

Of Code and Consequence: Assessing the Impact of Artificial Intelligence on International Criminal Law Norms Governing the Direct and Public Incitement to Genocide

Naman Anand*

Abstract

The rise of artificial intelligence (AI) and its seamless integration with today's defense technologies and equipment raise concerns about how this may impact the application of international legal norms related to the direct and public incitement of genocide. This work elaborates on the potential difficulties resulting from the rise of AI vis-à-vis the enforcement of the Rome Statute, specifically focusing on Article 25(3)(e) and interconnected soft law instruments such as the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD) and the Genocide Convention. The work concentrates on how the Rome Statute, which was designed during (and catered to) a pre-digital era, falls short in addressing the complex challenges posed by AI-driven portals. These types of portals can be utilized to plan and amplify calls for genocide in a manner that is not fully protected by the existing legal framework.

By looking at case law and engaging in theoretical and speculative analysis, this work delves into the various ways in which AI and virtual private networks (VPNs) can influence the dissemination of genocidal speech—specifically, by prioritizing user engagement over truthfulness or accuracy, necessitating a relook at what constitutes “direct and public incitement” under international law. The work also raises the question of attribution and how the atrocities committed due to such incitements might be prosecuted. Lastly, the work focuses on proposed amendments to the Rome Statute that could answer these new legal challenges while safeguarding freedom of expression.

Introduction

The immense growth of artificial intelligence (AI) has introduced various new ways of disseminating information. AI's impact has been felt across all major sectors, including mass media, politics, and law. While many aspects of this AI-led transformation of information systems are commendable, AI poses various challenges for international criminal law (ICL) concerning the direct and public incitement to genocide, ensuring that its provisions are efficacious in today's world. The proceeding sections assess AI's ability to incite genocide and address questions of international criminal liability that may arise from such acts.

The following sections proceed to examine this question, bearing in mind the provisions of Article 25(1) of the Rome Statute¹ (hereafter, "the Statute"). Article 25(1) establishes the principle of "personal jurisdiction" and reads as follows: "*The Court shall have jurisdiction over all natural persons.*"

More specifically, sub-paragraphs a) to c) of paragraph 3 of Article 25(1) establish the tenets of individual criminal liability and are essential for understanding the subject at hand:²

- a) refers to three forms of perpetration: direct perpetration (through oneself), co-perpetration (in collusion with another person), and indirect perpetration (through another);
- b) establishes three primary forms of participation: ordering, soliciting, or inducing the commission of the crime; and

* Visiting Fellow, Lauterpacht Centre for International Law, University of Cambridge; NYU Law (2024); I would like to thank Professor Theodor Meron CMG (Charles L Denison Professor of Law Emeritus, NYU Law) for his valuable comments and supervision of this article.

¹ Rome Statute of the International Criminal Court, July 17, 1998, 2187 U.N.T.S. 90, art. 25(1).

² K. Bowman, "Commentary on Article 25(1) of the Rome Statute," *Case Matrix Network* (updated June 30, 2016), <https://www.casematrixnetwork.org/cmn-knowledge-hub/icc-commentary-clicc/commentary-rome-statute/commentary-rome-statute-part-3#c4025>.

c) establishes criminal responsibility for aiding and abetting.

Paragraph e) expands such attributions to the crime of genocide. All references to the term “machines” in Article 25(1) include semi-autonomous and autonomous weapons that wholly or partially employ AI.

Some legal historians say the drafters’ decision to exclude juridical persons was contentious.³ During the Rome Conference for drafting the Statute, a French-led proposal advocated for jurisdiction over juridical persons, similar to the Nürnberg Charter.⁴ A heated debate ensued, reflecting the differences between civil and common law nations over concerns that the domestic legal systems of many countries did not address the criminal liability of juridical persons or that the concept of criminal liability of juridical persons was incompatible with the idea of an International Criminal Court (ICC), and the proposal was defeated.⁵ Some argue that the proposal was out of order since the Statute’s Article 25(3), the UN War Crimes Tribunal’s *Zyklon B* case⁶ and the ICC’s *Chiquita* controversy⁷ are examples of top industrialists or their agents held individually responsible for their body corporates’ actions under ICL.⁸

Article 25(1)(e) is read with 25(3)(e), which provides criminal punishment for natural persons for the direct and public incitement to genocide, which is the focus here. The article

³ Ibid.

⁴ M. Biggs, “International Criminal Law and Corporate Actors – Part 2: The Rome Statute and Its Aftermath,” *Doing Business Right Blog*, May 21, 2019, <https://www.asser.nl/DoingBusinessRight/Blog/post/international-criminal-law-and-corporate-actors-part-2-the-rome-statute-and-its-aftermath-by-maisie-biggs>.

⁵ Report of the Preparatory Committee on the Establishment of an International Criminal Court, United Nations Diplomatic Conference of Plenipotentiaries on the Establishment of an International Criminal Court, Rome, Italy, June 15 – July 17, 1998, A/CONF.183/2.

⁶ *United States v. Tesch*, British Military Court, Hamburg, Mar. 1946, in *Law Reports of Trials of War Criminals*, vol. 1, 93 (United Nations War Crimes Commission, 1947).

⁷ S. Maupas, “Chiquita ‘Contributed’ to Colombian Paramilitary Crimes, ICC Told,” *Justiceinfo.net*, May 18, 2017, <https://www.justiceinfo.net/en/33351-chiquita-contributed-to-colombian-paramilitary-crimes-icc-told.html>.

⁸ J. Meza, “ICC Criminal Jurisdiction on Corporations for Criminal Liability and/or Civil Liability for Corporations,” *Harvard International Law Journal Blog*, May 13, 2021, <https://journals.law.harvard.edu/ilj/2021/05/fcc-personal-jurisdiction-on-corporations-for-criminal-liability-and-or-civil-liability-for-reparations/>.

is similar in substance and tenor to Article III(c) of the 1948 Genocide Convention⁹ and the Statutes of the International Criminal Tribunal for the former Yugoslavia (ICTY) and the International Criminal Tribunal for Rwanda (ICTR).¹⁰ In this regard, to incite publicly means to do so in a manner that is directed at the public at large through mass communication, and to do so directly means to do so in a manner that is directed specifically at a group of individuals while asking the public at large to engage in criminal activity against that group—thus distinguishing it from ordinary “instigation” as covered by Article 25(3)(b).¹¹ Genocide is the only international crime for which public incitement has been criminalized, and it is considered an inchoate crime, thus distinguishing it from the forms of complicity seen in subparagraphs b), c), and d).¹²

AI Ambiguity: Why the Lack of Consensus on a Definition of AI is Problematic for ICL

As scholars note,¹³ the definition of AI has changed consistently over the past few years. Arguably, the term was first coined in 1955 and defined by Stanford Professor Emeritus John McCarthy, who defined it as “*the science and engineering of making intelligent machines.*”¹⁴

Arend Hintze further divides AI into four different types:¹⁵

⁹ Art. III(c), *Convention on the Prevention and Punishment of the Crime of Genocide*, Dec. 9, 1948, 78 U.N.T.S. 277.

¹⁰ M. Klamberg, “Commentary on Article 25(3)(e) of the Rome Statute,” *Case Matrix Network* (updated June 30, 2016), <https://www.casematrixnetwork.org/cmn-knowledge-hub/icc-commentary-clicc/commentary-rome-statute/commentary-rome-statute-part-3#c4025>.

¹¹ *Ibid.*

¹² *Ibid.*

¹³ M. Swart, “Constructing ‘Electronic Liability’ for International Crimes: Transcending the Individual in International Criminal Law,” *German Law Journal* 24, no. 3 (2023): 589–602.

¹⁴ *Human-Centred Artificial Intelligence (HAI) Stanford*, “Artificial Intelligence Definitions,” *HAI Stanford* (Sep. 2020), <https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf>.

¹⁵ A. Hintze, “Understanding the Four Types of AI, from Reactive Robots to Self-Aware Beings,” *The Conversation*, Nov. 13, 2016, <https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>.

- i) **Reactive Machines:** These are the most basic form of machines, which only react to a given input, have no memory, and do not use past facts or experiences to make decisions. The chess supercomputer Deep Blue is an example of this. Another example that utilizes neural networks (a machine-learning process that mimics the human brain¹⁶) is Google's AlphaGo.¹⁷ Scholars argue that only these machines should be made since humans are not particularly adept at programming simulations with an accurate or unbiased environment for such machines to run on.¹⁸
- ii) **Limited Memory:** These machines can look into the past and transiently include simple information in their self-programmed memories. As an example, Hintze cites self-driving cars and how they collect information about traffic, the speed of other vehicles, and road changes.¹⁹
- iii) **Theory of Mind:** Hintze states that the machines of the future should not only be able to form representations about the world (as seen in Limited Memory) but also be able to form representations about specific things, which Hintze calls "agents."²⁰ Hintze connects this idea to Premack and Woodruff's "Theory of the Mind" in psychology (that is, "the ability to impute mental states to [...] [one's] self or others."²¹
- iv) **Self-Awareness:** According to Hintze, the ultimate goal is to build self-aware or self-conscious machines that can understand their emotions and gauge the feelings of those around them.²²

¹⁶ Amazon Web Services, "What are Neural Networks?," accessed Sep. 9, 2023, <https://aws.amazon.com/what-is/neural-network/#:~:text=A%20neural%20network%20is%20a%20method%20in,layered%20structure%20that%20resembles%20the%20human%20brain>.

¹⁷ A. Hintze, "Understanding the Four Types" (n 15).

¹⁸ R. Brooks, "Intelligence without Representation," *Artificial Intelligence* 47, nos. 1-3 (Jan. 1991): 139–59.

¹⁹ A. Hintze, "Understanding the Four Types" (n 15).

²⁰ *Ibid.*

²¹ D. Premack and G. Woodruff, "Does the Chimpanzee Have a Theory of Mind?," *Behavioral and Brain Sciences* 1, no. 4 (1978): 515–26.

²² A. Hintze, "Understanding the Four Types" (n 15).

Understanding these four types is immensely important for ICL practitioners and judges since current technology is developing in a vacuum between types ii) and iii), whereby machines do not yet rely on “moral heuristics,” compelling them to make highly utilitarian decisions.²³ This connects to Lessig’s broader idea of “Code is Law” and the philosophical discussion on how law (in this case, ICL) is part of the architectures of control and forms a part of moral heuristics.²⁴ In simpler terms, moral heuristics will ensure that machines comply with the norms of ICL instead of maximizing utility.²⁵ This also ensures that machines have *mens rea*, an essential component of criminal law.²⁶

Prosecuting AI-Enabled Direct and Public Incitement to Genocide: First Steps

The above begs the question: How do we identify and prosecute atrocities by an actor that lacks a definition?

The first step would be to agree on what qualities of AI make it “AI,” as what was once considered part of that definition might not be considered part of it today.²⁷ Scholars believe that distinguishing²⁸ between Very Strong AI (Artificial Superintelligence – ASI), Strong AI (Artificial General Intelligence – AGI), and Weak AI (Artificial Narrow

²³ M. Pardo and D. Patterson, “Neuroscience and Legal Theory: Jurisprudence, Morality, and Economics (Chapter 3),” in *Minds, Brains, and the Law: The Conceptual Foundations of Law and Neuroscience*, eds. M. Pardo and D. Patterson (New York: Oxford University Press, 2013), 47–78.

²⁴ J. Quinn, “‘Code is Law’ during the Era of Blockchain,” *Forbes*, May 17, 2022, <https://www.forbes.com/sites/forbesbusinesscouncil/2022/05/17/code-is-law-during-the-age-of-blockchain/?sh=181030a22adb>.

²⁵ A. Woods, “Moral Judgements & International Crimes: The Disutility of Desert,” *Vanderbilt Journal of International Law* 52, no. 3 (2012): 633, 669.

²⁶ M. Swart, “Constructing ‘Electronic Liability,’” 591 (n 13).

²⁷ *Ibid.*

²⁸ N. Joshi, “7 Types of Artificial Intelligence,” *Forbes*, June 19, 2019, <https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/?sh=38a086c7233e>.

Intelligence – ANI), or as described above, types iv), ii), and i), respectively, may also gradually become necessary, just as culpability differs between minors and adults.²⁹

Once AI has been defined as an umbrella term, the second step (ideally, taking place simultaneously with step three, enumerated below) would be to create individual definitions for each of the three terms mentioned above in addition to the overall umbrella definition of AI. Rex Martinez differentiates between two varied approaches: *descriptive* and *prescriptive*.³⁰

The *descriptive* approach seeks to define a term by describing its elements or features by reflecting on its actual grammatical and colloquial usage. However, the definition may in and of itself enlarge or narrow down the meaning of the term compared to its colloquial usage.³¹ Martinez offers the following definition of AI: “Artificial intelligence includes (1) Reactive machines, (2) Limited Memory machines, (3) Theory of Mind systems, and (4) Self-awareness systems, or include[s] other systems that utilize autonomous deep learning.”³² He thus focuses on describing the individual components of AI loosely rather than prescribing them. Jeanne Frazier Price calls the *descriptive* approach the “fuzzy categories”³³ approach, as the definition is based on a loose, open-ended definition that concentrates on elucidating a wide range of characteristics rather than prescribing any requisite conditions.³⁴

Martinez also rightly points out both the pros and cons of this approach.³⁵ First, a disadvantage of the system is that it bundles different elements with various functions and

²⁹ L. Cunningham, “A Question of Capacity: Towards a Comprehensive and Consistent Vision of Children and Their Status under Law,” *UC Davis Journal of Juvenile Law and Policy* 10, no. 2 (2006): 275, 277–78; see also A. Spence, “Parental Liability,” *Insurance Law Journal* 309 (Oct. 1948): 787, 787–88.

³⁰ R. Martinez, “Artificial Intelligence: Distinguishing between Types & Definitions,” *Nevada Law Journal* 19, no. 3 (2019): 1015, 1037–39.

³¹ J. Price, “Wagging, Not Barking: Statutory Definitions,” *Cleveland State Law Review* 60, no. 4 (2013): 999, 1011.

³² R. Martinez, “Artificial Intelligence,” 1037 (n 30).

³³ *Ibid.*

³⁴ *Ibid.*

³⁵ *Ibid.*

levels of autonomy. While Martinez argues that this is “not inherently problematic,”³⁶ he doesn’t specify why. Martinez’s opinion is inherently problematic for ICL because distinguishing the level of human control and the severity and pervasiveness of the functions of a specific kind of AI are arguably the most critical factors for determining culpability and prosecuting AI-powered atrocities. Second, the fuzzy category approach undermines the concept of a definition itself as it leaves ample room for argumentation, which can create difficulties for legislative systems in Global South nations that are parties to the Statute.³⁷ The only identifiable boon of the approach is that, for all its faults, it does enable the creation of a distinction between Strong and Weak AI.

On the other hand, *prescriptive* definitions “proscribe” certain preconditions that work with a positive or negative approach. A positive approach seeks to proscribe certain preconditions necessary to include a particular thing within a specific definition (for example, the elements of statehood under the Montevideo Convention).³⁸ By contrast, a negative approach seeks to define certain elements necessary to exclude a particular thing from a specific definition—for example, the airport security check process under the United States Transportation Security Administration (TSA)—which presumes consent for a mechanized scan unless the passenger specifically rejects it, say, for a standard pat-down check.

To analogize, Martinez provides the following standard definition of Strong AI per this approach: “Artificial intelligence is a system, program, software, or algorithm that acts autonomously to think rationally [...], act rationally [...], make decisions, [...] provide outputs”³⁹ and suggests substituting “acts autonomously” to “follows instructions” for Weak

³⁶ Ibid.

³⁷ R. Matthan, “The Approach to AI Regulation for the Global South – The Difference in Balancing Two Regulatory Approaches to AI,” *Comparative Law Review International* 25, no. 1 (2024): 1–5.

³⁸ T. Grant, “Defining Statehood: The Montevideo Convention and Its Discontents,” *Columbia Journal of Transnational Law* 37, no. 2 (1999): 403–58.

³⁹ R. Martinez, “Artificial Intelligence,” 1025–28 (n 30).

AI,⁴⁰ which can help distinguish the levels of human control and determine culpability under ICL. The only major con of this approach is that it may create rigid, highly narrow, or inflexible definitions, which runs the risk of creating loopholes for excluding actors who may otherwise be held culpable under ICL. While there may be other ways to define AI, there is merit in Martinez's conclusion that these two approaches are the most suitable and, if used appropriately and on a case-by-case basis, can eliminate ambiguities and lead to meaningful precedents or judicial benchmarks and regularize the prosecution of AI-enabled direct and public incitement of genocide under ICL.

How AI-Powered Genocide Works

AI tools, particularly those used on social media platforms, are attaining notoriety for their ability to exploit an "us versus them" divide and fuel hate speech and aggression, which may turn into genocide.⁴¹ For example, China has been known to use smartphone applications to track Uighurs,⁴² and serious concerns have also been raised about India's Aadhaar (unique personal ID) program.⁴³ Such identification programs can be used to target specific communities and then commit atrocities against them. Platforms such as Facebook and WhatsApp have also been flagged for vitriolic speech ("pour fuel over their heads") across

⁴⁰ Ibid., 1038.

⁴¹ M. Swart, "Constructing 'Electronic Liability,'" 593–94 (n 13); G. Stanton, "The Ten Stages of Genocide," *Genocide Watch*, accessed Sep. 9, 2024, <https://www.genocidewatch.com/tenstages>.

⁴² S. Feldstein, "China's High-Tech Surveillance Drives Oppression of Uyghurs," *Bulletin of the Atomic Scientists*, Oct. 27, 2022, <https://thebulletin.org/2022/10/chinas-high-tech-surveillance-drives-oppression-of-uyghurs/#:~:text=Developed%20in%20partnership%20with%20private,surveillance%20through%20mandatory%20DNA%20sampling>.

⁴³ G. Krishna, "Is Genocide Violence in Manipur Linked to Chronology of Aadhar Number Database, an Unending Census Project?," *Mainstream Weekly*, July 22 and 29, 2023, <https://www.mainstreamweekly.net/article13639.html>.

several jurisdictions, including Myanmar, as the UN's Independent International Fact-Finding Mission on Myanmar⁴⁴ has found.

Command Responsibility as a Possible Framework for Liability: An Introduction

The command responsibility doctrine is used in ICL to hold senior officials liable for atrocities committed by their subordinates when they knew or should have reasonably known about them and failed to take reasonable remedial measures to prevent such conduct or to assist in reporting and prosecuting it. The broader concept of command responsibility originates in Article 1(1) of the 1899 Hague Convention,⁴⁵ which stipulates that “armed forces must be commanded by a person responsible for his subordinates.”⁴⁶ Complexity arises when AI blurs the lines between military command responsibility under Article 28(a) and non-military command responsibility under Article 28(b) of the Rome Statute.

Kortfält rightly notes that this leads to a situation whereby the doctrine can be interpreted in two ways.⁴⁷ The first interpretation is that the commander (or in the case of AI, the operator or facilitator) is liable for their participation in the commission of the principal crime, with a lower burden of proof if it is inchoate. Thus, those individuals become responsible through what Kortfält calls “commission by omission”; that is, a situation where an omission is deemed an act under ICL.⁴⁸ However, the downside to this doctrinal analysis is that in the draft of the Statute, an article that indicated a general responsibility for omissions

⁴⁴ Human Rights Council, *Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar*, U.N. Doc. A/HRC/39/CRP.2, Sep. 28, 2018; M. Swart, “Constructing ‘Electronic Liability,’” 593 (n 13).

⁴⁵ Convention (II) with Respect to the Laws and Customs of War on Land, art. 1(1), July 29, 1899, 32 Stat. 1803, 1 Bevans 247.

⁴⁶ L. Kortfält, “General Remarks on the Commentary on Article 28 of the Rome Statute,” *Case Matrix Network* (updated Aug. 18, 2017), <https://www.casematrixnetwork.org/cmn-knowledge-hub/icc-commentary-clicc/commentary-rome-statute/commentary-rome-statute-part-3#c4025>.

⁴⁷ L. Kortfält, “Commentary on Article 28 of the Rome Statute,” *Case Matrix Network* (updated Aug. 18, 2017), <https://www.casematrixnetwork.org/cmn-knowledge-hub/icc-commentary-clicc/commentary-rome-statute/commentary-rome-statute-part-3#c4025>.

⁴⁸ *Ibid.*

was suggested but was excluded from the final version⁴⁹—thus indicating the drafters’ original intention. Scholars argue that the present principles under Article 28 are only a remnant of the initially proposed rule.⁵⁰ However, a counterargument to this would be that Article 28 should be read along with Article 86(2) of Additional Protocol I to the Geneva Conventions, which codifies a duty for superiors to prevent, prohibit, and punish violations of ICL by subordinates and that failing to fulfill this duty gives rise to criminal responsibility.⁵¹ The other way to interpret this provision would be to consider the mere holding of a superior position equates to contributing to the principal crime.

Relevant Case Law

This section focuses on four criminal tribunal judgments: *Prosecutor v. Tadić*,⁵² *Prosecutor v. Katanga*,⁵³ *Prosecutor v. Bemba*,⁵⁴ and *Prosecutor v. Delalić et al.*⁵⁵ Honing in on the central doctrinal elements that emanate from these cases (Control and Predictability and Intentionality and Foreseeability) may help define and distinguish the culpability of AI-powered genocide. For brevity, the section solely addresses the portions of the judgments that may be immediately relevant to this analysis and does not discuss the judgments in detail.

⁴⁹ Ibid.

⁵⁰ K. Ambos and O. Triffterer, eds., *The Rome Statute of the International Criminal Court: A Commentary*, 3rd ed. (München, Baden-Baden: CH Beck/Hart/Nomos, 2016), 808.

⁵¹ L. Kortfält, “General Remarks,” (n 46).

⁵² *Prosecutor v. Tadić*, Case no. IT-94-1-A, Appeals Chamber Judgement, International Criminal Tribunal for the Former Yugoslavia (ICTY), July 15, 1999. [Italicized portions of the judgment are underlined.]

⁵³ *Prosecutor v. Katanga*, ICC-01/04-01/07, Judgment on the Appeal against the Decision on the Implementation of Regulation 55 of the Regulations of the Court by Trial Chamber II of 21 Nov. 2012, International Criminal Court, Mar. 25, 2014. The judgment came down on Mar. 7, 2014.

⁵⁴ *Prosecutor v. Bemba*, Case no. ICC-01/05-01/08, Appeals Judgment (Redacted), International Criminal Court, June 8, 2018.

⁵⁵ *Prosecutor v. Delalić et al.*, Case no. IT-96-21-A, Appeals Judgement, International Criminal Tribunal for the Former Yugoslavia (ICTY), Feb. 20, 2001.

Prosecutor v. Tadić

The *Tadić* case pertained to the crimes committed in the Prijedor region of Bosnia and Herzegovina by Dusko Tadić, a local politician and leader affiliated with the Serb Democratic Party during the Yugoslav Wars. The ICTY charged him with committing crimes against humanity and violating the laws of war. The case is essential to this analysis because it involved the application of the doctrine of command responsibility to establish the accused's criminal liability. Indeed, the *Tadić* judgment was a seminal development in determining individual criminal responsibility under ICL,⁵⁶ which can help understand the broad contours of legal accountability in AI-enabled conflicts and can be utilized as a starting point to attribute culpability for AI-powered genocide.

More specifically, paragraph 137 of the 1999 decision, read in today's context, is highly relevant: “[I]t is by no means necessary that the controlling authorities should plan all the operations of the units dependent on them [...] or give specific instructions [...] control required by International Law may be deemed to exist when [...] the Party to the Conflict [...] has a role in organizing, coordinating, or planning military actions [...] or providing operational support.”⁵⁷ This paragraph implies that any role in the organization or the coordination of, including the mere installation and launch of any AI-powered technological entity or device that commits genocide, shall lead to the accrual of individual liability.

Prosecutor v. Katanga

⁵⁶ V. Vij, “Individual Criminal Responsibility Under Aiding and Abetting after the Specific Direction Requirement in the Taylor and Perisic Cases,” *Die Friedens-Warte* 88, no. 3/4 (2013): 8157–75.

⁵⁷ *Prosecutor v. Tadić*, Case no. IT-94-1-A (n 52), 137 (emphasis added).

Paragraphs 781–84 of the 2014 *Katanga* judgment are significant, as the ICC’s Trial Chamber focused extensively on the subject elements of Katanga’s role in perpetrating the crime at issue.

In the case of Weak AI, machines operate within a limited scope defined by the humans who are in control. Thus, AI operators can be held responsible through the principles established in the paragraph above. While some may be concerned by the use of VPNs or methods to mask the identities of the operator(s), a counterargument would be that the burden of proof lies with the prosecution, which doesn’t affect the operation of the law in any way. However, there is merit in the concern that this hurdle may lead to a low number of prosecutions (or even none), which in turn may lead to difficulties in crystallizing these legal doctrines into hard ICL. However, establishing culpability for Strong AI might be more complicated since it would naturally entail satisfying a higher burden of proof that the AI had substantial human control behind it. The notion of “contribution” may, however, apply if it can be demonstrated that the AI’s independent actions were facilitated without adequate safeguards.

Establishing culpability is not easy, given Strong AI’s more autonomous nature. However, an intriguing solution to this quagmire might lie in the doctrine of “piercing the corporate veil.”⁵⁸ This well-known doctrine allows courts to set aside limited liability and hold a corporation’s shareholders, promoters, or directors individually liable for the corporation’s acts or omissions.⁵⁹ For AI and ICL, the concept can be mainly utilized to establish individual liability by proving the existence of a failure in the duty of caring, programming, or monitoring AI. Thus, the culpability should be determined based on whether

⁵⁸ J. Zerk, *Towards a Fairer and More Effective System of Domestic Law Remedies*, 46 (OHCHR, 2012).

⁵⁹ Cornell Law School, “Piercing the Corporate Veil,” *Legal Information Institute*, accessed Sep. 9, 2024, https://www.law.cornell.edu/wex/piercing_the_corporate_veil#:~:text=Overview,most%20common%20in%20cl,ose%20corporations.

the operator shunned their responsibility under ICL by enabling conduct that led the autonomous or semiautonomous AI-enabled machine to foreseeably act in a manner that led to the violation of ICL, thus necessitating a situation whereby “piercing” the veil ensues between creators, operators, and facilitators of AI. While such a principle is usually seen in corporate law and would be unprecedented, it could likely be judicially incorporated as an analogical concept and then crystallized through a gradual litany of judicial pronouncements. Another critical question would be to establish a distinction between moral responsibility and legal liability, which would in turn require the creation of a broad benchmark (to be used on a case-by-case basis) to determine when the former turns into the latter.

Lessons from *Bemba*

The Jean-Pierre Bemba case⁶⁰ is an excellent example of situating the concept of command responsibility within the ICC’s jurisprudence and is vital for this analysis.

While the Trial Chamber found Bemba guilty in one of the ICC’s first convictions based on command responsibility, the Appeals Chamber overturned the conviction. The doctrine of command responsibility as applied in this case hinged on whether a military commander is responsible for preventing the commission of ICL violations by a militia under their effective control. The concept of effective control, knowledge of potential crimes, and the duty to prevent misuse can be extended to the realm of AI, especially in contexts where such actions or inactions could lead to ICL violations.

However, the main issue with considering *Bemba* as a helpful precedent is that the Appeals Chamber criticized the Trial Chamber (albeit narrowly) for effectively trying to

⁶⁰ *Prosecutor v. Bemba*, Case no. ICC-01/05-01/08 (n 54).

create a notion that vicarious liability exists in ICL.⁶¹ This opinion was refuted by the two dissenting judges (Monageng and Hofmański),⁶² who saw the majority's opinion as a dilution of command responsibility and its ability to permeate into individual criminal responsibility under the Statute. This, in turn, might allow similarly placed individuals to go scot-free in the future, thus setting a dangerous precedent for ICL.

Moreover, the Appeals Chamber decision leads to difficulties in scenarios where operational control is complicated by distance or logistics (e.g., Bemba was in Belgium when the troops were in the Central African Republic) or AI's autonomous nature. It also raises concerns about how expecting AI developers, operators, and facilitators to have complete control or foresee their actions might be excessive. Thus, the tenets of the *Bemba* decision, when applied within the context of AI, promote a more nuanced and cautious approach to using the principles of command responsibility under ICL.

Lessons from *Delalić et al.* (the “Celebici Camp” case)

In this case, Zejnil Delalić, a commander of the Bosnian Muslim forces during the Serb-Bosnian War, was tried along with his colleagues for crimes against humanity committed against prisoners of war held at his camp. This ICTY case is critical because it shows how superiors can be held responsible under command responsibility or vicarious liability.

The Trial Chamber held Delalić guilty by finding that he effectively controlled his subordinates. The Trial Chamber acknowledged and expounded on the concept of command responsibility, declaring it an integral part of Article 7(3) of the Statute.⁶³ The judgment is

⁶¹ *Ibid.*, 186–88.

⁶² Dissenting Opinion of Judge Sanji Mmasenono Monageng and Judge Piotr Hofmański, International Criminal Court, June 8, 2018, https://www.icc-cpi.int/sites/default/files/RelatedRecords/CR2018_02987.PDF.

⁶³ *Prosecutor v. Delalić et al.*, Case no. IT-96-21-T, International Criminal Tribunal for the Former Yugoslavia, Nov. 16, 1998.

also noteworthy for delineating the three elements of command responsibility in detail,⁶⁴ which is a highly useful precedent in the context of AI.

However, it is interesting to note that Delalić was acquitted on appeal.⁶⁵ The Appeals Chamber ruled that merely being in a position of power or creating an administrative entity does not automatically confer liability under ICL;⁶⁶ specifically, evidence of demonstrable control is necessary. The Appeals Chamber also addressed the concept of *de facto* authority under ICL.⁶⁷ This precedent is highly useful in the AI context, as it notes that mere creation, distribution, licensing/sublicensing, or any other form of ownership or transactional stake in an AI-enabled gadget or gizmo is insufficient. This is essential, as the Appeals Chamber used the criterion in Blaškić⁶⁸ and distinguished between a mere creator and an operator. It allows creators to pursue research and development activities for what might otherwise be valuable and allows actors to use legal military defense or ancillary technology without fear of ICL sanctions.

Prosecutor v. Ferdinand Nahimana, Jean-Bosco Barayagwiza, and Hassan Ngeze (the ICTR “Media Case”)

The ICTR “Media Case,”⁶⁹ one of the tribunal’s most widely cited and discussed judgments,⁷⁰ pertains to the media’s role in inciting genocide in 1994 during the Rwandan

⁶⁴ *Ibid.*, 346.

⁶⁵ *Prosecutor v. Delalić et al.*, Case no. IT-96-21-A (n 55).

⁶⁶ *Ibid.*, 192.

⁶⁷ *Ibid.*, 186–98.

⁶⁸ *Ibid.*, 58 n 254.

⁶⁹ *Prosecutor v. Nahimana, Barayagwiza, and Ngeze*, Case no. ICTR-99-52-T, Judgment and Sentence, International Criminal Tribunal for Rwanda, Dec. 3, 2003.

⁷⁰ J. McCoy, “Making Violence Ordinary: Radio, Music and the Rwandan Genocide,” *African Music* 8, no 3 (2009): 85–96; see also J. Metz, “Rwandan Genocide and the International Law of Radio Jamming,” *AJIL* 91, no. 4 (1997): 628–51.

Civil War. Nahimana and Barayagwiza were senior executives of the *Radio Télévision Libre des Mille Collines* (RTLM), while Ngeze was the founder of the *Kangura* newspaper. The Trial Chamber, after a long, drawn-out process over three years, found that *Kangura* and the RTLM were part of a “common media front” and “partners of a Hutu coalition,” the goal of which was to mobilize the majority Hutu community against the historically dominant Tutsis.⁷¹ The trial was the first since the *Streicher* trial at the International Military Tribunal (IMT) at Nürnberg to focus on the media’s role in the perpetration of war crimes.⁷²

In the “Media Case,” the ICTR found that the principal task of the two media entities was to forward the *Coalition pour la Défense de la République* (CDR) political party’s militant agenda by equating political interest with ethnic identity⁷³ and that the effect of the impartial media can be likened to poison,⁷⁴ with some nicknaming the RTLM as “Radio Machete.”⁷⁵ In 2007, the Appeals Chamber upheld most of the Trial Chamber’s observations while reducing the convicts’ prison sentences to thirty years and under.⁷⁶

In the context of AI-enabled genocide, the judgment can be made directly applicable to digital platforms that employ algorithms and content management systems that automate, create, distribute, and relay vitriolic material on a significant scale. The case also demonstrates that media operators (in this case, AI users) have a solemn duty to ensure that their usage of the technology does not create or exacerbate any illegalities under ICL. This judgment could also be used to pressure governments to commit to regularizing content

⁷¹ C. McKinnon, “Prosecutor v. Nahimana, Barayagwiza, & Ngeze, Case no. ICTR 99-52-A,” *AJIL* 98, no. 2 (2004): 325–30.

⁷² *Ibid.*, 328.

⁷³ *Ibid.*, 325.

⁷⁴ IRMCT, *Three Media Leaders Convicted for Genocide*, IRMCT Press Release (Mar. 12, 2003), <https://unictr.irmct.org/en/news/three-media-leaders-convicted-genocide>.

⁷⁵ *Ibid.*

⁷⁶ *Prosecutor v. Nahimana, Barayagwiza, & Ngeze*, Case no. ICTR-99-52-A, Appeals Judgment, International Criminal Tribunal for Rwanda, Nov. 28, 2007.

management and monitoring systems,⁷⁷ particularly in ethnically charged Global South nations.⁷⁸

Most importantly, the case underscores the need to embed ethical considerations within the framework of media operations. The focus of this case harkens back to the broader topic of ethical AI development; viz. ensuring that AI systems are compliant with ethical standards and human rights norms is essential to prevent them from being exploited for purposes similar to those seen in this case.

The Prosecution of AI-led Genocide: The ICC's Three Principal Hurdles

There are three concerning challenges to the ICC's ability to prosecute AI-led genocide.

The first challenge is the ICC's resource crunch. The Court works with a finite budget, subject to its member states' approval and contribution. The ICC's budget has not increased proportionately vis-à-vis the growing calls for the Court's intervention in multiple global conflicts.⁷⁹ Many scholars have bemoaned the lack of funding for investigating the "gravity" of crimes, which is particularly important in cases involving AI.⁸⁰ Naturally, the amount of resources that the ICC requires will increase with the case's complexity, which will mainly rise with the number of witnesses involved.⁸¹ Assessing AI's role may also

⁷⁷ J. Robinson, "AI's Unethical Underbelly," *Brown Political Review*, May 3, 2023, <https://brownpoliticalreview.org/2023/05/ais-unethical-underbelly/>.

⁷⁸ L. Garbe, L. Selvik, and P. Lemaire, "How African Countries Respond to Fake News and Hate Speech," *Information, Communication, and Society* 26, no. 1 (2023): 86–103.

⁷⁹ J. Anderson, "The ICC in Times of Budget Crunch," *JusticeInfo*, Dec. 13, 2021, <https://www.justiceinfo.net/en/85475-icc-times-budget-crunch.html>.

⁸⁰ S. Ford, "What Investigative Resources Does the International Criminal Court Need to Succeed?: A Gravity-Based Approach," *Washington University Global Studies Review* 16, no. 1 (2016): 1–70.

⁸¹ D. Groome, "No Witness, No Case: An Assessment of the Conduct and Quality of ICC Investigations," *Penn State Journal of Law & International Affairs* 3, no. 1 (2014): 1, 1–3.

require expert assistance, which may further drive up costs—given that these experts are paid at the UN P-4 pay scale.⁸²

Secondly, enforcing such judgments presents significant legal challenges, given the varying legal protections that body corporates and their employees in different jurisdictions enjoy. Even in the case of influential high-ranking individuals, states have shown recalcitrance or blatant disregard for the ICC’s jurisdiction; an excellent example is the response to the indictment of erstwhile Sudanese President Omar al-Bashir.⁸³ Such heavy dependence on the cooperation of states has hindered the Court’s ability to function efficaciously, mainly when sovereignty or political considerations are at stake. Naturally, these enforcement issues will become far more complicated when AI enters the picture, particularly with the ongoing lack of a global consensus on regulating AI.⁸⁴ AI’s unparalleled cross-border abilities might also make settling the jurisdiction question challenging.⁸⁵ Conversely, this may also be a boon for the ICC, as it can expand its jurisdiction to cover any AI-powered atrocities if any activity has occurred within an ICC member state’s jurisdiction.

Thirdly, developing standards and setting benchmarks for *de facto* responsibility and reprimandable conduct within the context of AI would require newfangled judicial and policy thinking and a dynamic interpretation of the Rome Statute and other ICL instrumentalities. Scholars suggest that this is an unparalleled opportunity for the ICC to engage in judicial activism and urge states to train AI systems to distinguish between “friends and foes” (aka

⁸² International Criminal Court, “Experts,” accessed Sep. 9, 2024, <https://www.icc-cpi.int/get-involved/experts#:~:text=Remuneration%20to%20experts%20is%20calculated,one%20of%20the%20following%20fields>; see also International Commission of Civil Servants, *United Nations Common System of Salaries, Allowances, and Benefits*, accessed Sep. 9, 2024, <https://icsc.un.org/Resources/SAD/Booklets/sabeng.pdf>.

⁸³ G. Barnes, “The International Criminal Court’s Ineffective Enforcement Mechanisms: The Indictment of President Omar Al Bashir,” *Fordham International Law Journal* 34, no. 6 (2011): 1585–619.

⁸⁴ E. Klein and S. Patrick, “Envisioning a Global Regime Complex to Regulate AI,” *Carnegie Endowment for International Peace Papers*, Mar. 21, 2024, <https://carnegieendowment.org/2024/03/21/envisioning-global-regime-complex-to-govern-artificial-intelligence-pub-92022>.

⁸⁵ T. Dias and R. Sagoo, “AI Governance in the Age of Uncertainty: International Law as a Starting Point,” *Just Security*, Jan. 2, 2024, <https://www.justsecurity.org/90903/ai-governance-in-the-age-of-uncertainty-international-law-as-a-starting-point/>.

“civilians and combatants”) and that the same can be done by stipulating guidelines for all member states or urging a conference of parties to the Statute to convene and establish a moral code for utilizing AI-led weapon systems.⁸⁶

Conclusion

The ICC’s three principal challenges must be addressed at length before the Court can prosecute an AI-led incitement to genocide. Whilst the general principles of ICL provide the foundational framework for overcoming such challenges, the practical difficulties of applying them must be considered and remedied through a comprehensive and reasonably resourced response by the ICC’s member states. Enhancing the Court’s capabilities, whether by increasing its budget, amending the Rome Statute, establishing a model code, or any other suitable means, will ensure that the ICC does not remain a silent spectator on the dark day when AI-led genocide transmogrifies into reality.

⁸⁶ O. Sari and S. Celik, “Legal Evaluation of the Attacks Caused by Artificial Intelligence-based Lethal Weapon Systems within the Context of the Rome Statute,” *Computer & Security Review*, 42, no. 1 (2021): 105564.