

Rethinking Norms in Times of Algorithmic Decision Making? – A Kantian Approach

Elisabeth Hoffberger-Pippan

German Institute for International and Security Affairs
(SWP), Berlin

Abstract: The use of AI-enabled weapon systems in the military context poses a myriad of legal and ethical challenges. Since technological developments are progressing fast, humanity should consider whether the increased use of algorithmic decision-making will ultimately shatter the core tenets of international law. The discussion on autonomy in weapon systems and its implications for international law is primarily taking place in the UN-Governmental Group of Experts on emerging technologies in the area of (lethal) autonomous weapon systems centering upon the requirement to maintain human control. What can be observed is that various states and also researchers refer to deontological rather than utilitarian ethics to justify the requirement of maintaining human control. This discussion is largely influenced by the findings and deliberations of Immanuel Kant, especially with regard to the concept of human dignity. His legacy not only impacts the debate in Geneva but goes beyond ultimately contributing to the ongoing process of creating “digital constitutionalization” of the normative order dealing with AI-enabled systems. In the end, AI-enabled weapon systems are not able to shatter the core tenets of international law. Rather, human agency will be preserved allowing humans to stay in control in both structural and substantive terms.

Algorithms, autonomous weapon systems, human control, Immanuel Kant

1. Introduction

Technological advancement in the fields of robotics and artificial intelligence¹ (AI) permeate nearly every aspect of our lives. From self-driving cars to medical robots, the list of highly autonomous capabilities seems never-ending. There is an increased probability that the advent of technology resembling human intelligence will trigger a debate about whether the core structures and foundational principles of international law have to be re-evaluated in order to keep pace with technological innovation. Such a debate is particularly relevant in armed conflict, especially in the context of autonomous weapon systems (AWS). First, an increased level of AI within AWS could lead to the assumption that AWS have an own international legal personality distinct from states.² In light of this, states would be induced to consider them direct addressees of international humanitarian law (IHL). Second, conferring international legal personality to AI-enabled weapon systems inevitably raises the question of whether such weapon systems must have their own liability regimes.³ Third, technological innovation could make AWS more attractive and their use could be legitimized by state practice gradually replacing humans from the battlefield with presumably devastating consequences for the civilian population.

Despite great power rivalry and a bitter arms race between key military players in the area of AI, the vast majority of states parties in the UN-Group of Governmental Experts (GGE),⁴ the main forum where the various challenges regarding AWS are currently discussed, seems to agree that a sufficient level of human involvement over AWS has to be maintained in order to avoid *inter alia* an anthropomorphization of AI-enabled systems.⁵ Among

¹ For a definition see Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Hoboken: Prentice Hall 2010) 1–5.

² See, for example, Valentina Petroca Talimonchik, “The Prospects for the Recognition of International Legal Personality of Artificial Intelligence”, *LAWS*, 10, no. 85 (2021): 1–11. With critical remarks regarding the idea of granting international legal personality to AI-enabled systems in more general terms see Simon Chesterman, “Artificial Intelligence and the Limits of Legal Personality”, *International and Comparative Law Quarterly*, 69, no. 4 (2020): 822–844.

³ See, for example, Gabriel Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems* (Cham – Heidelberg – New York – Dordrecht – London 2015) 3 – 15.

⁴ For more information on the GGE (LAWS) see United Nations, Office for Disarmament Affairs (UNODA), “Background on LAWS in the CCW”, <https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/> (last accessed 30 March 2022).

⁵ Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Meeting of the High Contracting Parties 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

these states are those explicitly mentioning ethical concerns to substantiate their legal arguments stressing that the use of AWS without sufficient human involvement would run counter to core ethical principles, such as human dignity.⁶ But there are also a number of states parties explicitly denying the relevance of ethical concerns and rebutting the argument that ethics and law are inextricably intertwined. These states are generally more open to technical innovation usually calling for a comparatively low level of human involvement.⁷ The arguments put forward by the second group could trigger a broader discussion in public, at the national level, but also in academia, whether humans will remain the core addressees of the law and the only duty bearers of legal obligations, respectively. But any claims arguing that AWS are able to shatter the core tenets of international law (by e.g. granting AI-enabled AWS international legal personality) would not be based on solid grounds. The legacy of *Immanuel Kant* and his approach to human dignity is imperative in order to understand the leading role humans will play in military operations in future.⁸ But *Immanuel Kant's* work not only impacts the debate on AWS and IHL. It will be demonstrated that his legacy extends to the broader debate on the governance of AI and AWS, respectively.

This article sets out to explore how ethical considerations, especially the debate on human dignity, feed into the legal discussion on AWS and IHL. The first chapter after the introduction focusses on ethical concerns in the debate on AWS in the context of IHL, especially the inextricable interlinkage

Indiscriminate Effects, Final Report (13 December 2019) CCW/MSP/2019/9, Annex III.

⁶ See, for example, the national commentary submitted by Austria on the Guiding Principles of 2019, <https://documents.unoda.org/wp-content/uploads/2020/09/20200901-Austria.pdf> (last accessed 30 March 2022).

⁷ See, for example, the digital recordings of the UN GGE on AWS, <https://conf.unog.ch/digitalrecordings/index.html?embed=-h&mrid=5EBAAABD-DD16-458E-A2C9-FEA476FCE99B> (last accessed 30 March 2022).

⁸ Daniele Amoroso, *Autonomous Weapons Systems and International Law* (Baden-Baden: Nomos 2020) 181–89; Aaron Johnson and Sidney Axinn, “The Morality of Autonomous Robots”, *Journal of Military Ethics* 12 (2013): 129–141; Amanda Sharkey, “Autonomous weapons systems, killer robots and human dignity”, *Ethics and Information Technology* (2019): 75–87; Nayef Al-Rhodan, “A Neurophilosophy of Autonomous Weapons and Warfare”, *Blog of the American Philosophical Association*, August 10, 2020, <https://blog.apaonline.org/2020/08/10/a-neurophilosophy-of-autonomous-weapons-and-warfare/> (last accessed 30 March 2022). With critical remarks see Dieter Birnbacher, “Are autonomous weapons systems a threat to human dignity?” in *Autonomous Weapons Systems: Law, Ethics, Policy*, (eds.) Nehal Bhuta et al. (Cambridge: Cambridge University Press 2016) 105–111. See also Duncan MacIntosh, “Autonomous Weapons and the Nature of Law and Morality: How Rule-of-Law-Values Require Automation of the Rule of Law”, *Temple International and Comparative Law Journal* 30, no. 1, (2016): 114.; Ozlem Ulgen, “Human Dignity in an Age of Autonomous Weapons: Are We in Danger of Losing an ‘Elementary Consideration of Humanity’?”, *Conference Paper No. 15/2016*, 2016 ESIL Annual Conference, Riga, 8–10 September 2016.

between ethics and IHL, the role of the Martens Clause and also the role of human rights in armed conflict. Furthermore, *Kant's* deliberations on and approach to human dignity will be discussed with a view to ascertaining how the concept of human dignity impacts the AWS debate. It will be demonstrated that various narratives regarding AWS have emerged looking at the issue from different angles and perspectives. But ultimately, the importance to maintain a sufficient level of human involvement over AWS gains common ground and general acceptance without denying potential benefits of AI. Chapter two then proceeds with an analysis of *Kant's* legacy beyond IHL by analyzing his influence on the broader discussion on the use of AI including civilian contexts. A larger policy consensus seems to emerge calling for the maintenance of sufficient human involvement which will shape and restrain future law-making at both the national and the international level. In chapter three it will be demonstrated that in light of technological innovation and an increased militarization of the world order, *Immanuel Kant's* legacy and his human-centric approach is now more relevant than ever giving humanity guidance on how to address the various challenges posed by AI and how to remain in control both in procedural and structural terms. Chapter four will then address the question of how the concept of human agency translates into concrete action, that is to say, how humans should and must be involved in the process of developing, deploying and using AWS in order to guarantee compliance with IHL and ethical standards.

2. Autonomous Weapon Systems and International Humanitarian Law: Kant's approach to human dignity

The UN-GGE, embedded in the larger framework of the Conventional Weapons Convention (CCW),⁹ is the primary forum for a discussion on AWS. Even though the CCW qualifies as an arms control treaty, it is considered an integral part of IHL and numerous provisions in the CCW thus relate to IHL obligations.¹⁰

2.1. Autonomous Weapon Systems and Human Agency

There is no internationally agreed upon definition of AWS. In 2012, the US Department of Defense (DoD) made a first attempt at defining AWS by

⁹ 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), Geneva, 10 October 1980, in force 2 December 1983, UNTS Vol. 1342 No. 22495.

¹⁰ See, for example, the preamble of the CCW.

issuing Directive 3000.09 describing them as weapon systems able to independently identify and engage targets without further human input.¹¹ By the same token, the International Committee of the Red Cross (ICRC) provided a working definition in a report of 2016 that is widely used and largely accepted today. According to the ICRC, an AWS is any weapon system “that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention.”¹² Weapon systems arguably qualifying as AWS are *inter alia* air defence systems, active denial systems, sentry weapons, homing munition and loitering munition.¹³ According to a UN-Report of 2021, AWS were allegedly used by the Turkish military against retreating forces loyal to Chalifa Belqasim Haftar in Libya, a country deeply entrenched in armed conflict since the defeat of Muammar al-Gaddafi in 2011. Not all types of AWS are enabled by AI, however. Air defence systems, for example, identify and engage targets such as incoming missiles based on specific, pre-defined radar signatures. But recent technological developments suggest a larger role of AI in future weapon technology. In 2021, Navy Secretary *Carlos Del Toro* qualified AI as one of his top priorities for US naval forces.¹⁴ Russian President *Vladimir Putin* emphasized the importance of military robots and unmanned aerial vehicles to determine who will emerge victorious from the battlefield and has declared to increase spending on AI research and development (R&D).¹⁵

A recurring concern of ethicists is that in case AWS are deployed, life-and-death decisions would be delegated to machines and that humans would be degraded to mere data objects.¹⁶ In order to preserve human dignity, a sufficient level of human involvement over AWS is thus often demanded. Some-

¹¹ US Department of Defense, “Directive 3000.09”, November 21, 2021, <https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf> (last accessed 30 March 2022).

¹² ICRC, Convention on Certain Conventional Weapons (CCW), Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), 11-15 April 2016, Geneva, Views of the International Committee of the Red Cross (ICRC) on autonomous weapon system, 11 April 2016. See the ICRC’s updated position from 2021, ICRC position on autonomous weapon systems, ICRC position and background paper of 2021.

¹³ Daniele Amoroso, *Autonomous Weapons Systems and International Law* (Baden – Baden: Nomos 2020) 8–21.

¹⁴ Commander Edgar Jatho and Joshua A. Kroll, “Artificial Intelligence: Too Fragile to Fight?”, *US Naval Institute*, February 2022, https://www.usni.org/magazines/proceedings/2022/february/artificial-intelligence-too-fragile-fight?mc_cid=31b7a8c6db&mc_eid=cb69914e13 (last accessed 30 March 2022).

¹⁵ Anna Nadibaidze, “Russian Perceptions on Military AI, Automation and Autonomy”, *Foreign Policy Research Institute*, January 2022, <https://www.fpri.org/wp-content/uploads/2022/01/012622-russia-ai-.pdf> (last accessed 30 March 2022).

¹⁶ Amanda Sharkey, “Autonomous weapons systems, killer robots and human dignity”, *Journal for Ethics and Information Technology* 21 (2019): 75.

times the term *human agency* is employed instead,¹⁷ which can be largely described as the capacity of humans to preserve deterministic processes.¹⁸ In the context of AWS, human agency encompasses the capacity of humans to be sufficiently involved in the process of developing, acquiring and most of all using AWS in order to guarantee their use in line with both ethical and legal concerns.¹⁹ Legally speaking, preserving human agency entails the capacity of humans to remain the sole addressees of IHL obligations and to be responsible in case the use of such weapons leads to a violation of the law.²⁰ Human agency can be operationalized in various ways, *inter alia* by preserving human control.²¹ Even though there is no agreed upon definition of the concept of human control, the term can largely be understood as the requirement of placing sufficient limits on a weapon's target, the operational environment as well as human-machine interaction in order to guarantee compliance with both ethical and legal concerns.²² Maintaining human agency by preserving human control in order to guarantee the protection of

¹⁷ For example, Peter Asaro argues that AWS may only be deployed in line with ethical concerns if meaningful human control (understood as a manifestation of human agency) can be maintained. See Peter Asaro; "Autonomous Weapons and the Ethics of Artificial Intelligence", <https://peterasaro.org/writing/Asaro%20Oxford%20AI%20Ethics%20AWS.pdf> (last accessed 30 March 2022).

¹⁸ Cf. Erasmus Mayr, *Understanding Human Agency* (Oxford: Oxford University Press 2011) 3.

¹⁹ Cf. Alice Spazian, Arthur Holland Michel and Alisha Anand, "UNIDIR on Lethal Autonomous Weapons, Mapping our Research to the Discussions of the GGE on LAWS" (Geneva, Switzerland 2021), <https://unidir.org/sites/default/files/2021-07/UNIDIR%20on%20Lethal%20Autonomous%20Weapons%20-%20Final.pdf> (last accessed 30 March 2022).

²⁰ *Ibid.*

²¹ Article 36, "Autonomous weapon systems: Evaluating the capacity for 'meaningful human control' in weapon review processes", Discussion Paper (November 2017), <https://article36.org/wp-content/uploads/2013/06/Evaluating-human-control-1.pdf> (last accessed 5 August 2021); see also Article 36, "Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons", Briefing Paper for delegates at the Review Conference of the Convention on Certain Conventional Weapons (CCW) (Geneva, 12-16 December 2016); Article 36, "Key Element of Meaningful Human Control", Background paper to comments prepared by Richard Moyes, Managing Partner, Article 36, for the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) (Geneva, 11-15 April 2016), <https://article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf> (last accessed 5 August 2021). It should be noted, however, that various states within the GGE still object to the term "human control" and favor other terms instead, such as "appropriate level of human judgment". See, for example, United States, Commentaries on the Guiding Principles (2020), https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2020/gge/documents/US_2020.pdf (last accessed 30 March 2022).

²² Cf. Vincent Boulanin et al., "Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control", *Stockholm International Peace Research Institute (SIPRI)* together with the ICRC, June 2020, https://www.sipri.org/sites/default/files/2020-06/2006_limits_of_autonomy.pdf (last accessed 30 March 2022).

dignity of humans is not only an ethical demand but it largely feeds into the legal discussion on AWS and IHL.²³

2.2. The interlinkages between ethics and law in armed conflict

Dogmatically speaking, there are three ways how ethical concerns feed into the discussion on AWS. First, IHL is a body of law deeply entrenched in ethical concerns. From its very inception IHL required a weighing between military exigencies and considerations of humanity.²⁴ Second, the Martens Clause is often considered the legal “entry point” for ethical considerations in the context of IHL.²⁵ Even though the legal nature of the Martens Clause remains disputed in international law, especially with regard to the question of whether direct rights and obligations can be derived from it, it is the prevailing view that the Martens Clause at least serves as an additional source to interpret other existing rules of IHL, such as the principles of distinction, proportionality and precautions in attack.²⁶ According to the clause, “even in cases not covered by specific international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience”²⁷. The terms “princi-

²³ For more details see Elisabeth Hoffberger-Pippan, “Autonomous Weapon Systems and Human Control: Politically Desired or also Legally Required?”, in *Max Planck Yearbook of United Nations Law*, (eds.) Erika de Wet, Kathrin Maria Scherr and Rüdiger Wolfrum (Leiden: Brill 2021).

²⁴ Cf. Nobuo Hayashi, *Military Necessity: The Art, Morality and Law of War* (Cambridge: Cambridge University Press 2020) 55.

²⁵ Cf. Eliav Lieblich and Eyal Benvenisti, “The obligation to exercise discretion in warfare: why autonomous weapons systems are unlawful” in *Autonomous Weapons Systems: Law, Ethics and Policy*, (eds.) Nehal Bhuta et al. (Cambridge: Cambridge University Press 2016) 269; Robert Sparrow, “Ethics as a source of law: The Martens clause and autonomous weapons” *Humanitarian Law and Policy Blog* (14 November 2017), <https://blogs.icrc.org/law-and-policy/2017/11/14/ethics-source-law-martens-clause-autonomous-weapons/> (last accessed 5 August 2021).

²⁶ Cf. Antonio Cassese, “The Martens Clause: half a loaf or simply pie in the sky?”, *European Journal of International Law* 11 (2000): 15 et seq.

²⁷ UN Report of the International Law Commission on the Work of its Forty-sixth Session (2 May–22 July 1994) GAOR A/49/10, 317. See also Article 63 Geneva Convention I, Article 62 Geneva Convention II, Article 142 GC III and Article 158 Geneva Convention IV. Geneva Convention for the amelioration of the condition of the wounded and sick in armed forces in the field, Geneva, 12 August 1949, in force 21 October 1950, UNTS Vol. 75 No. 970; Geneva Convention for the amelioration of the condition of the wounded, sick and shipwrecked members of the armed forces at sea, Geneva, 12 August 1949, in force 21 October 1950, UNTS Vol. 75 No. 971; Geneva Convention relative to the treatment of prisoners of war, Geneva, 12 August 1949, in force 21 October 1950, UNTS Vol. 75 No. 972; Geneva Convention relative to the protection of civilian persons in time of war, Geneva, 12 August 1949, in force 21 October 1950, UNTS Vol. 75 No. 973.

ples of humanity” and “dictates of public conscience” merit further consideration in the process of identifying legal challenges regarding the deployment of AWS and in ascertaining *Immanuel Kant’s* role in the debate. *Judge Weeramantry* in the Dissenting Opinion on the *Legality of the Threat or Use of Nuclear Weapons Advisory Opinion* contended that human rights should be taken into consideration when interpreting the scope and meaning of the principles of humanity and dictates of public conscience²⁸. It is often argued that human dignity is not only an ethical concept but that it constitutes an own human *right* from which other human rights derive.²⁹ Human dignity can be found *inter alia* in the Universal Declaration of Human Rights³⁰ and is also mentioned in the preambles of the International Covenant on Civil and Political Rights³¹ as well as the International Covenant on Economic, Social and Cultural Rights³². Following the line of argumentation of *Judge Weeramantry*, human dignity as a hybrid between ethics and the law *does* play a pivotal role in the debate on AWS and has arguably found its legal entry point by being intrinsically intertwined with the “principles of humanity” but also the “dictates of public conscience”.³³

Third, human dignity as a distinct human right also plays a pivotal role in armed conflict independently from IHL. As the International Court of Justice in the *Legality of the Threat or Use of Nuclear Weapons Advisory Opinion*³⁴ stated, human rights also apply in armed conflict. Even though the exact relationship between human rights and humanitarian law is still disputed in international law with a plethora of different legal opinions having emerged, the findings of the Court at least prove that human rights and therefore

²⁸ *Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion) (Dissenting Opinion of Judge Weeramantry)* [1996] ICJ Rep 429, 490. See also Theodor Meron, “The Martens Clause, Principles of Humanity, and Dictates of Public Conscience” *American Journal of International Law* 94 no. 1 (2000): 84.

²⁹ For more details see George P. Smith, *Dignity as a Human Right?* (London: Lexington 2019) 1-15.

³⁰ Universal Declaration of Human Rights, 10 December 1948, General Assembly Resolution 27 A.

³¹ International Covenant on Civil and Political Rights, New York, 16 December 1966, in force 23 March 1976, UNTS Vol. 999 No. 14668.

³² International Covenant on Economic, Social and Cultural Rights, New York, 16 December 1966, in force 3 January 1976, UNTS Vol. 993 No. 145431.

³³ See, for example, Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE LAWS) Joint ‘Commentary’ on Guiding Principles A, B, C and D, submitted by Austria, Belgium, Brazil, Chile, Ireland, Germany, Luxembourg, Mexico, and New-Zealand (2020), <https://documents.unoda.org/wp-content/uploads/2020/09/GGE20200901-Austria-Belgium-Brazil-Chile-Ireland-Germany-Luxembourg-Mexico-and-New-Zealand.pdf> (last accessed 30 March 2022).

³⁴ ICJ, *Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion* of 8 July 1996, ICJ Reports 1996, para 25. See also ICJ, *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory, Advisory Opinion*, 9 July 2004, ICJ Reports 2004.

human dignity in the AWS debate cannot be ignored.³⁵ The Court further clarified the relationship between IHRL and IHL by stating that “[t]he test of what is an arbitrary deprivation of life, however, then falls to be determined by the applicable *lex specialis*, namely, the law applicable in armed conflict which is designed to regulate the conduct of hostilities.”³⁶ The term *lex specialis* insinuates the supremacy of IHL towards IHRL. But a closer scrutiny of the terminology employed in the Court’s advisory opinion suggests the conclusion that both legal regimes apply in armed conflict and that IHRL merely has to be interpreted *in light of* IHL considerations. It is neither necessary nor conducive to elaborate in greater detail the relationship between IHL and IHRL in armed conflict. Suffice it to say that it is the prevailing view in legal research and scholarship that these two bodies of law are not mutually exclusive but complement each other.³⁷ Thus, in addition to being inextricably intertwined with IHL, human dignity as a hybrid between ethics and law also plays an independent role in armed conflict as human rights continually apply and complement IHL obligations.

After having identified the three avenues how human dignity feeds into the legal discussion on AWS, it is now time to delve deeper into the exact meaning of human dignity in the context of the military use of AI.

2.3. Kant’s approach to human dignity: the AWS debate

Immanuel Kant is representative for so-called deontological ethics. According to this philosophical concept, an action is ethical if it follows predetermined rules irrespective of the particular consequences.³⁸ Unlike Christian philosophers, *Kant* does not derive those pre-determined rules from a spiritual source but stipulates that humans are able to give those rules themselves as they are self-determined beings with a free will.³⁹ Most importantly, *Kant* considers humanity to be an end of itself by emphasizing that one must “[a]ct in such a way that you always treat humanity, whether in

³⁵ Regarding the relationship between human rights and humanitarian law see, for example Gerd Oberleitner, *Human Rights in Armed Conflict: Law, Practice, Policy* (Cambridge: Cambridge University Press 2015); Noam Lubell, “Challenges in applying human rights law to armed conflict”, *International Review of the Red Cross* 87 (2005): 737–754.

³⁶ ICJ, *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, para. 25.

³⁷ For more details see Heinz-Joachim Heintze, “Theories on the relationship between humanitarian and human rights law”, in *Research Handbook on Human Rights and Humanitarian Law*, (eds.) Robert Kolb and Gloria Gaggioli (Cheltenham: Edward Elgar Publishing 2013) 57–59.

³⁸ Michael A. Slothe, *From Morality to Virtue* (Oxford: Oxford University Press 1992) 47.

³⁹ Immanuel Kant, *The Moral Law: Kant’s Groundwork of the Metaphysic of Morals* (HJ Paton Hutchinson and Co 1969).

your own person or in the person of any other, never simply as a means, but always at the same time as an end.”⁴⁰ *Kant*’s approach to human dignity has largely influenced the application and interpretation of Article 1 para. 1 of the German Basic Law⁴¹ according to which “[h]uman dignity shall be inviolable.”⁴² Especially in the academic debate on AWS, a prominent judgment by the German Constitutional Court dealing with Article 1 of the German Basic Law gained significant attention.⁴³ A closer scrutiny of the background of the case will clarify the relevance of the judgment for the AWS debate, especially the requirement of a sufficient level of human involvement. In 2005, the German Bundestag adopted a new law *inter alia* empowering the Minister of Defense to order the downing of a civilian aircraft if the Minister had substantial reasons to believe that the airplane would be used against the lives of others (by e.g hijacking).⁴⁴ The law sparked significant outcry among lawyers but also the public. The German Constitutional Court declared the law to be unconstitutional by, *inter alia*, infringing upon Article 1 para. 1 of the German Basic Law. Even though more lives could be saved by shooting down an airplane that was about to cause the lives of others, the Court seems to have based its conclusion on the assumption that such consequentialist and utilitarian calculus was incompatible with the constitutionally guaranteed right to human dignity and that humans must never, under any circumstances, be objectivized.⁴⁵ Thus, the Court widely acknowledged *Kant*’s approach to human dignity. The conclusions drawn by the Court were finally brought into the debate on AWS in order to support ethical as well as legal arguments calling for the maintenance of human agency and the preservation of a sufficient level of human involvement over AWS.⁴⁶ Even though it will not be possible in this article to elaborate in more detail the necessary type and degree of human control in order to deploy AWS in line with ethical and legal concerns, suffice it to say that using AWS without human control would lead to an unjustified objectification of humans. It would

⁴⁰ Ibid. 90–93.

⁴¹ Basic Law for the Federal Republic of Germany in the revised version published in the Federal Law Gazette Part III, classification number 100–101, as last amended by Article 1 of the Act of 28 March 2019 (Federal Law Gazette I p. 404).

⁴² Basic Law for the Federal Republic of Germany.

⁴³ Deutsches Bundesverfassungsgericht (BVerfG), Judgment of the First Senate of 15 February 2006 - 1 BvR 357/05 -, paras. 1–156. Para. 3. See also Aviation Security Act, Luftverkehrssicherheitsgesetz vom 11. Januar 2005 (BGBl. I S. 78), das zuletzt durch Artikel 1 des Gesetzes vom 22. April 2020 (BGBl. I S. 840) geändert worden ist.

⁴⁴ Ibid. § 14 para. 3.

⁴⁵ Cf. Oliver Lepsius, “Human Dignity and the Downing of Aircraft: The German Federal Constitutional Court Strikes Down a Prominent Anti-terrorism Provision in the New Air-transport Security Act” *German Law Journal* 7, no. 9, (2006): 766.

⁴⁶ Daniele Amoroso, *Autonomous Weapons Systems and International Law*, 122.

be the machine – not humans – making life-and-death decisions based on numbers and mere algorithms. Subjectively it might be a different feeling if the enemy combatant affected is neutralized by a military robot or a human soldier. In the latter case, the enemy combatant could still hope for mercy, interact and simply engage with the human soldier whose way of thinking and acting might be more intelligible to other humans than machine language and behavior.⁴⁷ Arguably, humans would also lose their dignity if AI-enabled technology was endowed with international legal personality thus becoming official addressees of IHL obligations. This would entail an unjustified anthropomorphization of AWS.⁴⁸ Such dogmatic consequences would add to the gradual legitimization of AI-enabled technology deployed without a sufficient level of human control ultimately relegating humans to objects or – at best – bystanders. The US-Supreme Court also seems to be observant of the need to maintain dignity. In *Woodson vs. North Carolina*, a case dealing with the mandatory application of the death penalty for specific crimes, the US-Supreme Court held that in case mandatory death penalty is applied for particular crimes people would not be treated as “uniquely human beings, but as members of faceless, undifferentiated mass”⁴⁹. Even though the case is not dealing with AWS, it is indicative of the Court’s awareness for human dignity as a deontological concept requiring that humans must not be reduced to undifferentiated mass or – in the context of AWS – mere data points.⁵⁰ It goes without saying that the mandatory application of the death penalty is incomparable to a situation of armed conflict.⁵¹ But the case at least demonstrates the necessity of maintaining human agency, that is to say, a human’s capacity to preserve deterministic processes.

Immanuel Kant’s philosophical approach stands in clear contrast with utilitarian and consequentialist ethics according to which “the morally right

⁴⁷ Daniele Amoroso, *Autonomous Weapon Systems and International Law*, 182.

⁴⁸ For more on the issue of anthropomorphization of AWS in an ethical context see Esther D. Reed, “Truth, Lies and New Weapons Technologies: Prospects for Jus in Silico?”, *Studies in Christian Ethics* 35, no. 1 (2022): 68–86.

⁴⁹ Supreme Court of the United States, *Woodson vs. North Carolina*, 428 US 280 (1976) 304. Cited in Maya Brehm, “Defining the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems Under International Humanitarian and Human Rights Law”, 2017, *Geneva Academy*, https://www.geneva-academy.ch/joomlatools-files/docman-files/Briefing9_interactif.pdf (last accessed 30 March 2022). See also Daniele Amoroso, *Autonomous Weapons and International Law*, 182.

⁵⁰ Maya Brehm, “Defining the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems Under International Humanitarian and Human Rights Law”, 2017, *Geneva Academy*, https://www.geneva-academy.ch/joomlatools-files/docman-files/Briefing9_interactif.pdf (last accessed 30 March 2022).

⁵¹ Marco Sassòli, “Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified” *International Law Studies* 90, no. 38 (2014): 323–333.

action is the action that produces the most good”⁵². From a merely utilitarian perspective, AWS could be ethically acceptable since they can – as is often argued – increase the accuracy of weapon systems and thereby decrease the likelihood of civilian casualties, especially in case AI-enabled technology is deployed.⁵³ *Ronald Arkin* even argues that AWS would be able to comply with ethical requirements if they were programmed accordingly.⁵⁴ And even the terminology employed in Additional Protocol I of 1977 to the Geneva Conventions of 1949⁵⁵ seems to take a consequentialist rather than a deontological approach by stating that ‘an attack 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’⁵⁶ is an indiscriminate attack and thus prohibited by IHL.

Ethical concerns have played a major role in the debate on AWS in the GGE in Geneva. At least two different groups have crystallized: states, which support the idea of beneficial AI able to make outcomes more predictable and weapon systems more accurate⁵⁷ and states which warn against the dangers when outsourcing critical functions such as target selection and engagement to machines instead of retaining a sufficient level of human involvement.⁵⁸

⁵² Stanford Encyclopedia Philosophy, “Utilitarianism”,

<https://plato.stanford.edu/entries/utilitarianism-history/> (last accessed 30 March 2022).

⁵³ John Cherry and Christopher Korpela, “Enhanced distinction: The need for a more focused autonomous weapons targeting discussion at the LAWS GGE”, *ICRC Blog*, March 28, 2019, <https://blogs.icrc.org/law-and-policy/2019/03/28/enhanced-distinction-need-focused-autonomous-weapons-targeting/> (last accessed 30 March 2022); Jai Galliot and Austin Wyatt, “Risks and Benefits of Autonomous Weapon Systems: Perceptions among Future Australian Defence Force Officers”, November 24, 2020, <https://www.airuniversity.af.edu/JIPA/Display/Article/2425657/risks-and-benefits-of-autonomous-weapon-systems-perceptions-among-future-austra/> (last accessed 30 March 2022).

⁵⁴ Ronald Arkin, *Governing Lethal Behavior in Autonomous Robots* (Boca Raton: CRC Press 2009) 200.

⁵⁵ Protocol additional to the Geneva Conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (Protocol I), Geneva, 8 June 1977, in force 7 December 1978, UNTS Vol. 1125 No. 17512.

⁵⁶ Article 51 para. 5 lit. b) Protocol I. See also Ozlem. Ulgen, “Kantian Ethics in the Age of Artificial Intelligence and Robotics” *Questions of International Law* (2017) 23.

⁵⁷ See, for example, statements made by the Russian Federation, Working Paper of the Russian Federation, National Implementation of the Guiding Principles on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, <https://documents.unoda.org/wp-content/uploads/2020/09/Ru-Commentaries-on-GGE-on-LAWS-guiding-principles1.pdf> (last accessed 30 March 2022); United States, U.S. Commentaries on the Guiding Principles, <https://documents.unoda.org/wp-content/uploads/2020/09/20200901-United-States.pdf> (last accessed 30 March 2022).

⁵⁸ Austria, 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 Group of Governmental Experts on Emerging Technologies in the Area of Lethal

The first group of states does not argue that AWS should be employed without a sufficient level of human involvement. But their narrative is different as they – in contrast to the second group – reiterate the positive effects that can be eventually created by AI-enabled technology in armed conflict. The second group finds support by *inter alia* Christoph Heyns, former UN-Special Rapporteur on extrajudicial, summary or arbitrary executions. In 2016, during an informal meeting on AWS, he stated – in a clearly *Kantian* way – that “a human being in the sights of a fully autonomous machine is reduced to being an object – being merely a target. This is death by algorithm.”⁵⁹

Already in April 2013, *Christof Heyns* argued that taking humans entirely out of the loop is hardly compatible with human dignity.⁶⁰ In a statement in 2014, he emphasized that “giving machines greater power to take life and death decisions is demeaning”.⁶¹ The author *Patrick Lin* argues in a similar direction claiming that the delegation of targeting decisions to machines infringes upon human dignity of the targeted people, even in case such targeted people are legitimate military targets from an IHL perspective.⁶² Especially the deliberations made by *Christof Heyns* have largely fed into the discussion on the legality of the use of AWS and the compatibility with their deployment in armed conflict with ethical considerations.⁶³ Most importantly, they reflect the deontological, non-consequentialist approach that was already taken by *Immanuel Kant*.

The above observations clearly show that AWS must be deployed with a sufficient level of human involvement. Even though a number of states in the GGE favor a consequentialist approach towards the potential use of military AI, a large number of states parties maintain that the requirement of

Autonomous Weapons Systems 2020 Contribution of Austria to the Chair’s request on the Guiding Principles on emerging technologies in the area of LAWS, <https://documents.unoda.org/wp-content/uploads/2020/09/20200901-Austria.pdf> (last accessed 30 March 2022); Germany, General Statements made at an Informal Meeting of Experts on AWS, Geneva, 13–16 May 2014, at 1 et seq. See also Daniele Amoroso, *Autonomous Weapons Systems and International Law*, 182.

⁵⁹ See Johanna Friman, “Death by algorithm?”, Blogpost at ETAIROS, <https://etairos.fi/2020/06/24/death-by-algorithm/> (last accessed 30 March 2022).

⁶⁰ Human Rights Council, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns (9 April 2013) UN Doc. A/HRC/23/47, paras. 89–97.

⁶¹ Christoph Heyns, “The Challenge of Autonomous Weapons Systems to Legal Regulation”, paper presented at the Conference on Autonomous Weapons Systems: Law, Ethics, Policy, Academy of European Law (24–25 April 2014). See Dieter Birnbacher, “Are autonomous weapons systems a threat to human dignity?”, 115.

⁶² Patrick Lin, “The Right to Life and the Martens Clause”, Convention on Certain Conventional Weapons (CCW) meeting of experts on lethal autonomous weapons systems (LAWS), (Geneva, 13–17 April 2015), http://ethics.calpoly.edu/ccw_testimony.pdf (last accessed 30 March 2022).

⁶³ Cf. Daniele Amoroso, *Autonomous Weapons Systems and International Law*, 182 et seq.

human control is a necessary piece in the puzzle to protect humans from the increased use of military AI, irrespective of a weapon's particular outcome. Given the current stalemate in the GGE and the arduous process of finding at least a minimum consensus regarding a potential regulation of AWS, it is still unclear which ethical approach will gain traction. The strong support by *inter alia* Christof Heyns and the dogmatically proven strong undeniable interlinkage between ethics and IHL suggests that deontological aspects will at least not go unnoticed and will continue to counterbalance the rather utilitarian approach taken by some states in the GGE, such as the United States and Russia.

3. Kant's legacy: repercussions for norm-creating processes and governance perspectives

Kant's understanding of human dignity as an inherent right of all human beings and the concomitant deontological, non-consequentialist approach not only impacts IHL but what can also be observed is a significant influence on international law and the use of military AI in a more general context. As *Ozlem Ulgen* righteously noted *Kant* provides “a human-centric ethical framework whereby human existence and capacity are at the centre of a norm-creating moral philosophy guiding our understanding of moral conduct”⁶⁴. Various norm-creating processes have been influenced by the concept of upholding human agency instead of vacating the law to a modernist perception of the new legal order where humans and machines are considered equipollent actors in international law. In 2018, the European Parliament (EP) adopted a resolution on AWS emphasizing that the latter must not be used without human control. This line of argumentation entails the requirement to preserve human agency over the use of force but it may also have impact on the question of how norms dealing with AI-enabled technology should be created in future.⁶⁵ Likewise, in 2021, the EP adopted a resolution on AI and the applicability of international law emphasizing “that AI used in a military and a civil context must be subject to meaningful human control”⁶⁶. In the same year, the European Commission submitted a proposal to regulate AI clearly indicating that algorithmic decision-making will

⁶⁴ Ozlem Ulgen, “Kantian Ethics in the Age of Artificial Intelligence and Robotics”.

⁶⁵ European Parliament, European Parliament resolution of 12 September 2018 on autonomous weapon systems (2018/2752(RSP)), P8, TA (2018)0341.

⁶⁶ European Parliament resolution of 20 January 2021 on artificial intelligence: questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice, 2020/2013(INI), P9_TA(2021)0009, at 4.

still be subjected to human control and/or human oversight.⁶⁷ By the same token, the African Commission on Human and Peoples' Rights adopted a resolution in 2021, addressing robotics, AI and algorithmic decision-making.⁶⁸ Cognizant of the fact that algorithms will inevitably be pervasive in our daily lives, the African Commission also cautioned against transferring too many competences to machines. At its 47th regular session taking place from 21 June to 14 July 2021, the Human Rights Council (HRC) discussed the legal implications of AI and algorithmic decision-making.⁶⁹ On 5 July 2021, the International Commission of Jurists in front of the Council, warned that “no artificial intelligence solution can fully replace human judgment”⁷⁰. And from 5th July to 7th July 2021, the Council of Europe's Ad Hoc Committee on Artificial Intelligence (CAHAI) convened online to discuss human rights related issues in the context of AI.⁷¹ All these examples show that even though technological advancements in the fields of robotics and AI may yield various societal benefits, algorithms are not considered new actors partially replacing human beings and requiring a re-configuration of international law. Rather, human agency will be preserved and so will the dogmatic foundations and structural characteristics of international law.

Attempts to address algorithmic decision-making have not only been undertaken at the international but also the domestic level. Various states have developed laws and policy guidelines dealing with recent advancements in the fields of robotics and AI. Thus, the following section will discuss national laws and policies dealing with human agency and algorithmic decision-making in various contexts, which inevitably raises the question of whether we are currently witnessing the development of a so-called digital constitutionalization of the global (digital) order shaping and restraining how algorithmic decision-making will be perceived and what role humans will play in it.

⁶⁷ Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 21 April 2021, COM(2021) 206 final, 2021/0106(COD).

⁶⁸ Resolution on the need to undertake a Study on human and peoples' rights and artificial intelligence (AI), robotics and other new and emerging technologies in Africa - ACHPR/Res. 473 (EXT.OS/ XXXI) 2021.

⁶⁹ United Nations Human Rights Council, 47th regular session of the Human Rights Council (21 June to 14 July 2021), <https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session47/Pages/47RegularSession.aspx> (last accessed 30 March 2022).

⁷⁰ International Commission of Jurists, “ICJ calls for a holistic human rights approach to artificial intelligence before UN Human Rights Council”, July, 5, 2021, available at <https://www.icj.org/icj-calls-for-a-holistic-human-rights-approach-to-artificial-intelligence-before-un-human-rights-council/> (last accessed 30 March 2022).

⁷¹ Council of Europe, Outcome of the 5th CAHAI Plenary meeting, July 7, 2021, at <https://www.coe.int/en/web/artificial-intelligence/-/upcoming-cahai-plenary-meeting> (last accessed 30 March 2022).

4. Human Agency: Is it Time for a Digital Constitutionalization in International Law?

With more and more technological advancements on the rise, which are not necessarily related to military operations, it is arguable that a common international framework (rather in a political and less in a legal sense) or at least a certain degree of coherence among regulators when addressing these developments is about to emerge or has already done so, which could be described as “digital constitutionalization”. This concept is not new but recent technological developments will continue to provide new impetus to this debate. In his article, “The Constitutionalisation of the Digital Ecosystem”⁷² *Edoardo Celeste* touches upon the question of whether a digital constitutionalization is emerging and summarizes measures and trends to regulate digital technologies respectively as “normative counteractions” of states and the international community in order to respond to an increasingly digitalized world. The terminology employed by *Celeste* (*counteractions*) seems to imply already a particular approach that is being taken by the author: despite the fact that technological advancements will yield a significant number of societal benefits, solutions will have to be found in order to address and potentially regulate algorithmic decision-making and other technological capabilities.

The concept of a constitutionalist normative order beyond the state dimension has been addressed by *inter alia* *Alfred Verdross* in his book “The Constitution of the International Legal Community”⁷³ published in 1926. Even more importantly, it was *Kant* himself who argued that there is or there should be an overarching international normative order characterized by constitutional principles and structures framing international law and policymaking.⁷⁴ But unlike other proponents of a constitutionalist approach to international law, *Kant* stipulated that the normative global order should not be characterized by the maximization of utility (thereby repudiating the so-called *homo economicus*), but based his conclusions on the premise that the global order entails inherent values deeply rooted in the freedom of humans as creatures capable of acting with reason.⁷⁵ The findings of *Kant* together with his approach to human dignity are helpful in identifying an emerging process of

⁷² Edoardo Celeste, “The Constitutionalisation of the Digital Ecosystem: Lessons from International Law”, June 23, 2021, *Max Planck Institute for Comparative Public Law and International Law (MPEPIL) Research Paper No. 2021-16*, at 1.

⁷³ Alfred Verdross, *Die Verfassung der Völkerrechtsgemeinschaft* (Wien: Springer 1926).

⁷⁴ Immanuel Kant, *Perpetual Peace: A Philosophical Sketch* (1795), paras. 102–105.

⁷⁵ *Ibid.* With critical remarks see Marti Koskeniemi, “Constitutionalism as Mindset: Reflections on Kantian Themes About International Law and Globalization” *Theoretical Inquiries in Law* 8, no. 1 (2009): 11.

digital constitutionalization in the global order. Whilst it is undisputed that technological advancements will yield a number of benefits, international actors, ranging from different forums, such as the GGE on AWS, and organizations and organs thereof, such as the EP, the EC, the African Commission of Human and Peoples' Rights, to the Council of Europe, seem to have taken the position that technological advancements will have to serve humans and human decision-makers instead of – at least in the legal realm – replacing them as the primary actors and addressees of the law. This approach is not only reflected at the international level. A number of states have adopted laws or policy guidelines on ethics, AI and algorithmic decision-making. The various measures taken seem to guide in the same direction: the preservation of human agency over algorithmic decision-making. China, for example, published the “New Generation Artificial Intelligence Development Plan”⁷⁶, the first official document clarifying China’s position towards AI-enabled systems. Even though China primarily praises the various, almost infinite advantages of AI, it also emphasizes the importance of rules (without reconfiguring them in order to be more “open” to technological advancements, by, e.g. establishing own liability regimes for computers⁷⁷) and ethical principles providing guidance on how AI and algorithmic decision-making respectively will and should be used in society, China also emphasizes that future technological developments are primarily based on human-machine collaboration. Thus, it seems that China shows awareness for the requirement of human agency, whereas it does not consider it necessary to amend structural principles of the law so as to guarantee that AI technology would be granted more rights or even be endowed with legal personality.⁷⁸ However, it is also worthy of note that the Chinese Artificial Intelligence Development Plan does not explicitly mention ethical concerns in the military context. Rather, it mentions ethical dimensions in broader, mainly civilian contexts. Thus, it remains to be seen how the Chinese government positions itself towards ethical considerations in the military realm.⁷⁹ China’s position paper on reg-

⁷⁶ Notice of the State Council Issuing the New Generation of Artificial Intelligence Development Plan, State Council Document [2017] No. 35, https://www.unodc.org/res/ji/import/policy_papers/china_ai_strategy/china_ai_strategy.pdf (last accessed 30 March 2022).

⁷⁷ This idea has also been rejected by the European Parliament in a resolution of October 2020. See European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence, 2020/2014(INL)), P9_TA (2020)0276, para. 6.

⁷⁸ *Ibid.*

⁷⁹ For more details see iPRAW, “Focus on National Regulations on LAWS and Military AI”, August 2021, accessible at https://www.ipraw.org/wp-content/uploads/2021/08/iPRAW-Report_NationalRegulations_August2021.pdf (last accessed 21 February 2022);

ulating military applications of AI of December 2021 may reveal more clarity in this regard. The document stipulates that the military use of AI poses a number of risks including ethical challenges. Furthermore, the position paper emphasizes the importance of maintaining human control over AI-enabled weapon systems.⁸⁰ Already in 2019, China created the National New Generation Artificial Intelligence Governance Committee publishing eight principles on the governance of AI-enabled technology.⁸¹ Other attempts at dealing with the ethical dimension of AI were made by the Beijing Academy of Artificial Intelligence, a conglomerate of major Chinese universities and the private sector, by publishing the Beijing AI Principles in 2019⁸². Again, the principles do not explicitly mention the potential military use of AI and concomitant ethical concerns but rather focusses on general AI principles including human dignity.⁸³

By the same token, Russia made its first statements on AI in 2018⁸⁴ indicating that AI and algorithmic decision-making respectively, should comply “with existing requirements”⁸⁵. It is difficult to assess whether Russia intends to change existing legal structures thereby amending the core tenets of international law. The statements made in its report of 2018 seem to suggest that in principle human agency will be preserved and new technological developments should rather serve humans for various purposes, from the civil sector to the military realm.⁸⁶

The US have also adopted a very detailed strategy on AI by adopting the National Artificial Intelligence Initiative⁸⁷ highlighting that AI will serve hu-

⁸⁰ China, Position Paper of the People’s Republic of China on Regulating Military Applications of Artificial Intelligence (AI), 2021, http://www.china-un.ch/eng/dbdt/202112/t20211213_10467517.htm (last accessed 20 February 2022).

⁸¹ Huw Roberts et al., “The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation”, *AI and Society*, 36 (2021): 68; the AI principles can be found at http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html (last accessed 21 February 2022).

⁸² Will Knight, “Why does Beijing suddenly care about AI ethics?”, *MIT Technology Review*, May, 31, 2019, available at <https://www.technologyreview.com/2019/05/31/135129/why-does-china-suddenly-care-about-ai-ethics-and-privacy/> (last accessed 21 February 2022). The original link to the principles is currently not accessible.

⁸³ *Ibid.*

⁸⁴ Конференция «Искусственный интеллект: проблемы и пути их решения – 2018», <https://mil.ru/conferences/is-intellekt.htm> (last accessed 30 March 2022). Seen in Stephanie Petrella, Chris Miller and Benjamin Cooper, “Russia’s Artificial Intelligence Strategy: The Role of State-Owned Firms” *Foreign Policy Research Institute* (2021): 80.

⁸⁵ Samuel Bendett, “Here’s How the Russian Military Is Organizing to Develop AI”, July, 20, 2018, <https://www.defenseone.com/ideas/2018/07/russian-militarys-ai-development-roadmap/149900/> (last accessed 30 March 2022).

⁸⁶ With some general remarks on Russia’s approach to AI in the military realm see Samuel Bendett, “Here’s How the Russian Military Is Organizing to Develop AI”, *supra* note 140.

⁸⁷ National Artificial Intelligence Act of 2020 (secs. 5001 – 5501), <https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210> (last accessed 30 March 2022).

manity in various fields, while at the same time abstaining from the development of harmful AI, that is to say, technology that may become “self-aware or uncontrollable”⁸⁸. By the same token, the US DoD adopted ethical principles on the use of AI highlighting inter alia that AI must be used responsibly, that is to say, that “DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities”⁸⁹. The US explicitly mentions the applicability of these ethical principles to the military realm. This line of argumentation also corresponds with the Organization of Economic Co-operation and Development’s (OECD) Principles on AI published in 2019⁹⁰. The principles explicitly mention core values, such as human dignity and emphasize the importance of upholding accountability in order to ensure that liability regimes are put in place reflecting adequately the responsibility of humans to guarantee that AI-enabled technology is developed and used in line with both ethical and legal obligations.⁹¹

In light of the above it can be observed that a principle is about to emerge or has already done so according to which human agency with regard to algorithmic decision-making will be preserved. Without doubt, the question of how human agency will be implemented and operationalized still needs to be further clarified and will highly depend on both institutional as well as cultural differences.⁹² Even more importantly, the US, China and Russia frequently emphasize the potentially beneficial *effects* of AI and thus rather seem to take a consequentialist rather than a deontological approach.⁹³ Notwithstanding, the approaches taken by the US, China and Russia emphasize

⁸⁸ Ibid. At 2136.

⁸⁹ United States Department of Defense, “DOD Adopts Ethical Principles for Artificial Intelligence”, press release, February 24, 2020, <https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/> (last accessed 20 February 2022).

⁹⁰ OECD, “Recommendation of the Council on Artificial Intelligence”, May, 22, 2019 <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> (last accessed 30 March 2022).

⁹¹ Ibid.

⁹² Huw Roberts, “The Chinese approach to artificial intelligence: an analysis of policy, ethics and regulation”, 68.

⁹³ With regard to the United States see United States, Commentaries on the Guiding Principles, September 1, 2020, <https://documents.unoda.org/wp-content/uploads/2020/09/20200901-United-States.pdf> (last accessed 20 February 2022); with regard to Russia see Anna Nadibaidze, “Russia’s Perspective on Human Control and Autonomous Weapons: Is the Official Discourse Changing?”, June 3, 2021, The Autonorms Project, Research Article, <https://www.autonorms.eu/russias-perspective-on-human-control-and-autonomous-weapons-is-the-official-discourse-changing-2/> (last accessed 20 February 2022); with regard to China see China, Position Paper of the People’s Republic of China on Regulating Military Applications of Artificial Intelligence (AI), 2021, http://www.china-un.ch/eng/dbdt/202112/t20211213_10467517.htm (last accessed 20 February 2022).

the importance of maintaining human agency over AI-enabled systems including systems used in the military realm and thus align with *Kant's* understanding of international law as a normative human-centric legal order. Instead of shattering the basic structures of international law, future technological developments will rather serve humans and human agency will be maintained. This does not mean, however, that developing AI-enabled technology is free from risk. On the contrary, an adequate human-machine relationship fitting into already existing legal structures as a set of norms created by humans for humans is of pivotal importance and must be adhered to, especially by great military powers, which often promise more than they can achieve on today's battlefields.

5. Human Agency translated into concrete action: the concept of human control/judgment

The foregoing analysis has revealed that both from an ethical and a legal perspective, humans must be sufficiently involved in the process of developing, deploying and using AWS. This rather abstract and still vague assessment necessarily begs the question how and according to what parameters humans should be precisely involved. Three different degrees of human control and/or human judgment⁹⁴ can be discerned feeding into the overall ethical and legal debate.

First, various authors have argued that programming an AWS suffices in order to maintain an adequate level of human control and/or judgment and thus guarantee compliance with ethical and legal standards. Already in 1942 *Isaac Asimov*, one of the most influential science-fiction writers, contemplated how robots could be programmed to comply with ethical and legal standards.⁹⁵ While *Asimov* primarily dealt with robots in a civilian context, *Ronald C. Arkin* developed his idea further by focusing on the use of robots in military operations. Similar to *Asimov*, *Arkin* argues that it is possible to program a robot in a way so as to guarantee compliance with *inter alia* ethical concerns.⁹⁶ Not only that, the author also speaks openly about “taking the

⁹⁴ For the purpose of this article these two words can be used interchangeably. It is worthy of note, however, that some delegations to the UN-GGE on AWS favor the term human control, while others favor the term human judgment. There are also delegations which are supporting neither of the two terms. For more information see International Panel on the Regulation of Autonomous Weapons (iPRAW), “Building Blocks for a Regulation on LAWS and Human Control”, 2021, https://www.ipraw.org/wp-content/uploads/2021/07/iPRAW-Report_Building-Blocks_July2021.pdf (last accessed 30 March 2022).

⁹⁵ Isaac Asimov, *Runaround* (Nightfall Inc. 1982), available at <http://faculty.las.illinois.edu/rushing/395/ewExternalFiles/Asimov--Robots.pdf> (last accessed 30 March 2022).

⁹⁶ Ronald C. Arkin, “The Case for Ethical Autonomy in Unmanned Systems”, available at

man out of the loop” as most IHL violations are allegedly committed by irrational humans and not by technical failures.⁹⁷ Arkin proposes the creation of an “ethical governor”, a robot capable of taking decisions in line with ethics and IHL and without any form of interference by humans (except for the programming stage).⁹⁸ But even latest technological innovation, such as AI-enabled software based on neural networks⁹⁹, is largely incapable of making adequate ethical judgments. One such example is “Delphi”, a research project by the Allen Institute for AI aimed at investigating how AI-enabled software would be able to make ethical judgments based on how it was programmed.¹⁰⁰ The system is rather simple as it mainly uses large amounts of text from the internet. Participants of the project should then read the text and indicate how they would behave in a certain way so as to guarantee their actions would comply with ethical principles. Delphi is fed with all these data (especially the answers given by humans) and should thus be able to make “correct” ethical judgments. But upon being asked whether genocide would be okay under certain circumstances, Delphi answered “yes” – despite having been fed with a huge amount of data from participants who have all verifiably indicated that genocide can never be justified.¹⁰¹ Without doubt, Delphi will not be the end of scientific research but it clearly demonstrates the limits and restraints placed upon humans when creating an AI-enabled ethical governor. Battlefield dynamics further exacerbate the problem as machines would be confronted with various, often unforeseen circumstances and would thus be likely incapable of acting in line with ethics or the law.¹⁰²

https://www.cc.gatech.edu/ai/robot-lab/online-publications/Arkin_ethical_autonomous_systems_final.pdf (last accessed 30 March 2022).

⁹⁷ Cf. Frank Pasquale, “Machines set loose to slaughter’: the dangerous rise of military AI”, *The Guardian*, 15 October 2020, available at <https://www.theguardian.com/news/2020/oct/15/dangerous-rise-of-military-ai-drone-swarm-autonomous-weapons> (last accessed 30 March 2022).

⁹⁸ Ronald C. Arkin, Patrick Ulam and Brittany Duncan, “An Ethical Governor for Constraining Lethal Action in an Autonomous System”, Technical Report GIT-GVU-09-02, available at <https://smartech.gatech.edu/bitstream/handle/1853/31465/09-02.pdf> (last accessed 30 March 2022).

⁹⁹ For more details on neural networks and machine learning see Kaushal Kumar, Gour Sundar Mitra Thakur, “Advanced Application of Neural Networks and Artificial Intelligence”, *International Journal for Information Technology and Computer Science* 6 (2012): 57–68.

¹⁰⁰ Ask Delphi, available at <https://delphi.allenai.org/> (last accessed 30 March 2022).

¹⁰¹ Matthew Gault, “Ethical AI Trained on Reddit Posts Said Genocide Is OK If It Makes People Happy”, *Vice*, 3 November 2021, <https://www.vice.com/en/article/v7dg8m/ethical-ai-trained-on-reddit-posts-said-genocide-is-okay-if-it-makes-people-happy> (last accessed 30 March 2022).

¹⁰² Cf. Noel E. Sharkey, “The evitability of autonomous robot warfare” *International Review of the Red Cross* 94 No. 886 (2012): 799.

The second approach determining the necessary level of human involvement in case AWS are deployed places more emphasis on external factors. While authors like *Arkin* maintain that adequate programming suffices to create an ethical governor capable of making correct assessments, proponents of the second approach argue that it is also necessary for humans to actively place restraints on the target, the operational environment as well as the concrete level of human-machine interaction.¹⁰³ For example, in case an AWS – albeit being programmed accordingly – is deployed in an urban environment with many civilians present, it will be almost impossible to use such a weapon system in line with IHL. Circumstances may change rapidly, civilians could suddenly take up their arms and directly participate in hostilities or enemy combatants could be rendered *hors de combat*.¹⁰⁴ Given these fast-changing dynamics with various unforeseen circumstances emerging in the fog of war, it could be the far better option not to e.g. deploy such weapon systems in urban environments at all or to evacuate buildings and to give sufficient warnings before a military operation takes place. But this needs to be undertaken by human soldiers on the ground as pre-programmed AWS are not able to take account of such complex environments with different tasks that need to be fulfilled. The ICRC, the Stockholm Peace Research Institute (SiPRI) but also various States in the UN-GGE on AWS seem to favor this approach.¹⁰⁵

The third approach is to require – depending on the operational circumstances – that humans must maintain options to intervene in a weapon system's mode of operation. Such an idea of human control and/or judgment should not be confused with drone operations, however. The crucial difference is that even in case humans retain options to intervene with regard to AWS, it would still be the weapon system undertaking the process of target selection and engagement.¹⁰⁶ Proponents of the third approach argue that solely placing limits on the types of target, the environment and human-machine interaction does not suffice in order to pay due regard to unforeseen

¹⁰³ Vincent Boulanin et al., “Limits on Autonomy in Weapon Systems: Defining Practical Elements of Human Control”, June 2020, <https://www.icrc.org/en/document/limits-autonomous-weapons> (last accessed 30 March 2022).

¹⁰⁴ Cf. Daniele Amoroso, *Autonomous Weapons Systems and International Law*, 61.

¹⁰⁵ ICRC Position on Autonomous Weapons, 12 May 2021, <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems> (last accessed 30 March 2022); National Statement by Germany Group of Governmental Experts on “Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS)” 3 – 13 August 2021, delivered by Ambassador Peter Beerwerth, <https://documents.unoda.org/wp-content/uploads/2021/08/Germany.pdf> (last accessed 30 March 2022).

¹⁰⁶ Cf. Alex Leveringhaus, *Ethics and Autonomous Weapons* (London, Palgrave MacMillan: 2016) 49.

circumstances in dynamic battlefield operations. The International Panel on the Regulation of Autonomous Weapons (iPRAW), for example, emphasizes that a “kill box” or “boxed autonomy”¹⁰⁷, where merely external factors are regulated accordingly without proscribing the possibility for humans to intervene, might have devastating consequences in urban environments and other highly sensitive and dynamic areas.¹⁰⁸ Imagine an AWS being deployed in already evacuated city against enemy combatants. What if the information passed to the AWS according to which all houses are entirely evacuated and all remaining persons are thus to be considered enemy combatants was wrong due to technical malfunction? While human oversight might in this case help drastically reduce or even eliminate unjustified civilian casualties, an AWS deployed without human oversight would probably not be able to adapt its behavior and mode of action so as to guarantee compliance with the law. By the same token, SiPRI and the ICRC also argue that solely placing restraints on the types of target, environment and human-machine interaction might not suffice to comply with ethical and legal standards under certain circumstances. On the contrary, especially in scenarios where the presence of civilians is likely, more safeguards could be necessary that would also entail the human operator’s capacity to intervene in the process of operation.¹⁰⁹

Thus, while it is clear that human agency will prevail over robots in more general terms and over AWS more specifically, the question of how exactly this assessment translates into “concrete action” largely remains unanswered as much depends on the concrete operational context. While it seems rather clear that programming itself does not suffice to guarantee the deployment of AWS in line with ethical and legal concerns, it is still unclear when options for intervention are needed and when more leeway can be granted to machines by e.g. solely placing limits and constraints on external factors, such as targets and the environment. It is safe to argue, however, that the more likely it is that civilians are present, the higher the degree of human

¹⁰⁷ For more details on this concept see Daniele Amoroso and Guglielmo Tamburrini, “Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues”, *Current Robotics Reports* 1 (2020): 187–194.

¹⁰⁸ iPRAW, “Focus on Human Control”, 2019, https://www.ipraw.org/wp-content/uploads/2019/08/2019-08-09_iPRAW_HumanControl.pdf (last accessed 30 March 2022).

¹⁰⁹ Vincent Boulanin et al., “Limits on Autonomy in Weapon Systems: Defining Practical Elements of Human Control”, June 2020, <https://www.icrc.org/en/document/limits-autonomous-weapons> (last accessed 30 March 2022); ICRC Position on Autonomous Weapons, 12 May 2021, <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems> (last accessed 30 March 2022).

involvement over the use of force – potentially including the requirement of human oversight and options to intervene respectively – is necessary.¹¹⁰

6. Conclusion

Technological developments are pervasive in our daily lives. They will bring about new advantages and techniques able to fulfill tasks a lot better than humans. But these developments also raise fundamental questions, including whether algorithmic decision-making is able or necessitates a re-thinking of the law itself. The debate on AWS and recent developments in the area of AI including the military sector serve as an ideal example proving that even in case of autonomous decision-making, there are strong indications that human agency will be preserved. However, increasingly outsourcing core competences to machines bears the risk that humans may lose control over weapons systems and might thus be incapable of complying with their legal and ethical obligations. Thus, arguing that human agency and humans respectively as the sole addressees of the law will remain by no means indicates that the deployment of e.g. AWS is completely without risks. But the core tenets of international law cannot be shattered by the advent of such weapons systems. Most importantly, the role of ethics, especially human dignity and *Kant's* understanding thereof, has largely influenced the AWS debate where a number of states demand that human agency has to prevail and that sufficient safeguards – whether politically or legally binding – will be established to guarantee that such weapons systems are employed in line with the law. Even though it is true that some autonomous functions may have positive effects, such as increased accuracy, it is also true that according to various scholars and philosophers humans must not be reduced to sole objects of algorithmic decision-making irrespective of whether the consequences were “good” by having reduced collateral damage, for example. The deliberations by *Kant* imply that even in times of significant progress in AI research and development, the core structures of international will likely remain the same. *Kant's* approach to human dignity has influenced a number of human rights documents, which, together with other legal but also political sources, will have significant impact on the question of how humans should position themselves towards technological developments. Another driving factor speaking in favor of the preservation of human agency are national approaches to AI and algorithmic decision-making. It could be observed that the great military powers, such as China, Russia, and the United States, despite aiming to invest into AI research and development, have little

¹¹⁰Ibid.

interest in re-configuring international law. Rather, they thrive to maintain human agency and use technology for their own military benefits but – as they claim – also for the benefit of humanity. This does not mean, however, that we should not be cautious when it comes to AI-enabled technology. It could likely be misused, especially by autocratic regimes and especially if used in a military context, raising significant legal and ethical concerns but also international security challenges. Moreover, the assessment that human agency will prevail leaves the question of how humans could be sufficiently involved in the process of developing, deploying or using AWS, largely unanswered. While a number of theories and concepts have been developed thus far, the UN-GGE was hitherto incapable of agreeing on substantial rules dealing with AWS and the question of how they could be used in line with both ethical and legal concerns. It is obvious, however, that more limits and constraints must be foreseen in environments where the presence of civilians is likely including options to intervene and/or abort mission. In light of the aforementioned it is important to emphasize that we should be careful about technological developments. However, we should not be concerned that the structure of international would need to be fundamentally changed. It is still humans calling the tune and it is our responsibility to put additional standards in place that guarantee such weapon systems will be deployed in line with ethical standards and IHL requirements.

