

ARTICLE

Special Issue: International Law and Digitalization

# Constructing “Electronic Liability” for International Crimes: Transcending the Individual in International Criminal Law

Mia Swart

Edge Hill University, Ormskirk, United Kingdom  
Email: [swartm@edgehill.ac.uk](mailto:swartm@edgehill.ac.uk)

(Received 11 April 2023; accepted 11 April 2023)

## Abstract

It is increasingly clear that autonomous agents can commit international crimes such as torture and genocide. This article aims to construct ‘electronic liability’ for such international crimes. It will argue that it is not sufficient to hold the persons or programmers behind the autonomous agents liable, but that it should be possible to hold the autonomous agents that commit international crimes liable. It will examine ways in which legal personality can be attributed to machines and argue that if there is a continuum of potential subjects of ICL, then the argument for electronic personhood and liability of machines is as compelling as for other non-humans such as corporate entities and animals. It will be argued that the ICC will potentially only be able to meaningfully prosecute international crimes committed by autonomous agents if it is willing to accommodate strict liability and other faultless models of liability that have so far been anathema to international criminal justice.

**Keywords:** Electronic liability; international crimes; artificial intelligence

## A. Introduction

Before us sat the ultimate plaything, the dream of ages, the triumph of humanism—or its angel of death.<sup>1</sup>

The early skeptics who feared that the development of autonomous agents could have highly destructive consequences have not been unduly apocalyptic. Increasingly, it is acknowledged that non-human autonomous agents can commit not only garden-variety domestic crimes but also the most serious international crimes.

The most well-known examples of autonomous agents committing crimes are drones killing civilians during armed conflict. There is a growing literature on the damage that drones and other autonomous weapons can inflict on civilians and other non-military targets.<sup>2</sup> But there is increasing recognition that autonomous agents can also commit other international crimes, including

<sup>1</sup>IAN McEWAN, *MACHINES LIKE ME* 3 (2019).

<sup>2</sup>See generally Frédéric Mégret, *The Humanitarian Problem with Drones*, 5 UTAH L. REV. 1283 (2013); Meredith Hagger & Tim McCormack, *Regulating the Use of Unmanned Combat Vehicles: Are General Principles of International Humanitarian*

some of the most paradigmatic of international crimes namely genocide and torture. Robots are employed for interrogation purposes which means they can inflict torture. And bots are used to spread hate speech and intensify discrimination which have triggered ethnic violence and genocide.<sup>3</sup>

But can a human be held accountable for the autonomous actions of a non-human machine? More complex yet: can an autonomous agent be held accountable for its actions? This article will consider the consequences of the ability of AI to commit international crimes and what the appropriate model of liability could be. It will consider these questions against the backdrop of the fundamental legal constraint that only natural persons have legal personality before the International Criminal Court (ICC). It will proceed from the view that it is not sufficient to hold the persons or programmers behind the autonomous agents liable but that it should be possible to hold the autonomous agents that commit international crimes liable.

The reference to “machines” in this article should be understood to include autonomous “agents” which are semi or fully autonomous. The article acknowledges the various degrees to which actors can be autonomous or non-autonomous. It will be argued that whereas it is easier to attribute liability to autonomous agents that are controlled, to a greater or smaller extent, by humans, it is possible to construct a framework for the liability of machines that are able to also operate in a “fully” autonomous way, in other words without human intervention.

The article will start by considering the kinds of international crimes that are committed by machines and then ask whether it follows that machines can be held accountable and if so what form liability should take given the current accountability gap. It will be argued that the requirement of fault does not present an unsurmountable obstacle to finding liability in international criminal law (ICL). It will be suggested that strict liability could form an acceptable basis for holding machines accountable for international crimes. Strict liability, it will be argued, can serve some of the same purposes of international criminal law that are served by fault liability.

Awarding legal personality to machines and then attributing strict liability to machines in the context of international crimes will necessitate a rethinking of the centrality of individual criminal liability and the insistence on human agency in international criminal law. It will also require a rethinking of the purposes of international criminal law and the extent to which the preoccupation with the individual continues to serve these purposes. This holds dramatic consequences for legal subjectivity. It will be argued that if there is a continuum of potential subjects of ICL, then the argument for electronic personhood and liability of machines is as compelling as for other non-humans such as corporate entities and animals. Because autonomous warfare represents a paradigm shift from traditional forms of warfare, it needs a legal approach that is qualitatively different from what existed in the past, also with regard to how international criminal law accommodates such warfare.

## B. Definitions

The definition of artificial intelligence (AI) has not remained stable.<sup>4</sup> The definition has changed and evolved multiple times. A definition formulated in 2018 holds that artificial intelligence is “a branch of computer science focused on programming machines to perform tasks that replicate or augment aspects of human cognition.”<sup>5</sup> Allen defines AI as the “ability of a computer system to

---

*Law Sufficient?*, 21 J.L. INFO. & SCI. 1 (2012); Cesáreo Gutiérrez Espada & María José Cervell Hortal, *Autonomous Weapons Systems, Drones and International Law*, 2 REVISTA DEL INSTITUTO ESPAÑOL DE ESTUDIOS DE ESTRATÉGICOS 1 (2013).

<sup>3</sup>See Joshua Uyheng & Kathleen M. Carley, *Bots and Online Hate During the Covid-19 Pandemic: Case Studies in the United States and the Philippines*, 3 J. COMPUTATIONAL SOC. SCI. 445 (2020).

<sup>4</sup>STUART RUSSELL & PETER NORVIG, *ARTIFICIAL INTELLIGENCE: A MODERN APPROACH* (2016).

<sup>5</sup>Dan Coats, *The AIM Initiative: A Strategy for Augmenting Intelligence Using Machines* (2018), <https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf>.

solve problems and to perform tasks that would otherwise require human intelligence.”<sup>6</sup> A House of Lords Select Committee on Artificial Intelligence defined AI as “technologies with the ability to perform tasks that would otherwise require human intelligence.”<sup>7</sup>

The absence of a single, agreed-upon definition of AI does not necessarily obstruct one from holding AI liable for committing or aiding and abetting international crimes. Most definitions of AI have certain common elements such as the feature that the intelligence of AI actors resemble human intelligence.<sup>8</sup> The variance in definition does not affect the key questions of the criminal accountability of AI because all AI actors, regardless of definition, lack the key element of fault or *mens rea* for example. Regardless of the definition of AI one chooses, one would not be able to insist on the traditional elements of criminal law and one would have to construct a new model of electronic liability.

There is also no consensus on a single definition for the term “autonomous,” which is an element of most definitions of AI. In some cases, autonomy is assumed to mean the ability of a system to operate successfully without human intervention. Christof Heyns, former Special Rapporteur on extrajudicial, summary, or arbitrary executions, described autonomous weapons as weapons “that, once activated, can select and engage targets without further human intervention.”<sup>9</sup> But stating that it operates “without human intervention” does not fully capture the nuance and complexity of autonomy. In other cases, autonomy is conflated with the lack of human control. The French definition qualifies human intervention as “human supervision, meaning there is absolutely no link (communication or control) with the military chain of command.”<sup>10</sup> The French definition of autonomy still caters for the possible presence of human involvement: under this definition, autonomous weapons may even be deployed under some form of human control but essentially operates independently from humans.

### C. How AI Commits International Crimes

Autonomous agents can commit core international crimes of genocide, war crimes, crimes against humanity and the crime of aggression (in other words, international crimes over which the ICC has subject matter jurisdiction)<sup>11</sup> as well as non-core international crimes such as human trafficking. Antonio Cassese describes the core crimes as “a category comprising the most heinous offences.”<sup>12</sup>

The crimes committed by Autonomous Weapons Systems (AWS) have received a good deal of attention. AWS is understood as any *weapon system* that can independently select and attack targets, including some existing weapons.<sup>13</sup> AWS can commit war crimes such as the indiscriminate killing of civilians. Less attention has been paid to how the war crime of torture and other international crimes can be committed by autonomous agents. Autonomous agents have committed human trafficking as well as selling, buying and possessing banned drugs.<sup>14</sup>

<sup>6</sup>Greg Allen, *Understanding AI Technology*, JOINT ARTIFICIAL INTELLIGENCE CENTER, Apr. 2020, at 5.

<sup>7</sup>SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE, HOUSE OF LORDS, AI IN THE UK: READY, WILLING AND ABLE? (2018).

<sup>8</sup>Bernard Marr, *The Key Definitions of Artificial Intelligence*, FORBES (Feb. 14, 2018), <https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/?sh=5d57eb294f5d>.

<sup>9</sup>Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Execution*, U.N. Doc. A/HRC/23/47 (Apr. 9, 2013).

<sup>10</sup>République Française (2016). Working paper of France: “Characterization of A Laws”. In *Meeting of experts on lethal autonomous weapons systems (LAWS)* (on file with author)

<sup>11</sup>See Rome Statute of the International Criminal Court art. 5 (1998); Christine Schwöbel-Patel, *The Core Crimes of International Criminal Law*, in THE OXFORD HANDBOOK OF INTERNATIONAL CRIMINAL LAW (2020) (providing a critical view).

<sup>12</sup>ANTONIO CASSESE, INTERNATIONAL CRIMINAL LAW 148 (2008).

<sup>13</sup>Neil Davison, *A Legal Perspective: Autonomous Weapon Systems Under International Humanitarian Law*, 30 UNODA OCCASIONAL PAPERS 5 (2018).

<sup>14</sup>JOHN FREDERICK ARCHBOLD, CRIMINAL PLEADING, EVIDENCE AND PRACTICE (2018).

Autonomous agents can be used to commit various of the core international crimes. Most of the literature has focused on the dangers that AWS can inflict in wartime or conflict. But autonomous agents can also “commit” international crimes outside the context of armed conflict. From the perspective of accountability, the ways in which the crimes of torture and genocide can be committed by autonomous agents are particularly interesting and important. By way of example, two contexts in which the actions of autonomous agents can become relevant under international criminal law are outlined in turn.

### I. Torture

To start with, AI can potentially commit torture. This can occur either physically or psychologically. Autonomous agents can be said to “commit” torture when AI developers integrate AI capabilities into an interrogation system. Thomasen defines a robot interrogator as “any automated technology that examines an individual through questioning . . . for purposes of eliciting incriminating statements or confessions.”<sup>15</sup> McAllister describes the environment that makes torture by “robots” or autonomous systems possible: “increasing developments in human-computer interaction, research and physiological measurement devices, combined with the inability of humans to act as effective lie detectors and the traditional reliance on technology to enhance human capacities.”<sup>16</sup> Automated deception detectors have indeed been used for some time in the form of prototype robotic guards, for example, for border control in the United States.<sup>17</sup> McAllister, describing the reasons governments find autonomous interviewers attractive, stated, “Indeed, autonomous interviewers would permit governments and agencies reliable, low-intrusive, and cost-effective means of interviewing large groups of people quickly or detecting weapons at common security points.”<sup>18</sup>

According to Thomasen, the robot interrogator is expected to be “a tireless, bias-free, faster, more effective and more accurate version of a human interrogator.”<sup>19</sup> Because sensor technologies can be streamlined into the interrogator, it is easier to conceal their ability to search suspects.<sup>20</sup> Using AI to inflict torture is further attractive because the deployer of the AI may be able to detach themselves both emotionally and physically.<sup>21</sup> The fact that the deployer can distance himself or herself physically means that they will not be performing the *actus reus* under current definitions of torture which requires the victim to be under the control or in the custody of the torturer.<sup>22</sup>

As robots cannot experience pain or empathy and cannot feel compassion, the mere presence of an interrogation robot may cause the subject of interrogation to talk out of fear. What further adds to the effectiveness of robot interrogation is that the interrogatee knows that the robot cannot understand pain or experience empathy and is therefore unlikely to act with mercy and stop the interrogation.<sup>23</sup> But this does not mean that robots cannot be programmed to emulate human traits. What makes robots suitable interrogators is that they have the capacity to use personality traits such as flattery and intimidation to manipulate the interlocutor.<sup>24</sup>

<sup>15</sup>Kristen Thomasen, *Liar Liar Pants on Fire! Examining the Constitutionality of Enhanced Robo-Interrogation 2* (Working Paper, 2012). See also *id.* at 1, n. 2 (“Humans are generally poor lie detectors (not usually accurate above 60%).”).

<sup>16</sup>Amanda McAllister, *Stranger than Science Fiction: The Rise of AI Interrogation in the Dawn of Autonomous Robots and the Need for an Additional Protocol to the UN Convention Against Torture*, 101 MINN. L. REV. 2540 (2017).

<sup>17</sup>Jordan Pearson, *The CIA Used Artificial Intelligence to Interrogate Its Own Agents in the 80’s*, VICE (Sept. 22, 2014), <https://www.vice.com/en/article/qkvz85/the-cia-used-artificial-intelligence-to-interrogate-its-own-agents-in-the-80s> (explaining that the CIA already experimented with the first robot interrogator in 1983).

<sup>18</sup>*Id.*

<sup>19</sup>Kristen Thomasen, *Examining the Constitutionality of Robot Enhanced Interrogation*, in ROBOT LAW (2016).

<sup>20</sup>*Id.*

<sup>21</sup>McAllister, *supra* note 17.

<sup>22</sup>Rome Statute of the International Criminal Court art. 7(2)(e) (1998).

<sup>23</sup>*Id.* at arts. 19–20.

<sup>24</sup>Thomasen, *supra* note 20.

The idea of an autonomous robot in an interrogative space which self-sufficiently conducts an interrogation invokes a multitude of questions, including the question of what constitutes lawful interrogation, torture, and torture or cruel, degrading and inhuman treatment.<sup>25</sup> Further, AI actors can be the direct perpetrators of international and transnational crimes. AI can also be used as a tool by humans when AI aids and abets criminal activities. In these cases, examples of AI aiding and abetting criminal activities include smuggling (by using unmanned vehicles for example), as well as torture, sexual offences, fraud and theft. As Sehrawat puts it: “[T]hey can be held responsible under different modes of liability, such as for attempting, assisting, facilitating, aiding, abetting, planning, or instigating the commission of a war crime.”<sup>26</sup>

## II. Genocide

The second example of an international crime that might be committed, or assisted, by autonomous agents is genocide. AI tools used on social media can significantly exacerbate the stages of genocide. The first stage of genocide pertains to classifying groups,<sup>27</sup> that is the creation of an “us” versus “them” dichotomy between people of different race, ethnicity, religion or nationality. Such divisions may escalate into removing or denying a group’s citizenship and stripping them of their civil and human rights.<sup>28</sup> AI can exacerbate such divisions by facilitating and accelerating the spread of hatred and deepfakes. Through the use of bots, the content of social media messages can be tailored specifically to vilify targeted groups.<sup>29</sup>

The case of international crimes committed against the Uighurs is an example of humans using technology to commit genocide. Millions of Uighurs are currently experiencing the consequences of the misuse of AI-technologies. China has used an intrusive surveillance smartphone application with the name “Integrated joint operations platform” in order to track Uighurs.<sup>30</sup> In the region of Xinjiang, machines developed with AI-technologies are able to check and scan IDs at checkpoints and alert the police or military in case surveillance cameras detect what Chinese officials call “suspicious activity.”<sup>31</sup> The use of facial recognition technology has also contributed to China’s ability to target Uighurs.

A further example is the genocide in Myanmar. Some believe that Facebook exacerbated this genocide.<sup>32</sup> Members of the Myanmar military used Facebook to systematically target the country’s Muslim minority, the Rohingya. In the view of human rights groups, the anti-Rohingya propaganda incited large scale murder and rape.<sup>33</sup> For example, a Facebook post posted in 2018 showed a photograph of a boatload of Rohingya refugees with the words: “Pour fuel and

<sup>25</sup>McAllister, *supra* note 17, at 2563.

<sup>26</sup>Vivek Sehrawat, *Autonomous Weapons System and Command Responsibility*, 31 FLA. J. INT’L L. 315, 316 (2021).

<sup>27</sup>*What was the Holocaust?*, THE WIENER HOLOCAUST LIBRARY, <https://www.theholocaustexplained.org/what-was-the-holocaust/what-was-genocide/eight-stages-of-genocide/>. Scholars have approached the question of the identification of different levels of genocide in different ways. See generally Shelley Burleson *Spatiality of the Stages of Genocide: The Armenian Case*, 10 GENOCIDE STUDIES & PREVENTION 39, 42 (2016); see also Sheri P. Rosenberg, *Genocide is a Process, Not an Event*, 7 GENOCIDE STUD. & PREVENTION 16 (2012).

<sup>28</sup>Gregory Stanton, *The Ten Stages of Genocide*, GENOCIDE WATCH (2016), <https://www.genocidewatch.com/tenstages>.

<sup>29</sup>Nuha Albadi, Maram Kurdi & Shivakant Mishra, *Hateful People or Hateful Bots? Detection and Characterization of Bots Spreading Religious Hatred in Arab Social Media*, 3 PROC. ACM HUM.-COMPUT. INTERACTION (2019).

<sup>30</sup>Marine Milard & Sophie Smith, *How AI Can Either Exacerbate or Prevent Genocides: Reflection Based on the 10 Stages of Genocide*, BUDAPEST CENTRE FOR MASS ATROCITIES PREVENTION (2021), [https://www.genocideprevention.eu/files/10\\_stages\\_AI.pdf](https://www.genocideprevention.eu/files/10_stages_AI.pdf).

<sup>31</sup>The Chinese government further employs a wide network of surveillance cameras using facial recognition. The tools to do so were created by the telecommunication giant, Huawei. This feature is capable of detecting the faces of a person from the Uighur minority and—most alarming—to alert Chinese officials in case of “unusual behaviour” or if a person goes beyond a certain authorised-area.

<sup>32</sup>Paul Mozur, *A Genocide Incited on Facebook, with Posts from Myanmar’s Military*, N.Y. TIMES (Oct. 15, 2018), <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>.

<sup>33</sup>*Id.*

set fire so that they can meet Allah faster.”<sup>34</sup> A report by the UN’s Independent International Fact-Finding Mission on Myanmar that was established to make recommendations on genocide and other crimes committed in Myanmar concluded that Facebook has been a useful instrument for those seeking to spread hate against the Rohingya.<sup>35</sup>

#### D. The Need for Electronic Liability

The need to construct a model of electronic liability is not self-evident. In some cases where autonomous agents or machines directly or indirectly commit international crimes or aid and abet such crimes, it will be clear that there is an identifiable human being “behind” the machine, operating or programming the machine’s actions either directly or indirectly. However, it will more often be the case that the person behind the machine is not identifiable or very difficult to identify. The identity of the perpetrator is likely to be deliberately hidden or obfuscated. As a result, the human actors will escape international criminal liability.

Further, an approach that attaches liability to the human behind the machine is also not realistic from the point of view of the current stage of development of autonomous agents and weapons. Machines or systems of machines have already reached the level of autonomy that they can identify an enemy target and take lethal action without human input.<sup>36</sup> Examples of such autonomous machines include aerial vehicles, submersible vehicles as well as ground-based vehicles with an attached lethal weapon.<sup>37</sup> And drones are theoretically able to be controlled autonomously or by a human controller.

This creates an accountability gap that can be filled by constructing electronic liability. This will prevent the perpetrators of crimes from escaping liability by relying on the anthropocentric features of criminal law. It is important not to create incentives for persons to hide behind autonomous agents or for persons to use autonomous agents to hide the true perpetrators of crimes. Electronic liability will allow the “unseen” manufacturer or programmer of an autonomous agent to still pay the price or suffer the consequences of the actions of the machine or autonomous agent. However, it should be noted that imposing electronic liability will not mean that the persons directing the machine, to the extent that they can be identified, cannot or should not be found liable *in addition* to the machine. The imposition of electronic liability for the commission of international crimes through machines will simply help fill the current accountability gaps in the law.

In addition, electronic liability will prevent perpetrators from escaping liability by relying on narrow definitions of crimes. In the case of torture for example, the *actus reus* requires that the torturer cannot detach themselves physically from the person being tortured. The Rome Statute defines torture as, “[T]he intentional infliction of severe pain or suffering, whether physical or mental, upon a person in the custody or under the control of the accused; except that torture shall not include pain or suffering arising only from, inherent in or incidental to lawful sanctions.”<sup>38</sup>

<sup>34</sup>Dan Milmo, *Rohingya Sue Facebook for £150bn over Myanmar Genocide*, GUARDIAN (Dec. 6, 2021).

<sup>35</sup>Human Rights Council, *Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar*, at 323, U.N. Doc. A/HRC/39/CRP.2 (Sept. 28, 2018) (“For example, the Mission encountered over 150 online public social media accounts, pages and groups that have regularly spread messages amounting to hate speech against Muslims in general or Rohingya in particular. Given Facebook’s dominance in Myanmar, the Mission paid specific attention to a number of Facebook accounts that appear to be particularly influential considering the number of followers . . .”).

<sup>36</sup>Drew Charters, *Killing on Instinct: A Defense of Autonomous Weapon Systems for Offensive Combat*, 4 VITERBI CONVERSATIONS IN ETHICS (May 19, 2020), <https://vce.usc.edu/volume-4-issue-1/killing-on-instinct-a-defense-of-autonomous-weapon-systems-for-offensive-combat/>.

<sup>37</sup>*Id.*

<sup>38</sup>Rome Statute of the International Criminal Court art. 7(2)(e) (1998). See generally M. CHERIF BASSIOUNI, THE STATUTE OF THE INTERNATIONAL CRIMINAL COURT (1998).

The current definition of torture therefore requires the torture victim to be under the control or custody of the accused, implying physical proximity, and does not allow for torturing someone from a remote location.<sup>39</sup> Using machines to torture therefore makes it easier for torturers to obfuscate liability and circumvent the law. Former UN Special Rapporteur on Torture Nils Melzer has remarked that many countries have invested “significant resources towards developing methods of torture which can achieve purposes of coercion, intimidation, punishment, humiliation or discrimination without causing readily identifiable physical harm or traces.”<sup>40</sup>

## E. Overcoming the Obstacles to AI Criminal Liability

The most significant obstacles for electronic liability to be recognized are the following: First, the idea that international criminal law is—only—based on individual criminal responsibility (and that for the purpose of individual criminal responsibility, individuals are understood to be humans); second, the fact that machines are not considered to possess legal personality; and third, the requirement of fault in finding criminal liability. This section shows how these obstacles can be overcome.

### I. Individual Criminal Responsibility

One of the most significant contributions the discipline of international criminal law has made to the field of international law generally was the attribution of criminal liability to individuals. Whereas international courts such as the International Court of Justice has attributed liability exclusively to states, the *ad hoc* international criminal tribunals revolutionised criminal responsibility in international law by not only allowing individual criminal liability but *basing* its jurisdiction on individual criminal liability.

As a result, humans are still the measure of all things in international criminal law. According to Lostal, the ICC has “clearly espoused the human exceptionalism theory as shown by the contextual, systemic interpretation and subsequent practice around the notion of ‘person.’”<sup>41</sup> For example, when the preamble of the Rome Statute mentions victims, it does so by making reference to “children, women and men.”<sup>42</sup> According to Lostal, other norms of the legal framework refer to persons as a shorthand to indicate “human being.” Article 1 of the Statute claims that the Court “shall have the power to exercise its jurisdiction over persons for the most serious crimes of international concern”; Article 26 refers to jurisdiction over “persons”; and Rule 123 RPE relates to measures to ensure the presence of the “person” concerned at the confirmation hearing. The term “persons” is also omnipresent in the Rome Statute’s definitions of crimes.

Prosperi and Terrosi have similarly described existing international criminal law as “essentially anthropocentric.”<sup>43</sup> It can be said that international criminal law *defines itself* in terms of individual criminal responsibility. Article 25 of the Rome Statute states that the court shall have jurisdiction over natural persons and sets out the circumstances under which an individual shall be held responsible. Article 25(2) states: “A person who commits a crime within the jurisdiction of the Court shall be individually responsible and liable for punishment in accordance with this Statute.” Individual criminal responsibility is applicable where an individual directly commits

<sup>39</sup>See Owen Bowcott, *UN Warns of the Rise of “Cybertorture” to Bypass Physical Ban*, GUARDIAN (Feb. 21, 2020), <https://www.theguardian.com/law/2020/feb/21/un-rapporteur-warns-of-rise-of-cybertorture-to-bypass-physical-ban> (discussing the inadequacies of the current definition of torture).

<sup>40</sup>*Id.*

<sup>41</sup>Marina Lostal, *De-objectifying Animals: Could They Qualify as Victims Before the International Criminal Court?*, 19 J. INT’L CRIM. JUST. 583, 592 (2021).

<sup>42</sup>Rome Statute of the International Criminal Court pmb. (1998).

<sup>43</sup>Luigi Prosperi & Jacopo Terrosi, *Embracing the “Human Factor”: Is There New Impetus at the ICC for Conceiving and Prioritizing Intentional Environmental Harms as Crimes Against Humanity?*, 15 J. INT’L CRIM. JUST. 509, 510 (2017).

a crime or directly contributes to it through ordering, planning, instigating, inciting, co-perpetration, joint criminal enterprise, aiding and abetting.<sup>44</sup>

However, although the notion of “person” has so far been interpreted as “human person,” the wording of these provisions of the Rome Statute would be open to a more inclusive interpretation. As person is, in other (international) legal contexts, often interpreted as including both natural and legal persons, it would be possible also to include artificial persons. The anthropocentrism of international criminal law could thus be overcome by adopting a broader interpretation of already existing norms. However, given the dramatic implications of extending this notion to AI actors, the Rome Statute should be amended to include AI explicitly.

## II. Legal Personality

Even if the anthropocentrism of international criminal law can be overcome, liability can only be attributed to machines once machines have been assigned legal personality. Although non-humans do not currently possess legal personality before the ICC or most domestic jurisdictions, the concept of legal personality has been altered many times throughout history to keep up with societal developments. Andrew Clapham reminds us that not that long ago a doctrinal debate was raging over the question whether individuals can be the subject of international law.<sup>45</sup>

The proposal to recognize legal personality for machines or “electronic personhood” has been controversial. Such legal personality would mean that particular highly sophisticated robots and software agents are the addressees of legal duties and obligations and the holders of legal rights.<sup>46</sup> The debate has been fuelled by a resolution adopted by the European Parliament on Civil Law Rules on Robotics. This resolution made recommendations to the European Commission according to which, in the long run, legislators should award legal personality to some very advanced AI systems.<sup>47</sup> The underlying idea is that if it is increasingly difficult to trace harm triggered by AI back to any kind of human behaviour, it would follow that one can hold AI itself liable.<sup>48</sup>

The idea that machines can hold legal personality has been heavily criticized. Much of the criticism and resistance has its roots in ethical considerations. It is believed that an “attempt to put machines on an equal footing with human beings and afford them the same or similar rights is apt to help delude the fundamental difference between human beings and things.”<sup>49</sup>

One’s willingness to grant legal personality to autonomous agents will therefore largely depend on how one views the “person” in legal personhood. Those who believe that legal personality consists of nothing more than the formal capacity to bear a legal right<sup>50</sup> will be more comfortable to bestow legal personality on autonomous agents than those who think that there is always a necessary connection between moral and legal persons and who attribute metaphysical qualities to personhood.<sup>51</sup>

The definition of legal personality that separates morality from subjectivity is the most inclusive definition of personhood in that it can, potentially, be all-embracing. As Naffine argues, “anything can be a legal person because legal persons are stipulated as such or defined into existence.”<sup>52</sup>

<sup>44</sup>Bert Swart, *Modes of international Criminal Liability*, in *THE OXFORD COMPANION TO INTERNATIONAL CRIMINAL JUSTICE* (Antonio Cassese ed., 2009).

<sup>45</sup>Andrew Clapham, *The Role of the Individual in International Law*, 21 *EUR. J. INT’L L.* 25, 28 (2010). See also ASTRID KJELDGAARD-PEDERSEN, *THE INTERNATIONAL LEGAL PERSONALITY OF THE INDIVIDUAL* 139 (2018) (challenging the states-only conception of legal personality under international law).

<sup>46</sup>Christiane Wendehorst, *Strict Liability for AI and Other Emerging Technologies*, 11 *J. EUR. TORT L.* 150, 155 (2020).

<sup>47</sup>European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics, Feb. 16, 2017, 2017 O.J. (C 252).

<sup>48</sup>Wendehorst, *supra* note 47, at 155.

<sup>49</sup>*Id.* at 156.

<sup>50</sup>Ngairé Naffine, *Who are Law’s Persons? From Cheshire Cats to Responsible Subjects*, 66 *MOD. L. REV.* 346, 350 (2003).

<sup>51</sup>*Id.* at 366.

<sup>52</sup>*Id.* at 351.



According to this understanding of personhood, legal persons can include animals, fetuses, the dead, the environment, corporations, or whatever else the law accepts into the community of persons.<sup>53</sup>

According to Novelli, Bongiovanni and Sartor, many authors support the idea of conferring legal personality on AI. They mention not only the advanced cognitive abilities of AI actors but also other “new elements” introduced by AI:

Yet what seem to characterize AIs, to the point of introducing new elements into the debate on legal personality, are the sociotechnical profiles resulting from the deployment of artificial intelligence agents, e.g., the marked unpredictability of their decision-making processes and the impact (both positive and negative) that these processes may have on people’s lives, on society, and on the market; the ability of such systems to communicate and network; the involvement of different human players in the production and implementation of such systems, each with different potential responsibilities; and the difficulty, sometimes the impossibility, of tracking the relevant human players.<sup>54</sup>

In some domestic jurisdictions such as New Zealand, there have been developments relating to awarding legal personality to non-humans such as lands and rivers.<sup>55</sup> And in the context of animal rights, it has been argued that acknowledging nonhuman species as possessing limited legal personality.<sup>56</sup> Awarding legal personality to animals is extremely rare but not unheard of. In 2015, a judge in Argentina awarded legal personality to an Orangutan.<sup>57</sup> But US courts have not been similarly adventurous. In June 2022, a US court refused to recognize the legal personality of an elephant called Happy, illustrating that the law on whether animals enjoy personality is still inconsistent.<sup>58</sup> These developments show that, despite existing controversies, the legal concept of personhood is currently broadening. Attributing legal personality to AI would thus be in line with this trend.

### III. Fault as Requirement for Liability

The third apparent obstacle to AI criminal responsibility is the fault requirement for liability. Criminal liability requires both a physical element (*actus reus*) and a mental element (*mens rea*). Criminal law requires the presence of *mens rea* before it can be said that a crime has been committed.<sup>59</sup> *Mens rea* refers to a blameworthy mental state and can consist of intention or negligence. It is this mental element that presents the sticking point in terms of holding machines accountable under international criminal law.

To overcome this difficulty, the concept of strict liability constitutes one of the possible legal frameworks for AI accountability. Strict liability has been defined as “crimes which do not require

<sup>53</sup>*Id.*

<sup>54</sup>Novelli, Bongiovanni & Sartor, *infra* note 94, at 200.

<sup>55</sup>See Devon O’Neil, *Parks are People Too*, OUTSIDE (Aug. 3, 2016), <https://www.outsideonline.com/2102536/parks-are-people-too>; Bryant Rousseau, *In New Zealand, Lands and Rivers Can Be People (Legally Speaking)*, N.Y. TIMES (July 13, 2016), <https://www.nytimes.com/2016/07/14/world/what-in-the-world/in-new-zealand-lands-and-rivers-can-be-people-legally-speaking.html>.

<sup>56</sup>Veerle Platvoet, *The Attribution of Limited Legal Personality to Nonhuman Species*, 10 J. ANIMAL ETHICS 49 (2020).

<sup>57</sup>*Orangutan Sandra Granted Personhood Settles into New Florida Home*, GUARDIAN, (Nov. 7, 2019), <https://www.theguardian.com/world/2019/nov/07/sandra-orangutan-florida-argentina-buenos-aires>.

<sup>58</sup>*Happy the Elephant is Not a Person, Says Court on Key US Animal Rights Case*, GUARDIAN (June 15, 2022), <https://www.theguardian.com/us-news/2022/jun/14/elephant-person-human-animal-rights-happy>.

<sup>59</sup>Rome Statute of the International Criminal Court art. 30 (1998) (defining the mental element).

*mens rea* or even negligence as to one or more elements in the *actus reus* are known as offences of strict liability.<sup>60</sup> As AI systems are not capable of meeting existing criminal law principles<sup>61</sup> and the requirements of both factual and mental elements in particular, many believe strict liability is the most appropriate liability model in this context. The next section outlines this possibility in more detail.

However, strict liability would be a novelty in the criminal law context. So far, there has been much resistance to strict liability as a basis for criminal liability.<sup>62</sup> In the domestic context, courts have argued that strict liability should never be the basis for retributive punishment, and that is also a weak basis for deterrence.<sup>63</sup> South African courts, for example, are hostile to strict liability and will only deviate from the principle of no liability without fault if there are clear and convincing indications.<sup>64</sup> Moreover, the more serious a crime, the less likely it is that domestic criminal law systems will allow for strict liability. As Grant states plainly, serious crimes require intention. Domestic systems such as the US thus quite comfortably attach strict liability to traffic offences,<sup>65</sup> but attaching strict liability to serious offences such as murder or homicide is highly controversial.

Because of the centrality of fault in our understanding of crime, some argue that the absence of fault means that machines are not capable of committing crimes, but that terms such as “mal-functioning” should be used instead.<sup>66</sup> But, as McAllister writes, to only attach liability to those things or beings philosophically capable of intent would defeat the state parties’ original intent in drafting and adopting the CAT: to prohibit all torture and cruel, inhuman and degrading treatment, not merely torture by those things.<sup>67</sup> This argument applies not only to torture but to international crimes more generally.

## F. Models of Electronic Liability

There are two potential models of electronic liability: strict liability and command responsibility. Both are addressed in turn.

### I. Strict liability

Strict liability does not fit easily into traditional understandings of international criminal law. In the *Bemba* case, the ICC Pre-Trial Chamber went as far as stating that the Rome Statute disapproves of strict liability. When the Pretrial Chamber examined the requirement that “the suspect either knew or, owing to the circumstances at the time, should have known” about the relevant crimes, the Chamber stated that “the Rome Statute does not endorse the concept of strict liability,”<sup>68</sup> meaning that “attribution of criminal responsibility for any of the crimes that fall within the jurisdiction of the Court depends on the existence of the relevant state of mind or degree of

<sup>60</sup>DAVID ORMEROD & KARL LAIRD, SMITH, HOGAN, AND ORMEROD’S CRIMINAL LAW 146 (2021).

<sup>61</sup>Nora Osmani, *The Complexity of Criminal Liability of AI Systems*, 14 MASARYK UNIV. J.L. & TECH. 53, 57 (2020).

<sup>62</sup>See Larry Alexander, *Is There a Case for Strict Liability?*, 12 CRIM. L. & PHIL. 531 (2018).

<sup>63</sup>*S v. Arenstein* 1964 (1) SA 361 (A); *S v. Qumbella* 1966 (4) SA 256 (A).

<sup>64</sup>In *S v. Qumbella*, the Court said: “The legislature must make strict liability appear plainly.” Indeed, the Appellate Division has even set up the requirement of fault as a presumption. *S v. Qumbella* 1966 (4) SA 256 (A). In *S v. Arenstein*, the court stated: “the general rule is that *actus non facit reum nisi mens sit rea*, and that in construing statutory prohibitions or injunctions, the legislature is presumed, in the absence of clear and convincing indications to the contrary not to have intended innocent violations thereof to be punishable.” *S v. Arenstein* 1964 (1) SA 361 (A)

<sup>65</sup>Noah Kazis, *Tort Concepts in Traffic Crimes*, 125 YALE L.J. 1131 (2016).

<sup>66</sup>Xavier J. Ramírez García de León, *Requirement of Mens Rea for War Crimes in the Light of the Development of Autonomous Weapons Systems*, 21 ANUARIO MEXICANO DE DERECHO INTERNACIONAL 442 (2021).

<sup>67</sup>McAllister, *supra* note 17, at 2557.

<sup>68</sup>*Prosecutor v Bemba*, ICC-01/05-01/08, Pre-Trial Chamber II, Decision on the Confirmation of Charges, 15 June 2009, ¶ 427.

fault.”<sup>69</sup> This statement by the ICC can be interpreted as meaning that *individuals* will not be held strictly liable. In contrast, due to its individual-centredness of the Rome Statute, the court has so far been silent on how to deal with the liability of non-humans such as machines.

Fault-based liability is often retributory in its aims. Because strict liability cannot result in retribution, it is considered not fitting in the context of international criminal law which adopts, at least in part, a model of retributive justice. But retributive justice is just one of various models of justice, and retribution just one of various purposes served by international criminal justice. Apart from retribution, international criminal law is also believed to serve the purposes of deterrence, promoting peace and security, strengthening accountability, creating a historical record, and truth-telling.<sup>70</sup> Victims’ participation and victims’ protection can be added to the traditional list of purposes.<sup>71</sup> Strict liability can serve many of these purposes as well as fault-based liability does. And in some cases, the policy considerations that will be served by electronic liability, such as enhancing public trust, promoting legal certainty and risk control,<sup>72</sup> are as important as the more traditional purposes, such as deterrence.

A key benefit of introducing strict liability for AI would be its strong “symbolic” value and the fact that it is likely to both enhance public trust in the mass roll-out of AI and to put an end to legal uncertainty.<sup>73</sup> Further, although controversial,<sup>74</sup> the goal of deterrence is already a well-established goal of international criminal law, and strict liability can promote deterrence as much as fault liability can. In domestic law, deterrence as public policy consideration has already been accepted as justification for applying a strict liability standard. In the context of product liability, courts have confirmed that public policy demands that liability for loss should fall or be placed where it will most effectively deter such loss from recurring. In the landmark US decision on product liability, *Escola v. Coca-Cola Bottling Co.*, the court stated:

Even if there is no negligence, however, public policy demands that responsibility be fixed wherever it will most effectively reduce the hazards to life and health inherent in defective products that reach the market. It is evident that the manufacturer can anticipate some hazards and guard against the recurrence of others, as the public cannot. Those who suffer injury from defective products are unprepared to meet its consequences. The cost of an injury and the loss of time or health may be an overwhelming misfortune to the person injured, and a needless one, for the risk of injury can be insured by the manufacturer and distributed among the public as a cost of doing business.<sup>75</sup>

In the context of autonomous agents with the potential of committing international crimes, Zech argues that strict liability can be an instrument for risk distribution.<sup>76</sup> The risk lies with the injurer. The risk controller must consider whether the expected benefit of an activity exceeds its risk.<sup>77</sup> Social media companies such as Facebook and Twitter that allow the proliferation of hateful content should be held liable rather than the end user because the end user is not able to successfully predict and protect himself or herself against the harm.

<sup>69</sup>*Id.*

<sup>70</sup>See Mark Klamberg, *What are the Objectives of International Criminal Procedure?*, 79 NORDIC J. INT’L L. 279 (2010).

<sup>71</sup>RICHARD MAY & MARIEKE WIERDA, *INTERNATIONAL CRIMINAL EVIDENCE* 17 (2002). See also Mirjan Damaška, *What is the Point of International Criminal Justice?*, 83 CHI.-KENT L. REV. 329 (2008).

<sup>72</sup>Herbert Zech, *Liability for AI: Public Policy Considerations*, 22 ERA FORUM (2021).

<sup>73</sup>Wendehorst, *supra* note 47, at 178.

<sup>74</sup>Many scholars have expressed criticism of deterrence as an effective aim of international criminal law. See e.g., John Dietrich, *The Limited Prospects of Deterrence by the International Criminal Court: Lessons from Domestic Experience*, 88 INT’L. SOC. SCI. REV. 1 (2014).

<sup>75</sup>*Escola v. Coca-Cola Bottling Co.*, 24 Cal. 2d 453, 462 (1944).

<sup>76</sup>Zech, *supra* note 73, at 152.

<sup>77</sup>*Id.*

In the context of damage caused by machines where fault is not a useful construct in finding liability, liability will nevertheless be hooked onto causation. As Weinrib writes, under strict liability, causation is decisive to a defendant's liability.<sup>78</sup> Essentially, the requirement of fault falls away, and causation becomes a more important requirement. The requirement of causation prevents strict liability from running rampant. It acts as a check or limitation on strict liability.

A strict liability approach would solve many of the problems attached to searching for human agents "behind" autonomous agents and holding them accountable. An approach that holds those who design, program, or create autonomous agents liable is not necessarily just. The complexity of the autonomous agent's programming could make it possible that the designer, developer, or deployer would neither know nor be able to predict the AI's criminal act or omission.<sup>79</sup> For this reason, liability should not rest on knowledge or intent because it might create an incentive for human agents to avoid finding out what exactly the machine learning system is doing.<sup>80</sup> It is also not true that robots will do only what they are programmed to do. As Grut writes, "[P]rograms with millions of lines of code are written by teams of programmers, none of whom knows the entire program; hence, no individual can predict the effect of a given command with absolute certainty, since portions of large programs may interact in unexpected, untested ways . . . ."<sup>81</sup>

In addition, there remains the question of whether autonomous robots would even obey orders or be capable of recognizing a chain of command.<sup>82</sup>

A final concern with attributing strict liability to machines is finding an appropriate sanction. When a criminal defendant is deemed strictly liable in criminal law or in tort law, he or she may be ordered to pay compensatory damages.<sup>83</sup> Autonomous agents, like corporations, cannot be imprisoned, but they can be made to feel the brunt of any misconduct through a panoply of sanctions.<sup>84</sup> In the absence of the option of imprisonment, a finding of "electronic liability" can be punished by imposing a fine. Yet, some will find it difficult to comprehend how a monetary fine as an equitable punishment for a violation of a *jus cogens* norm.<sup>85</sup> Fines coupled with other remedies, such as guarantees of non-repetition, might thus be more appropriate than "mere" fines.

## II. Command Responsibility

In addition to strict liability, command responsibility can potentially provide a framework to address the accountability gap caused by AI. The model of command responsibility is included in Article 28 of the Rome Statute and employed in *Bemba*. Command responsibility rests on the presumption of the negligence of the commander(s) who authorized the deployment of an autonomous weapon that commits an illegal act.

Command responsibility extends to actions committed by the forces under the commander's "effective control."<sup>86</sup> While the superior's duty is not so much active, it is a kind of liability that arises from violating the duty to prevent illegal actions of a party, actions over which the superior

<sup>78</sup>Ernest J Weinrib, 'Causation and Liability', 63 CHICAGO-KENT LAW REVIEW (1987) 416.

<sup>79</sup>Thomas C. King, Nikita Aggarwal, Mariarosaria Taddeo & Luciano Floridi, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, 26 SCI. & ENG'G ETHICS 89, 95 (2020).

<sup>80</sup>*Id.* (quoting Rebecca Williams' presentation of Written Evidence to the House of Lords Select Committee on Artificial Intelligence (AIC0206) (2017)).

<sup>81</sup>Chantal Grut, *The Challenge of Autonomous Lethal Robotics to International Humanitarian Law*, 18 J. CONFLICT & SEC. L. 5, 20 (quoting Gary E. Marchant, Braden Allenby, Ronald Arkin, Edward T. Barrett, Jason Borenstein, Lyn M. Gaudet, Orde Kittrie, Patrick Lin, George R. Lucas, Richard O'Meara & Jared Silberman, *International Governance of Autonomous Military Robots*, 12 COLUM. SCI. & TECH. L. REV. 272, 283 (2011)).

<sup>82</sup>McAllister, *supra* note 17, at 2564.

<sup>83</sup>Alexander, *supra* note 63.

<sup>84</sup>*Id.* at 1118.

<sup>85</sup>McAllister, *supra* note 17, at 2563.

<sup>86</sup>Rome Statute of the International Criminal Court art. 28 (1998).

exercises professional control.<sup>87</sup> The vicarious criminal liability that results from command responsibility implicates a commander in many of the acts committed by subordinate forces that violate international law. In the context of the battlefield, the subordinate forces may include AWS in its arsenal of capabilities.<sup>88</sup> In this context, command liability amounts to operator liability.<sup>89</sup>

In addition, Corn has suggested that command responsibility could be extended to the procurement officials who bring AWS into a government's inventory. This approach would ensure that "decision-making officials and not technicians or legal advisers" and the ones who endorse the developing technological knowhow of AWS are the individuals who are held accountable should any unlawful outcomes result.<sup>90</sup>

Although applying command responsibility to the AI context is thus a feasible option, it remains doubtful as to whether such an approach would be an appropriate solution to the accountability gap. For example, Human Rights Watch has expressed concerns that it is "arguably unjust" to hold commanders to account for the action of machines "over which they could not have sufficient control."<sup>91</sup> Rather than command responsibility, strict liability appears to be the better option.

## G. Conclusion

The ICC will potentially only be able to meaningfully prosecute international crimes committed by autonomous agents if it is willing to accommodate strict liability and other faultless models of liability that have so far been anathema to international criminal justice. In order to do so, it would have to make the giant leap of moving away from fault as the central requirement for criminal liability. It would also have to open up its notion of legal subjectivity.

The Rome Statute follows the philosophy articulated at Nuremberg that men, not abstract entities, commit crimes against international law.<sup>92</sup> But the insistence on finding human agency behind international crimes will not serve the victims of drone attacks and international crimes committed by autonomous agents. Individuals will ultimately only be sufficiently or meaningfully protected if the legal personality is not only enjoyed by individuals or humans.<sup>93</sup>

Accepting and constructing "electronic liability" will necessitate an (undoubtedly seismic) shift away from the individual-centredness of ICL. But amending the ICC Statute is nothing new. It is necessary to now move to amend some of the provisions that have previously been considered foundational, even sacrosanct. According to David Luban, one of the legacies of Nuremberg was enlarging the reach of the law.<sup>94</sup> He writes that the lawmakers at Nuremberg "viewed their own words and deeds from the perspectives of a distant more pacific age."<sup>95</sup> It can be asked whether the drafters of the Rome Statute were similarly prescient and forward-looking when they restricted legal personality to natural persons.

<sup>87</sup>Guénaél Mettraux, *Command Responsibility in International Law – The Boundaries of Criminal Liability for Military Commanders and Civilian Leaders* (Jan. 2008) (Ph.D. thesis, London School of Economics) (on file with the London School of Economics).

<sup>88</sup>James Kraska, *Command Accountability for AI Weapons Systems in the Law of Armed Conflict*, 97 INT'L L. STUD. 407, 438 (2021).

<sup>89</sup>Swati Malik, *Autonomous Weapons Systems: The Possibility and Probability of Accountability*, 35 WIS. INT'L L. J. 609, 636.

<sup>90</sup>Geoffrey S. Corn, *Autonomous Weapon Systems: Managing the Inevitability of "Taking the Man out of the Loop,"* in AUTONOMOUS WEAPONS SYSTEMS: LAW, ETHICS, POLICY 209 (Nehal Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu & Claus Kreß eds., 2016).

<sup>91</sup>*Losing Humanity: The Case Against Killer Robots*, HUMAN RIGHTS WATCH (Nov. 19, 2012), <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>.

<sup>92</sup>*France v. Goering*, 22 I.M.T. 411, 466 (1946).

<sup>93</sup>See Claudio Novelli, Giorgio Bongiovanni & Giovanni Sartor, *A Conceptual Framework for Legal Personality and its Application to AI*, 13 JURISPRUDENCE 194 (2022).

<sup>94</sup>David Luban, *The Legacies of Nuremberg*, 54 SOC. RSCH. 779, 779 (1987).

<sup>95</sup>*Id.*

To accommodate electronic liability, the Rome Treaty should be amended to explicitly extend its personal jurisdiction to legal persons. Articles 1 and 25(1) should be amended to include legal persons.<sup>96</sup> This might require extensive amendments to the Rome Statute, but the alternative would be that the ICC becomes increasingly irrelevant when it comes to fighting impunity for the most serious crimes known to mankind.

If a thing without a tangible form, such as a corporation, can be a legal person, then it is no great conceptual leap to also confer legal personality on a thing that does have physical existence. The roots of the insistence to confine personality to individuals should be re-examined. It should be asked whether the purposes of ICL—both the normative principles and the more policy-oriented—are best served by rigidly clinging to individuals as the only subjects of ICL.

**Acknowledgement.** Professor Swart would like to thank the editors of this special volume for helpful comments.

**Competing Interests.** The author declares no competing interest.

**Funding Statement.** No specific funding has been declared in relation to this article.

---

<sup>96</sup>For more on ways in which to amend the Rome Statute, see Joe DelGrande, *Corporate Accountability: Prosecuting Corporations for the Commission of International Crimes of Atrocity*, 53 N.Y.U. J. INT'L L. & POLS. 144 (2021).