the filtering of social-media data, resulting in over 200 arrests.[16] (It is not clear whether or not the technologies used in that context may be characterized as AI.)

It does not seem to strain credulity to anticipate that the provision of humanitarian services in war – both protection and relief activities[17] – may rely in some contexts on AI-related tools and techniques.[18] Applications that might be characterized as relying on possible technical antecedents to AI-related tools and techniques include predictive-mapping technologies used to inform populations of outbreaks of violence, track movements of armed actors, predict population movements, and prioritize response resources.[19]

[11] See N Strout, 'How the Army Plans to Use Space and Artificial Intelligence to Hit Deep Targets Quickly' *Defense News* (5 August 2020) www.defensenews.com/digital-show-dailies/smd/2020/08/05/how-the-army-plans-to-use-space-and-artificial-intelligence-to-hit-deep-targets-quickly/.

[12] See J Keller, 'The Army's Futuristic Heads-Up Display Is Coming Sooner than You Think' (*Task & Purpose*, 20 November 2019) https://taskandpurpose.com/military-tech/army-integrated-visual-augmentation-system-fielding-date.

[13] See CP Trumbull IV, 'Autonomous Weapons: How Existing Law Can Regulate Future Weapons' (2020) 34 *EmoryILR* 533, 544–550.

[14] See L Xuanzun, 'China Launches World-Leading Unmanned Warship' Global Times (22 August 2019) www.globaltimes.cn/content/1162320.shtml.

[15] See DA Lewis, 'AI and Machine Learning Symposium: Why Detention, Humanitarian Services, Maritime Systems, and Legal Advice Merit Greater Attention' (*Opinio Juris*, 28 April 2020) http://opiniojuris.org/2020/04/28/ai-and-machine-learning-symposium-ai-in-armed-conflict-why-detention-humanitarian-services-maritime-systems-and-legal-advice-merit-greater-attention/ (hereafter Lewis, 'AI and Machine Learning'); T Bridgeman, 'The Viability of Data-Reliant Predictive Systems in Armed Conflict Detention' (*ICRC Humanitarian Law and Policy Blog*, 8 April 2019) https://blogs.icrc.org/law-and-policy/2019/04/08/viability-data-reliant-predictive-systems-armed-conflict-detention/ ; A Deeks, 'Detaining by Algorithm' (*ICRC Humanitarian Law and Policy Blog*, 25 March 2019) https://blogs.icrc.org/law-and-policy/2019/03/25/detaining-by-algorithm/; A Deeks, 'Predicting Enemies' (2018) 104 Virginia LR 1529.

[16] CBS News, 'Israel Claims 200 Attacks Predicted, Prevented with Data Tech' *CBS News* (12 June 2018) www.cbsnews.com/news/israel-data-algorithms-predict-terrorism-palestinians-privacy-civil-liberties/.

[17] See ICRC, *Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field* (2nd ed. 2016) paras 807–821 https://ihl-databases.icrc.org/ihl/full/GCI-commentary (hereafter ICRC, *Commentary*).

[18] See Lewis, 'AI and Machine Learning' (n 15).

[19] See UNHCR, 'The Jetson Story' (UN High Commissioner for Refugees Innovation Service) http://jetson.unhcr.org/story.html; N Manning, 'Keeping the Peace: The UN Department of Field Service's and Peacekeeping Operations Use of Ushahidi' (*Ushahidi Blog*, 8 August 2018) www.ushahidi.com/blog/2018/08/08/keeping-the-peace-the-un-depart

3. *International Debates on 'Emerging Technologies in the Area of Lethal Autonomous Weapons Systems'*

Perhaps especially since 2013, increased attention has been given at the international level to issues around autonomous weapons. Such weapons may or may not involve AI-related tools or techniques. A significant aspect of the debate appears to have reached a kind of normative deadlock.[20] That impasse has arisen in the recent main primary venue for intergovernmental discourse: the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems (GGE), which was established under the Convention on Certain Conventional Weapons (CCW)[21] in 2016.

GGE debates on the law most frequently fall under three general categories: international humanitarian law/law of armed conflict (IHL/LOAC) rules on the conduct of hostilities, especially on distinction, proportionality, and precautions in attacks; reviews of weapons, means, and methods of warfare;[22] and individual and State responsibility.[23] (The primary field of international law developed by States to apply to conduct undertaken in relation to armed conflict is now often called IHL/LOAC; this field is sometimes known as the *jus in bello* or the laws of war.)

Perhaps the most pivotal axis of the current debate concerns the desirability (or not) of developing and instantiating a concept of 'meaningful human control' or a similar formulation over the use of force, including autonomy in configuring, nominating, prioritizing, and applying force to targets.[24] A close reading of States' views expressed in the GGE suggests that

ment-of-field-services-and-peacekeeping-operations-use-of-ushahidi. See also A Duursma and J Karlsrud, 'Predictive Peacekeeping: Strengthening Predictive Analysis in UN Peace Operations' (2019) 8 *Stability IJ Sec & Dev* 1.

[20] This section draws heavily on DA Lewis, 'An Enduring Impasse on Autonomous Weapons' (*Just Security*, 28 September 2020) www.justsecurity.org/72610/an-enduring-impasse-on-autonomous-weapons/ (hereafter Lewis, 'An Enduring Impasse').

[21] Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (with Protocols I, II, and III) (signed 10 October 1980, entry into force 2 December 1983) 1342 UNTS 137.

[22] See GGE, 'Questionnaire on the Legal Review Mechanisms of New Weapons, Means and Methods of Warfare' (29 March 2019) Working Paper by Argentina to the Group of Governmental Experts on Lethal Autonomous Weapons Systems CCW/GGE.1/2019/WP.6; GGE, *'The Australian Article 36 Review Process' (30 August 2018) Working Paper by Australia to the* Group of Governmental Experts on Lethal Autonomous Weapons Systems CCW/GGE.2/2018/WP6; GGE, *'Strengthening of the Review Mechanisms of a New Weapon, Means or Methods of Warfare' (4 April 2018) Working Paper by Argentina to the Group of Governmental Experts on Lethal Autonomous Weapons Systems* CCW/GGE.1/2018/WP2; GGE, *'Weapons Review Mechanisms' (7 November 2017) Working Paper by the Netherlands and Switzerland to the Group of Governmental Experts on Lethal Autonomous Weapons Systems* CCW/GGE.1/2017/WP5; German Defense Ministry, 'Statement on the Implementation of Weapons Reviews under Article 36 Additional Protocol I by Germany' (The Convention on Certain Conventional Weapons (CCW) Third Informal Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, 11–15 April 2016) https://perma.cc/4EFG-LCEM; M Meier, 'US Delegation Statement on "Weapon Reviews"' (The Convention on Certain Conventional Weapons (CCW) Informal Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, 13 April 2016) www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2016/meeting-experts-laws/statements/13April_US.pdf.

[23] M Brenneke, 'Lethal Autonomous Weapon Systems and Their Compatibility with International Humanitarian Law: A Primer on the Debate' (2018) 21 *YBIHL* 59.

[24] See M Wareham 'Stopping Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control' *(Human Rights Watch, August 2020)* www.hrw.org/sites/default/files/media_2020/08/arms0820_web.pdf; AM Eklund, 'Meaningful Human Control of Autonomous Weapon Systems: Definitions and Key Elements in the Light of International Humanitarian Law and International Human Rights Law' (Swedish Defense Research Agency FOI, February 2020) www.fcas-forum.eu/publications/Meaningful-Human-Control-of-Autonomous-Weapon-Systems-Eklund.pdf; V Boulanin and others, 'Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control' (Stockholm International Peace Research Institute and International Committee of

governments hold seemingly irreconcilable positions beyond some generically formulated principles, at least so far, on whether existing law is fit for purpose or new law is warranted.[25] That said, there might be a large enough contingent to pursue legal reform, perhaps outside of the CCW.

4. *Technical Opacity Coupled with Military Secrecy*

Both inside and outside of the GGE, armed forces continue to be deeply reluctant to disclose how they configure sensors, algorithms, data, and machines, including as part of their attempts to satisfy legal rules applicable in relation to war. In a nutshell, a kind of 'double black box' may emerge where human agents encase technical opacity in military secrecy.[26]

The specific conduct of war as well as military-technological capabilities are rarely revealed publicly by States and non-state parties to armed conflicts. Partly because of that, it is difficult for people outside of armed forces to reliably discern whether new technological affordances create or exacerbate challenges (as critics allege) or generate or amplify opportunities (as proponents assert) for greater respect for the law and more purportedly 'humanitarian' outcomes.[27] It is difficult to discern, for example, how and to what extent the human agents composing a party to an armed conflict in practice construct and correlate proxies for legally relevant characteristics – for example, those concerning direct participation in hostilities as a basis for targeting[28] or imperative reasons of security as a ground for detention[29] – involved in the collection of data and the operation of algorithms. Nor do parties routinely divulge what specific dependencies exist within and between the computational components that their human agents adopt regarding a particular form of warfare. Instead, by and large, parties – at most – merely reaffirm in generic terms that their human agents strictly respect the rules.

the Red Cross, June 2020) www.sipri.org/sites/default/files/2020-06/2006_limits_of_autonomy_0.pdf (hereafter Boulanin and others, 'Limits on Autonomy'); ICRC, 'Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centred Approach' (International Committee of the Red Cross, 6 June 2019) www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach; T Singer, *Dehumanisierung der Kriegführung: Herausforderungen für das Völkerrecht und die Frage nach der Notwendigkeit menschlicher Kontrolle* (2019); Advisory Council on International Affairs and Advisory Committee on Issues of Public International Law, *Autonomous Weapon Systems; the Need for Meaningful Control* (No. 97 AIV/ No. 26 CAVV, October 2015) (views adopted by Government) www.advisorycouncilinternationalaffairs.nl/documents/publications/2015/10/02/autonomous-weapon-systems; Working Paper by Austria, 'The Concept of "Meaningful Human Control"' (The Convention on Certain Conventional Weapons (CCW) Second Informal Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, 13–18 April 2015) https://perma.cc/D35A-RP7G.

[25] See, e.g., Lewis, 'An Enduring Impasse' (n 20).

[26] See generally AH Michel, 'The Black Box, Unlocked: Predictability and Understandability in Military AI' (UN Institute for Disarmament Research, 2020) https://unidir.org/publication/black-box-unlocked (hereafter Michel, 'The Black Box, Unlocked').

[27] See, e.g., Lewis, 'An Enduring Impasse' (n 20).

[28] See Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) (signed 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3 (Additional Protocol I) Art 51(3) (hereafter AP I); Protocol Additional to the Geneva Conventions of 12 August 1949 and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II) (signed 8 June 1977, entered into force 7 December 1978) 1125 UNTS 609 (Additional Protocol II) Article 13(3) (hereafter AP II).

[29] See Geneva Convention relative to the Protection of Civilian Persons in Time of War (signed 12 August 1949, entry into force 21 October 1950) 75 UNTS 287 (GC IV) Article 78, first para.

III. OVERVIEW OF INTERNATIONAL LAW APPLICABLE TO ARMED CONFLICT

International law is the only binding framework agreed to by States to regulate acts and omissions related to armed conflict. In this respect, international law is distinguishable from national legal frameworks, corporate codes of conduct, and ethics policies.

The sources, or origins, of international law applicable in relation to armed conflict include treaties, customary international law, and general principles of law. Several fields of international law may lay down binding rules applicable to a particular armed conflict. As mentioned earlier, the primary field developed by States to apply to conduct undertaken in relation to armed conflict is IHL/LOAC. Other potentially relevant fields may include the area of international law regulating the threat or use of force in international relations (also known as the *jus ad bellum* or the *jus contra bellum*), international human rights law, international criminal law, international refugee law, the law of State responsibility, and the law of responsibility of international organizations. In international law, an international organization (IO) is often defined as an organization established by a treaty or other instrument governed by international law and possessing its own international legal personality.[30] Examples of IOs include the United Nations Organization (UN) and the North Atlantic Treaty Organization (NATO), among many others.

Under contemporary IHL/LOAC, there are two generally recognized classifications, or categories, of armed conflicts.[31] One is an international armed conflict, and the other is a non-international armed conflict. The nature of the parties most often distinguishes these categories. International armed conflicts are typically considered to involve two or more States as adversaries. Non-international armed conflicts generally involve one or more States fighting together against one or more non-state parties or two or more non-state parties fighting against each other.

What amounts to a breach of IHL/LOAC depends on the content of the underlying obligation applicable to a particular human or legal entity. Depending on the specific armed conflict, potentially relevant legal entities may include one or more States, IOs, or non-state parties. IHL/LOAC structures and lays down legal provisions concerning such thematic areas as the conduct of hostilities, detention, and humanitarian services, among many others.

For example, under certain IHL/LOAC instruments, some weapons are expressly prohibited, such as poisoned weapons,[32] chemical weapons,[33] and weapons that injure by fragments that escape detection by X-rays in the human body.[34] The use of weapons that are not expressly prohibited may be tolerated under IHL/LOAC at least insofar as the use of the weapon comports with applicable provisions. For instance, depending on the specific circumstances of use and the relevant actors, those provisions may include:

[30] See Draft Articles on Responsibility of International Organizations with Commentary (Report of the Commission to the General Assembly on the Work of Its Sixty-Third Session, 2011) Ybk Intl L Comm, Volume II (Part 2) A/CN.4/SER.A/2011/Add 1 (Part 2), Article 2(a) (hereafter (D)ARIO).

[31] See ICRC, *Commentary* (n 17) paras 201–342, 384–502.

[32] Regulations Respecting the Laws and Customs of War on Land, Annex to Convention (IV) Respecting the Laws and Customs of War on Land (signed 18 October 1907, entered into force 26 January 1910) 36 Stat 2295, Article 23(a).

[33] Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction (signed 3 September 1992, entered into force 29 April 1997) 1975 UNTS 45, Article I(1).

[34] Protocol on Non-detectable Fragments (Protocol I) to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be deemed to be Excessively Injurious or to have Indiscriminate Effects (signed 10 October 1980, entered into force 2 December 1983) 1342 UNTS 147.

- the obligation for parties to distinguish between the civilian population and combatants and between civilian objects and military objectives and to direct their operations only against military objectives;[35]
- the prohibition on attacks which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated;[36]
- the obligation to take constant care to spare the civilian population, civilians, and civilian objects in military operations;[37] and
- obligations to take certain precautions concerning attacks.[38]

International law sets out particular standard assumptions of responsibility for the conduct of States and IOs. It is on the basis of those assumptions that specific IHL/LOAC provisions exist and are applied.[39] In other words, international law pertaining to armed conflict exists and is applied in respect of States and IOs based on the interrelationships between the 'primary' substantive IHL/LOAC provisions and the 'secondary' responsibility institutions. Regarding both State responsibility and IO responsibility, standard assumptions of responsibility are rooted in underlying concepts of attribution, breach, circumstances precluding wrongfulness, and consequences.[40] Those assumptions are general in character and are assumed and apply unless excluded, for example through an individual treaty or rule.[41]

A use in an armed conflict of an AI-related tool or technique may (also or separately) give rise to individual criminal responsibility under international law. Such personal criminal responsibility may arise where the conduct that forms the application of an AI-related tool or technique constitutes, or otherwise sufficiently contributes to, an international crime. For example, under the Rome Statute of the International Criminal Court (ICC), the court has jurisdiction over the crime of genocide, crimes against humanity, war crimes, and the crime of aggression.[42] A use of an AI-related tool or technique may form part or all of the conduct underlying one or more of the crimes prohibited under the ICC Statute.

Concerning imposition of individual criminal responsibility, it may be argued that standard assumptions of responsibility are based (at least under the ICC Statute) on certain underlying concepts.[43] Those concepts may arguably include jurisdiction;[44] ascription (that is, attribution of conduct to a natural person);[45] material elements (in the sense of the prohibited conduct forming the crime);[46] mental elements (including the requisite intent and knowledge);[47] modes

[35] AP I (n 28) Article 48.
[36] Ibid Article 51(5)(*b*).
[37] Ibid Article 57(1).
[38] Ibid Article 57(2).
[39] See JR Crawford, 'State Responsibility' in R Wolfrum (ed), *Max Planck Encyclopedia of Public International Law* (2006) (hereafter Crawford, 'State Responsibility').
[40] Ibid; Draft Articles on Responsibility of States for Internationally Wrongful Acts, with Commentary (Report of the Commission to the General Assembly on the Work of its Fifty-Third Session, 2001) Ybk Intl L Comm, Volume II (Part Two) A/CN.4/SER.A/2001/Add 1 (Part 2) (hereafter (D)ARSIWA); (D)ARIO (n 30).
[41] Crawford, 'State Responsibility' (n 39).
[42] Rome Statute of the International Criminal Court (signed 17 July 1998, entered into force 1 July 2002) 2187 UNTS 3 (ICC Statute), Articles 5, 10–19.
[43] See DA Lewis, 'International Legal Regulation of the Employment of Artificial-Intelligence-Related Technologies in Armed Conflict' (2020) 2 *Moscow JIL* 53, 61–63.
[44] See ICC Statute, Articles 5–19.
[45] See ICC Statute, Articles 25–26.
[46] See ICC Statute, Articles 6–8 *bis*.
[47] See ICC Statute, Article 30.

of responsibility (such as aiding and abetting or command responsibility);[48] grounds for excluding responsibility;[49] trial;[50] penalties (including imprisonment of the responsible person);[51] and appeal and revision.[52] It may be argued that it is on the basis of the assumptions related to those concepts that the provisions of the ICC Statute exist and are applied.

IV. PRECONDITIONS ARGUABLY NECESSARY TO RESPECT INTERNATIONAL LAW

In this section, I outline some preconditions underlying elements that are arguably necessary for international law to be respected in relation to a use in an armed conflict of an AI-related tool or technique. I assume that the employment of the technology is governed (at least in part) by international law. By respecting international law, I mean the bringing of a binding norm, principle, rule, or standard to bear in relation to a particular employment of an AI-related tool or technique in a manner that accords with the object and purpose of the relevant provision, that facilitates observance of the provision, and that facilitates incurrence of responsibility in case of breach of the provision.

At least three categories of actors may be involved in respecting international law in relation to a use in an armed conflict of an AI-related tool or technique. Each category is arguably made up, first and foremost, of human agents. In addition to those human agents, the entities to which those humans are attached or through which they otherwise (seek to) implement international law may also be relevant.

The first category is made up in part of the humans who are *involved* in relevant acts or omissions (or both) that form the employment of an AI-related tool or technique attributable to a State or an IO. This first category of actors also includes the entity or entities – such as the State or the IO or some combination of State(s) and IO(s) – to which the employment is attributable. The human agents may include, for example, software engineers, operators, commanders, and legal advisers engaging in conduct on behalf of the State or the IO.

The second category of actors is made up in part of humans *not involved* in the employment in an armed conflict of an AI-related tool or technique attributable to a State or an IO but who may nevertheless (seek to) ensure respect for international law in relation to that conduct. This second category of actors also includes entities – such as (other) States, (other) IOs, international courts, and the like – that may attempt, functionally through the humans who compose them, to ensure respect for international law in relation to the conduct.

The third category of actors is made up in part of humans who (seek to) apply international law – especially international law on international crimes – to relevant conduct of a natural person. These humans may include, for example, prosecutors, defense counsel, and judges. This third category of actors also includes entities (mostly, but not exclusively, international or domestic criminal tribunals) that may seek, functionally through the humans who compose them, to apply international law to natural persons.

In the rest of this section, I seek to elaborate some preconditions regarding each of these three respective categories of actors.

[48] See ICC Statute, Articles 25, 28.
[49] See ICC Statute, Articles 31–33.
[50] See ICC Statute, Articles 62–76.
[51] See ICC Statute, Article 77.
[52] ICC Statute, Articles 81–84.

1. *Preconditions Concerning Respect for International Law by Human Agents Acting on Behalf of a State or an International Organization*

In this sub-section, I focus on employments in armed conflicts of AI-related tools or techniques attributable to one or more States, IOs, or some combination thereof. In particular, I seek to outline some preconditions underlying elements that are arguably necessary for the State or the IO to respect international law in relation to such an employment.

Precondition #1: Humans Are Legal Agents of States and International Organizations
The first precondition is that humans are arguably the agents for the exercise and implementation of international law applicable to States and IOs. This precondition is premised on the notion that existing international law presupposes that the functional exercise and implementation of international law by a State or an IO in relation to the conduct of that State or that IO is reserved solely to humans.[53] According to this approach, this primary exercise and implementation of international law may not be partly or wholly reposed in non-human (artificial) agents.[54]

Precondition #2: Human Agents of the State or the International Organization Sufficiently Understand the Performance and Effects of the Employment
The second precondition is that human agents of the State or the IO that engages in conduct that forms an employment in an armed conflict of an AI-related tool or technique arguably need to sufficiently understand the technical performance and effects of the employed tool or technique in respect of the specific circumstances of the employment and in relation to the socio-technical system through which the tool or technique is employed.[55] For this precondition to be instantiated, the understanding arguably needs to encompass (among other things) comprehension of the dependencies underlying the socio-technical system, the specific circumstances and conditions of the employment, and the interactions between those dependencies, circumstances, and conditions.

[53] See Informal Working Paper by Switzerland (30 March 2016), 'Towards a "Compliance-Based" Approach to LAWS [Lethal Autonomous Weapons Systems]' (Informal Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, 11–15 April 2016) https://perma.cc/WRJ6-CCMS (expressing the position that '[t]he Geneva Conventions of 1949 and the Additional Protocols of 1977 were undoubtedly conceived with States and individual humans as agents for the exercise and implementation of the resulting rights and obligations in mind.') (hereafter Switzerland, 'Towards a "Compliance-Based" Approach'); see also Office of the General Counsel of the Department of Defense (US), *Department of Defense Law of War Manual* [June 2015, updated Dec. 2016], s 6.5.9.3, p 354 (expressing the position that law-of-war obligations apply to persons rather than to weapons, including that 'it is persons who must comply with the law of war') (hereafter US DoD OGC, *Law of War Manual*).

[54] For an argument that algorithmic forms of warfare – which may apparently include certain employments of AI-related tools or techniques – cannot be subject to law writ large, see G Noll, 'War by Algorithm: The End of Law?', in M Liljefors, G Noll, and D Steuer (eds), *War and Algorithm* (2019).

[55] See generally L Suchman, 'Configuration' in C Lury and N Wakeford (eds), *Inventive Methods* (2012). For an analysis of the 'technical layer', the 'socio-technical layer,' and the 'governance layer' pertaining to autonomous weapons systems, see I Verdiesen, F Santoni de Sio, and V Dignum, 'Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight'(2020) *Minds and Machines* https://doi.org/10.1007/s11023-020-09532-9. For an analysis of US 'drone operations' (albeit admittedly not pertaining to AI as such) informed in part by methods relevant to socio-technical configurations, see MC Elish, 'Remote Split: A History of US Drone Operations and the Distributed Labor of War' (2017) 42(6) *Science, Technology, & Human Values* 1100. On certain issues related to predicting and understanding military applications of artificial intelligence, see Michel, 'The Black Box, Unlocked' (n 26). With respect to machine-learning algorithms more broadly, see J Burrell, 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' (January–June 2016) *Big Data & Society* 1-12. For recent arguments concerning limits on autonomy in weapons systems in particular, see Boulanin and others, 'Limits on Autonomy' (n 24).

Precondition #3: Human Agents of the State or the International Organization Discern the Law Applicable to the Employment
The third precondition is that human agents of the State or the IO that engages in conduct that forms an employment in an armed conflict of an AI-related tool or technique arguably need to discern the law applicable to the State or the IO in relation to the employment. The applicable law may vary based on (among other things) the specific legal provisions applicable to the State or the IO through different sources, or origins, of international law. (As noted above, those sources may include treaty law, customary international law, and general principles of international law, among others.)

Precondition #4: Human Agents of the State or the International Organization Assess the Legality of the Anticipated Employment Before the Employment
The fourth precondition is that human agents of the State or the IO that engages in conduct that forms an employment in an armed conflict of an AI-related tool or technique assess – before the employment is initiated – whether the anticipated employment would conform with applicable law in relation to the anticipated specific circumstances and conditions of the employment.[56] In line with this precondition, only those employments that pass this legality assessment may be initiated and only then under the circumstances and subject to the conditions necessary to pass this legality assessment.

Precondition #5: Human Agents of the State or the International Organization Impose Legally Mandated Parameters Before and During the Employment
The fifth precondition is that human agents of the State or the IO that engages in conduct that forms an employment in an armed conflict of an AI-related tool or technique need to impose – before and during the employment – limitations or prohibitions or both as required by applicable law in respect of the employment. To instantiate this precondition, human agents of the State or the IO need to discern and configure the particular limitations or prohibitions by interpreting and applying international law in respect of the employment. Factors that the human agents might need to consider could include (among many others) interactions between the socio-technical system's dependencies and the specific circumstances and conditions of the employment.[57]

Suppose those dependencies, circumstances, or conditions (or some combination thereof) materially change after the employment is initiated. In that case, the human agents of the State or the IO arguably need to discern and configure the limitations or prohibitions (or both) in light of those changes.

To the extent, if any, required by the law applicable in relation to a specific employment or generally, human agents of the State or the IO may need to facilitate at least partial interaction by one or more humans with the system during the employment. Such interactions may take such forms (among others) as monitoring, suspension, or cancellation of some or all of the employment.[58]

[56] See N Goussac, 'Safety Net or Tangled Web: Legal Reviews of AI in Weapons and War-Fighting' (*ICRC Humanitarian Law and Policy Blog*, 18 April 2019) https://blogs.icrc.org/law-and-policy/2019/04/18/safety-net-tangled-web-legal-reviews-ai-weapons-war-fighting/; Lewis, 'Legal Reviews' (n 2).

[57] For broader critiques and concerns – including some informed by socio-technical perspectives – related to (over-) reliance on algorithmic systems, see, among others, R Benjamin, *Race after Technology* (2019); SU Noble, *Algorithms of Oppression* (2018); BD Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (July–Dec. 2016) *Big Data & Society* 1-21; C O'Neil, *Weapons of Math Destruction* (2016).

[58] See, e.g., with respect to precautions in attacks in situations of armed conflict, AP I (n 28) Article 57(2)(b).

Precondition #6: Human Agents of the State or the International Organization Assess (Il)Legality after the Employment

The sixth precondition is that human agents of the State or the IO that engages in conduct that forms an employment in an armed conflict of an AI-related tool or technique arguably need to assess, after employment, whether or not the employment complied with applicable law. To instantiate this precondition, those human agents need to discern (among other things) which humans engaged in which elements of relevant conduct, the circumstances and conditions pertaining to that conduct, and whether the anticipated and actual performance and effects of the socio-technical system underlying the employment conformed with the legally mandated parameters.

Precondition #7: Human Agents of the State or the International Organization Assess Potential Responsibility for Violations Arising in Connection with the Employment

The seventh precondition concerns suspected violations that may arise in relation to an employment in an armed conflict of an AI-related tool or technique by or on behalf of a State or an IO. The precondition is that human agents of the State or the IO that undertook the conduct assess whether or not the conduct constitutes a violation – and, if they assess a violation occurred, human agents of the State or the IO (also) evaluate whether the international legal responsibility of the State or the IO is engaged. To make the assessment required by this precondition, human agents of the State or the IO need to discern, first, whether or not the conduct that forms the employment is attributable to the State or the IO (or to some combination of one or more State(s) or IO(s) or both).[59] If attribution is established, human agents of the State or the IO need to discern whether a breach occurred. This exercise entails assessing the conduct against applicable law. Finally, if the occurrence of a breach is established, human agents of the State or the IO evaluate whether or not the circumstances preclude the wrongfulness of the breach.[60]

Precondition #8: Human Agents of the State or the International Organization Facilitate Incurrence of Responsibility

The eighth precondition concerns situations in which a breach – the wrongfulness of which is not precluded by the circumstances – is established. The precondition is that, where such a breach is established, human agents of the State or the IO arguably need to facilitate incurrence of responsibility of the State or the IO concerning the breach. As part of the process to facilitate such incurrence of responsibility, human agents of the State or the IO may arguably need to impose relevant consequences on the State or the IO. Those consequences may relate, for example, to cessation or reparation (or both) by the State or the IO.[61]

Summary

Suppose that the various premises underlying the above-elaborated preconditions are valid. In that case, the absence of one or more of the following conditions may be preclusive of an element integral to respect for international law by the State or the IO:

[59] For an exploration of certain legal aspects of attribution in relation to 'cyber operations' (which may or may not involve AI-related tools or techniques), see HG Dederer and T Singer, 'Adverse Cyber Operations: Causality, Attribution, Evidence, and Due Diligence' (2019) 95 *ILS* 430, 435–466.

[60] See (D)ARSIWA (n 40) ch V; (D)ARIO (n 30) ch V.

[61] See (D)ARSIWA (n 40), Articles 30–31; (D)ARIO (n 30), Articles 30–31.

https://doi.org/10.1017/9781009207898.037 Published online by Cambridge University Press

1. An exercise and implementation of international law by human agents of the State or the IO in relation to the conduct that forms an employment in an armed conflict of an AI-related tool or technique;
2. A sufficient understanding by human agents of the State or the IO of the technical performance and effects of the employed AI-related tool or technique in relation to the circumstances of use and the socio-technical system through which the tools or techniques are employed;
3. Discernment by human agents of the State or the IO of the law applicable to the State or the IO in relation to the employment;
4. An assessment by human agents of the State or the IO whether the anticipated employment would conform with applicable law in relation to the anticipated specific circumstances and conditions of the employment;
5. Imposition by human agents of the State or the IO of limitations or prohibitions (or both) as required by applicable law in respect of the employment;
6. An assessment by human agents of the State or the IO after employment as to whether or not the employment complied with applicable law;
7. An assessment by human agents of the State or the IO as to whether or not the conduct constitutes a violation, and, if so, (also) an evaluation by human agents of the State or the IO as to whether or not the international legal responsibility of the State or the IO is engaged; or
8. Facilitation by human agents of the State or the IO of the incurrence of responsibility – including imposition of relevant consequences on the State or the IO – where such responsibility is established.

2. *Preconditions Concerning Non-Involved Humans and Entities Related to Respect for International Law by a State or an International Organization*

In this sub-section, I seek to outline some preconditions underlying elements that are arguably necessary for non-involved humans and related entities to (help) ensure respect for international law by a State or an international organization whose conduct forms an employment in an armed conflict of an AI-related tool or technique. Such non-involved people might include, for example, legal advisers from another State or another IO or judges on an international court seized with proceedings instituted by one State against another State.

Precondition #1: Humans Are Legal Agents

As with the previous sub-section, the first precondition here is that humans are arguably the agents for the exercise and implementation of international law applicable to the State or the IO whose conduct forms an employment of an AI-related tool or technique.[62] This precondition is premised on the notion that existing international law presupposes that the functional exercise and implementation of international law to a State or an IO by a human (and by an entity to which that human is connected) not involved in relevant conduct is reserved solely to humans. According to this approach, that primary exercise and implementation of international law may not be partly or wholly reposed in non-human (artificial) agents.

[62] See Switzerland, 'Towards a "Compliance-Based" Approach', above (n 53); US DoD OGC, *Law of War Manual*, above (n 53).

https://doi.org/10.1017/9781009207898.037 Published online by Cambridge University Press

Precondition #2: Humans Discern the Existence of Conduct that Forms an Employment of an AI-Related Tool or Technique

The second precondition is that humans not involved in the conduct of the State or the IO arguably need to discern the existence of the conduct that forms an employment in an armed conflict of an AI-related tool or technique attributable to the State or the IO. To instantiate this precondition, the conduct must be susceptible to being discerned by (non-involved) humans.

Precondition #3: Humans Attribute Relevant Conduct of One or More States or International Organizations to the Relevant Entity or Entities

The third precondition is that humans not involved in the conduct of the State or the IO arguably need to attribute the conduct that forms an employment in an armed conflict of an AI-related tool or technique by or on behalf of the State or the IO to that State or that IO (or to some combination of State(s) or IO(s) or both). To instantiate this precondition, the conduct undertaken by or on behalf of the State or the IO must be susceptible to being attributed by (non-involved) humans to the State or the IO.

Precondition #4: Humans Discern the Law Applicable to Relevant Conduct

The fourth precondition is that humans not involved in the conduct of the State or the IO arguably need to discern the law applicable to the conduct that forms an employment in an armed conflict of an AI-related tool or technique attributable to the State or the IO. To instantiate this precondition, the legal provisions applicable to the State or the IO to which the relevant conduct is attributable must be susceptible to being discerned by (non-involved) humans. For example, where an employment of an AI-related tool or technique by a State occurs in connection with an armed conflict to which the State is a party, humans not involved in the conduct may need to discern whether the State has become party to a particular treaty and, if not, whether a possibly relevant rule reflected in that treaty is otherwise binding on the State, for example through customary international law.

Precondition #5: Humans Assess Potential Violations

The fifth precondition is that humans not involved in the conduct that forms an employment in an armed conflict of an AI-related tool or technique attributable to the State or the IO arguably need to assess possible violations by the State or the IO concerning that conduct.

To make that assessment, (non-involved) humans need to discern, first, whether or not the relevant conduct is attributable to the State or the IO. To instantiate this aspect of the fifth precondition, the conduct forming the employment in an armed conflict of an AI-related tool or technique must be susceptible to being attributed by (non-involved) humans to the State or the IO.

If attribution to the State or the IO is established, (non-involved) humans need to discern the existence or not of the occurrence of a breach. To instantiate this aspect of the fifth precondition, the conduct forming the employment in an armed conflict of an AI-related tool or technique by the State or the IO must be susceptible to being evaluated by (non-involved) humans as to whether or not the conduct constitutes a breach.

If the existence of a breach is established, (non-involved) humans need to assess whether or not the circumstances preclude the wrongfulness of the violation. To instantiate this aspect of the fifth precondition, the conduct forming the employment in an armed conflict of an AI-related tool or technique must be susceptible to being evaluated by (non-involved) humans as to whether or not the specific circumstances preclude the wrongfulness of the breach.

Precondition #6: Humans (and an Entity or Entities) Facilitate Incurrence of Responsibility
The sixth precondition is that humans (and an entity or entities) not involved in the conduct that forms an employment in an armed conflict of an AI-related tool or technique attributable to the State or the IO arguably need to facilitate incurrence of responsibility for a breach the wrongfulness of which is not precluded by the circumstances. In practice, responsibility may be incurred through relatively more formal channels (such as through the institution of State-vs.-State legal proceedings) or less formal modalities (such as through non-public communications between States).

As part of the process to facilitate incurrence of responsibility, (non-involved) humans arguably need to impose relevant consequences on the responsible State or IO. Typically, those humans do so by acting through a legal entity to which they are attached or through which they otherwise (seek to) ensure respect for international law – for example, consider legal advisers of another State, another IO, or judge on an international court. The consequences may relate to (among other things) cessation and reparations.

Regarding cessation, the responsible State or IO is obliged to cease the act, if it is continuing, and to offer appropriate assurances and guarantees of non-repetition, if circumstances so require.[63] To instantiate this aspect of the sixth precondition, the conduct forming the employment in an armed conflict of an AI-related tool or technique must be susceptible to being evaluated by (non-involved) humans as to whether or not the conduct is continuing; furthermore, the conduct must (also) be susceptible to being subject to an offer of appropriate assurances and guarantees of non-repetition, if circumstances so require.

Regarding reparation, the responsible State or IO is obliged to make full reparation for the injury caused by the internationally wrongful act.[64] To instantiate this aspect of the sixth precondition, the conduct forming the employment in an armed conflict of an AI-related tool or technique must be susceptible both to a determination by (non-involved) humans of the injury caused and to the making of full reparations in respect of the injury.

Summary
Suppose that the various premises underlying the above-elaborated preconditions are valid. In that case, the absence of one or more of the following conditions may be preclusive of an element integral to (non-involved) humans and entities helping to ensure respect for international law by a State or an IO where the latter's conduct forms an employment in an armed conflict of an AI-related tool or technique:

1. An exercise and implementation by (non-involved) humans of international law applicable to the State or IO in relation to the conduct;
2. Discernment by (non-involved) humans of the existence of the relevant conduct attributable to the State or the IO;
3. An attribution by (non-involved) humans of the relevant conduct undertaken by or on behalf of the State or the IO;
4. Discernment by (non-involved) humans of the law applicable to the relevant conduct attributable to the State or the IO;
5. An assessment by (non-involved) humans of possible violations committed by the State or the IO in connection with the relevant conduct; or

[63] See (D)ARSIWA (n 40) Article 30; (D)ARIO (n 30) Article 30.
[64] See (D)ARSIWA (n 40) Article 31; (D)ARIO (n 30) Article 31.

6. Facilitation by (non-involved) humans of an incurrence of responsibility of the responsible State or the responsible IO for a breach the wrongfulness of which is not precluded by the circumstances.

3. *Preconditions Concerning Respect for the ICC Statute*

In the above sub-sections, I focused on respect for international law concerning employments in armed conflicts of AI-related tools and techniques by or on behalf of a State or an IO, whether the issue concerns respect for international law by those involved in the conduct (IV 1) or whether it concerns those not involved in the conduct (IV 2). In this sub-section, I seek to outline some preconditions underlying elements that are arguably necessary for respect for the ICC Statute. As noted previously, under the ICC Statute, individual criminal responsibility may arise for certain international crimes, and an employment in an armed conflict of an AI-related tool or technique may constitute, or otherwise contribute to, such a crime. In this section, I use the phrase 'ICC-related human agents' to mean humans who exercise and implement international law in relation to an application of the ICC Statute. Such human agents may include (among others) the court's prosecutors, defense counsel, registrar, and judges.

Precondition #1: Humans Are Legal Agents

The first precondition is that humans are arguably the agents for the exercise and implementation of international law applicable in relation to international crimes – including under the ICC Statute – arising from conduct that forms an employment in an armed conflict of an AI-related tool or technique.[65] (Of the four categories of crimes under the ICC Statute, strictly speaking only war crimes by definition must necessarily be committed in connection with an armed conflict. Nonetheless, the other three categories of crimes under the ICC Statute may be committed in connection with an armed conflict.) This precondition is premised on the notion that existing international law presupposes that the functional exercise and implementation of international law to the conduct of a natural person is reserved solely to humans (and, through them, to the entity or entities, such as an international criminal tribunal, to which those humans are attached). According to this approach, this primary exercise and implementation of international law may not be partly or wholly reposed in non-human (artificial) agents.

Precondition #2: Humans Discern the Existence of Potentially Relevant Conduct

The second precondition is that ICC-related human agents arguably need to discern the existence of conduct that forms an employment in an armed conflict of an AI-related tool or technique ascribable to a natural person. For this precondition to be instantiated, such conduct must be susceptible to being discerned by relevant ICC-related human agents.

Precondition #3: Humans Determine Whether the ICC May Exercise Jurisdiction

The third precondition is that ICC-related human agents arguably need to determine whether or not the court may exercise jurisdiction in relation to an employment in an armed conflict of an AI-related tool or technique ascribable to a natural person. The court may exercise jurisdiction only over natural persons.[66] Furthermore, the ICC may exercise jurisdiction only where the

[65] See Switzerland, 'Towards a "Compliance-Based" Approach', above (n 53); US DoD OGC, *Law of War Manual*, above (n 53).
[66] ICC Statute, Article 25(1).

relevant elements of jurisdiction are satisfied.[67] To instantiate the third precondition, conduct that forms an employment in an armed conflict of an AI-related tool or technique ascribable to a natural person must be susceptible to being evaluated by relevant ICC-related human agents as to whether or not the conduct is attributable to one or more natural persons over whom the court may exercise jurisdiction.

Precondition #4: Humans Adjudicate Individual Criminal Responsibility

The fourth precondition is that ICC-related human agents arguably need to adjudicate whether or not an employment in an armed conflict of an AI-related tool or technique ascribable to a natural person subject to the jurisdiction of the court constitutes, or otherwise contributes to, an international crime over which the court has jurisdiction. For the fourth precondition to be instantiated, such conduct must be susceptible to being evaluated by relevant ICC-related human agents – in pre-trial proceedings, trial proceedings, and appeals-and-revision proceedings – as to whether or not (among other things) the conduct satisfies the 'material'[68] and 'mental'[69] elements of one or more crimes and whether the conduct was undertaken through a recognized mode of responsibility.[70]

Precondition #5: Humans Facilitate the Incurrence of Individual Criminal Responsibility

The fifth precondition is that ICC-related human agents arguably need to facilitate incurrence of individual criminal responsibility for an international crime where such responsibility is established. As part of the process to facilitate the incurrence of such responsibility, relevant ICC-related humans need to (among other things) facilitate the imposition of penalties on the responsible natural person(s).[71] For the fifth precondition to be instantiated, the conduct underlying the establishment of individual criminal responsibility needs to be susceptible to being subject to the imposition of penalties on the responsible natural person(s).

Summary

Suppose that the various premises underlying the above-elaborated preconditions are valid. In that case, the absence of one or more of the following conditions – in relation to an employment in an armed conflict of an AI-related tool or technique that constitutes, or otherwise contributes to, an international crime – may be preclusive of respect for the ICC Statute:

1. An exercise and implementation of international law by one or more relevant ICC-related human agents concerning the conduct;
2. Discernment by one or more relevant ICC-related human agents of the conduct that forms an employment in an armed conflict of an AI-related tool or technique ascribable to a natural person;
3. A determination by one or more relevant ICC-related human agents whether or not the court may exercise jurisdiction in respect of an employment in an armed conflict of an AI-related tool or technique ascribable to a natural person;
4. An adjudication by relevant ICC-related human agents whether or not an employment in an armed conflict of an AI-related tool or technique ascribable to a natural person subject

[67] See ICC Statute, Articles 5–19.
[68] See ICC Statute, Articles 6–8 *bis*.
[69] See ICC Statute, Article 30.
[70] See ICC Statute, Articles 25, 28.
[71] ICC Statute, Article 77.

https://doi.org/10.1017/9781009207898.037 Published online by Cambridge University Press

to the jurisdiction of the court constitutes, or otherwise contributes to, an international crime over which the court has jurisdiction; or

5. Facilitation by one or more relevant ICC-related human agents of an incurrence of individual criminal responsibility – including the imposition of applicable penalties on the responsible natural person(s) – where such responsibility is established.

V. CONCLUSION

An employment in an armed conflict of an AI-related tool or technique that is attributable to a State, an IO, or a natural person (or some combination thereof) is governed at least in part by international law. It is well established that international law sets out standard assumptions of responsibility for the conduct of States and IOs. It is also well established that it is on the basis of those assumptions that specific legal provisions exist and are applied in respect of those entities. International law also arguably sets out particular standard assumptions of criminal responsibility for the conduct of natural persons. It may be contended that it is on the basis of those assumptions that the ICC Statute exists and is applied.

Concerning the use of AI in armed conflicts, at least three categories of human agents may be involved in seeking to ensure that States, IOs, or natural persons respect applicable law. Those categories are the human agents acting on behalf of the State or the IO engaging in relevant conduct; human agents not involved in such conduct but who nevertheless (seek to) ensure respect for international law in relation to that conduct; and human agents who (seek to) ensure respect for the ICC Statute. Each of those human agents may seek to respect or ensure respect for international law in connection with a legal entity to which they are attached or through which they otherwise act.

'Responsible AI' is not a term of art in international law, at least not yet. It may be argued the preconditions arguably necessary to respect international law – principally in the sense of applying and observing international law and facilitating incurrence of responsibility for violations – ought to be taken into account in formulating notions of 'responsible AI' pertaining to relevant conduct connected with armed conflict. Regarding those preconditions, it may be argued that, under existing law, humans are the (at least primary) legal agents for the exercise and implementation of international law applicable to an armed conflict. It may also be submitted that, under existing law, an employment in an armed conflict of an AI-related tool or technique needs to be susceptible to being (among other things) administered, discerned, attributed, understood, and assessed by one or more human agent(s).[72]

Whether – and, if so, the extent to which – international actors will commit in practice to instantiating the preconditions arguably necessary for respecting international law pertaining to an employment in an armed conflict of an AI-related tool or technique will depend on factors that I have not expressly addressed in this chapter but that warrant extensive consideration.

[72] See DA Lewis, 'Three Pathways to Secure Greater Respect for International Law Concerning War Algorithms' (Harvard Law School Program on International Law and Armed Conflict, December 2020) https://dash.harvard.edu/bitstream/handle/1/37367712/Three-Pathways-to-Secure-Greater-Respect.pdf?sequence=1&isAllowed=y; V Boulanin, L Bruun, and N Goussac, 'Autonomous Weapon Systems and International Humanitarian Law: Identifying Limits and the Required Type and Degree of Human–Machine Interaction' (Stockholm International Peace Research Institute, 2021) www.sipri.org/sites/default/files/2021-06/2106_aws_and_ihl_0.pdf.

https://doi.org/10.1017/9781009207898.037 Published online by Cambridge University Press