EMBODIED ARTIFICIAL INTELLIGENCE AND *JUS AD BELLUM* NECESSITY: INFLUENCE AND IMMINENCE IN THE DIGITAL AGE

FRANCIS GRIMAL AND MICHAEL J. POLLARD*

ABSTRACT

"*Prevention is better than the cure. . .*" — Erasmus, 16th Century[1]

In "re-opening" the classic debate surrounding a state's wider right of self-defense (in light of emerging technologies, and via the "lens" of influence communications), the authors controversially "close" the following discussion in favor of allowing Embodied Artificial Intelligence (EAI) to lawfully authorize pre-emptive acts of self-defense in response to non-imminent threats of a grave use of force. The authors provide a twofold justification for adopting this highly provocative stance.

First, they argue that the introduction of EAIs will facilitate a unique recalibration of the necessity and last resort requirements of self-defense which would "enable" certain pre-emptive actions to be re-categorized as "anticipatory." Secondly, the authors contend that because EAIs will be able to "compute" post-bellum considerations as part of their preparatory calculations, the potential unlawfulness of pre-emptive actions are further mitigated. In short, the utilization of EAI's will ensure that a greater range of humanitarian protections can

---

* Francis Grimal is a Reader in Public International Law, University of Buckingham, UK, and Michael J. Pollard PhD in Public International Law, University of Buckingham, UK. The authors would like to extend their sincerest thanks to Professor Christopher Waters, University of Windsor, Ontario, Professor Dr. Tom Ruys, Ghent University, and Lieutenant Colonel Hamish MacMillan of the U.K. Ministry of Defence Joint Information Activities Group for all their considerable advice and invaluable feedback throughout the preparation of this Article. Finally, the authors would like to extend their deepest gratitude to all at *Georgetown Journal of International Law*, to Madeline Bauer Editor-in-Chief and, in particular, to Miles Malley and Sophie Mehta for all their considerable suggestions and recommendations during the edits—the authors are extremely grateful & are delighted to be working with the Journal on this third installment of the series.

1. *See, e.g.*, ROYAL COLLEGE OF NURSING, https://www.rcn.org.uk/get-involved/campaign-with-us/prevention-is-better-than-cure (last visited Jan. 10, 2022) (writing that "[t]he phrase 'prevention is better than cure' is often attributed to the Dutch philosopher Desiderius Erasmus in around 1500. It is now a fundamental principle of modern health care and inherent within health and social care strategies across the UK"); *see also Prevention Is Better Than Cure*, THE OXFORD DICTIONARY OF PHRASE AND FABLE, https://www.oxfordreference.com/view/10.1093/acref/9780198609810.001.0001/acref-9780198609810-e-5664 (last visited Jan. 10, 2022).

be provided to the civilian population when future acts of self-defense are deemed necessary.

I. INTRODUCTION

In a briefing on July 28, 2021, a top U.S. military commander explained how the Pentagon is utilizing Artificial Intelligence (AI) to achieve information dominance and enable military planners to look far beyond the temporal urgency of seconds, minutes or hours, to instead predict a behavior or action in the coming days ahead.[2] As a result, in the following analysis (the third in a series of articles),[3] the

2. *See* Gen. Glen D. VanHerck, NORTHCOM Commander, NORTHCOM Commander Glen D. VanHerck Conducts Press Briefing on North American Aerospace Defense Command and U.S. Northern Command Global Information Dominance Experiments, (July 28, 2021) (transcript available at https://www.defense.gov/Newsroom/Transcripts/Transcript/Article/2711594/northcom-commander-gen-glen-d-vanherck-conducts-press-briefing-on-north-america/) [hereinafter GIDE].

3. For previous analyses *see* Francis Grimal & Michael J. Pollard, *"Embodied AI" and the Direct Participation in Hostilities: A Legal Analysis*, 51 GEO. J. INT'L L. 513 (2020) [hereinafter Grimal & Pollard (2020); Francis Grimal & Michael J. Pollard, *The Duty To Take Precautions in Hostilities, and*

authors use the exclusive nature of Embodied AI (EAI) to revisit the concept of imminence within the *jus ad bellum*, and recalibrate the "necessity threshold" of self-defense. The authors' proposed recalibration would lawfully permit a state to act pre-emptively against a non-imminent and latent threat of a grave use of force.

To reach such an undeniably controversial conclusion,[4] the authors begin by identifying the previously anonymous concept of "perpetual self-defense" — a notion which is squarely routed within the ever-expanding domain of influence communications and propaganda.[5] As its name suggests, "perpetual self-defense" is the recognition that all states are inherently (and as a default setting) "on the defensive" (be it DEFCON 4 or DEFCON 1)[6] particularly, in the context of influence communications and propaganda where states are constantly attempting to manipulate and even coerce their adversaries in the short, medium, and long term. Undeniably, the reader of this Article may immediately raise "objection" and quite rightly assert that while the recourse to influence communications and propaganda is an unsightly form of statecraft, it is one that falls outside the stringent regulation of the *ad bellum* framework.

The authors counter this, however, by identifying a second previously anonymous concept — that of "self-defense by proxy." "Self-defense by proxy" is used by the authors to denote a form of indirect pre-emptive self-defense that is best achieved by utilizing influence communications. Perhaps somewhat analogous to an indirect use of force, self-defense by proxy is where the authors envisage the use of influence communications and propaganda as part of a long-term strategy to destabilize a state adversary.

By way of example, one might envisage a sliding scale where, at the minor end, a state could choose to "bombard" its adversary, State B, with "anti-vax" propaganda in relation to the uptake of immunizations against Covid 19 (though undeniably harmful, such an act is one which clearly does fall outside of the *ad bellum* remit). In contrast, however, at

the Disobeying of Orders: Should Robots Refuse?, 44 FORD. INT. L. J. 671 (2021) (considering EAIs in greater detail in Part II) [hereinafter Grimal & Pollard (2021)].

4. Regarding the controversial nature of this conclusion, s*ee* the authors' discussion *infra* Part II.

5. Noting this concept is considered in greater detail throughout and specifically discussed *infra* Part IV.

6. DEFCON is the United States' Nuclear Defense Condition System. There are five levels of military 'readiness'; 5, Low/ Normal; 4, Above normal; 3, Airforce ready to mobilize in 15 minutes; 2, Armed forces primed and ready to be deployed within hours; 1, Maximum readiness, capable of acting/ responding immediately. For a useful discussion/ analysis *see, e.g.*, Scott D. Sagan, *Nuclear Alerts and Crisis Management*, 9 INT'L SEC. 99, 100-02 (1985).

the major end of the scale, State A might attempt to destabilize or even topple State B's government by "brainwashing" its civilian population (whether human or EAI) to the extent that that population employs physical force against the incumbent administration. A recent event which might be used as an example to demonstrate an action that sits somewhere between these two extremes might be something akin to the violence that was witnessed in Washington, D.C. in January 2021.[7] According to at least one report,[8] there is clear evidence to suggest that this violence, which followed President Biden's election victory, was fueled by a misinformation and disinformation campaign spread on social media platforms — noting, however, that in this instance there is no suggestion the lies were spread by the state, or were in some other way state sponsored.[9]

Nevertheless, with both "perpetual self-defense" and "self-defense by proxy" firmly in mind (bleak and dystopian though that may appear) the authors contend that in each concept there remains a strict constant — the state is acting pre-emptively against a non-imminent and latent threat of force. Therefore, pre-emptive self-defense by an EAI (via influence communications) is not only perhaps already lawfully acceptable, but also strategically desirable. Consequently, though each concept may initially appear somewhat abstract, they perhaps simply represent a crystallization of previously unidentified *lege ferenda* into tangible *lex lata*.[10]

Accepting that some forms of pre-emptive self-defense (particularly within the confines of influence communications and propaganda) are already *fait accompli*, the authors uniquely propose (for pre-emption to

7. As noted, for example, by Jan Wolfe, in January 2021 supporters of Donald Trump attacked the seat of the federal government in the U.S. Capital in an attempt to overturn the election result. In total, four people died and approximately 140 police officers were injured as a result of the violence. *See* Jan Wolfe, *Democracy under siege: An hour-by-hour look at the assault on the U.S. Capitol*, REUTERS (Jan. 4, 2022), https://www.reuters.com/world/us/democracy-under-siege-an-hour-by-hour-look-assault-us-capitol-2022-01-04/.

8. *See* Craig Silverman et al., *Facebook Hosted Surge of Misinformation and Insurrection Threats in Months Leading Up to Jan. 6 Attack, Records Show*, DEFENSE ONE (Jan. 4, 2022), https://www.defenseone.com/ideas/2022/01/facebook-hosted-surge-misinformation-and-insurrection-threats-months-leading-jan-6-attack-records-show/360333/.

9. Indeed, as noted by Wolfe, *supra* note 7, ¶ 8, then vice-president Mike Pence, adhering to his legal responsibility to certify the election result, made it clear that neither he or anyone else in government would be officially supporting the call from Trump and his supporters for the election result to be overturned.

10. The authors readily note that in order to establish the existence of an embryonic formulation of a new customary norm, the requisite state practice and *opinio juris* would need to be present.

be lawfully permissible),[11] a coupling of the necessity test at the moment of action, and *not* at the moment of threat. Doing so would remove the conceptual incompatibility between pre-emption, and the necessity requirement of self-defense — the stretching of time, against a non-imminent threat, makes justification of an action being one of last resort almost nigh impossible.

Currently, one envisages the EAI "running" the necessity test in 2021 against a perceived threat which may not materialize until 2029. This makes necessity very difficult to reconcile — given the EAI is applying necessity *now* rather than later. Instead, the authors conceive that the necessity requirement is physically applied and forecast in the context of 2029 by the virtue of the fact that the EAI can, and indeed has, calculated *every* single move — thus making that legal determination compatible with the requirement of "last resort." Conversely, and as accepted by the authors, stretching the temporal aspects at one end of the spectrum naturally invites a similar re-appraisal at the other. Consequently, the authors propose that as part of each EAI assessment, *post bellum* considerations should also form part of the overall calculation. In short, where an EAI identifies that lesser harms are more likely to occur as a result of acting sooner, it should be permitted to act — providing such actions adhere to the authors proposed "test" resulting from the necessary fusion of *ad bellum* and *in bello* norms, and to established *post bello* values.

At the heart of this discussion is the authors' firm belief that the current application of self-defense doctrine is somewhat "skewed." For example, the reader is no doubt already instinctively familiar with the legal requirement that lawful recourse to force in self-defense must be in response either to an actual "armed attack" (as codified in Article 51 U.N. Charter)[12], or (as is dictated by customary law) where a state is in imminent danger of suffering an armed attack.[13] However, if a state *must* wait until it suffers losses (or at least until the would-be aggressor has made it exceptionally clear that they intend to launch an armed

11. Which it is generally not currently considered to be.

12. U.N. Charter art. 51 [hereinafter U.N. Charter] reads, "Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security. Measures taken by Members in the exercise of this right of self-defence shall be immediately reported to the Security Council and shall not in any way affect the authority and responsibility of the Security Council under the present Charter to take at any time such action as it deems necessary in order to maintain or restore international peace and security."

13. Noting that imminence may also potentially relate to a threat of a grave use of force.

attack) an act of self-defense may, in reality, represent nothing more than a reprisal or retaliatory act — noting that these are generally considered taboo under international law.[14]

Nevertheless, as the authors have noted elsewhere, (and as Commander General VanHerck is no doubt very aware) EAIs offer humankind an opportunity to peer further into the future than has been previously possible.[15] Going forward, not only will EAIs be better placed to predict future events with increasing accuracy but more significantly, they will undoubtedly be able to manipulate these events in order to shape future outcomes. With that in mind, the present authors forward the hypothesis that an EAI *should* be lawfully permitted to act where it identifies lesser harms that are more likely to occur as a result of acting sooner rather than later — a natural consequence of an EAI's capability to accurately calculate comparatively enormous amounts of data.[16]

Perhaps a useful illustration regarding the authors' proposed approach are the inherent similarities both practical (and conceptual) with the one thousand, five-hundred-year-old game of chess.[17] Initial advances in chess technology (perhaps most familiar to the reader) will be the advent of computer "participants" as evidenced by the mid-nineties infamous battle between the then reigning human chess champion, Gary Kasparov, and the IBM supercomputer, known as

---

14. *See, e.g.*, Shane Darcy, *Retaliation and Reprisal*, *in* THE OXFORD HANDBOOK OF THE USE OF FORCE IN INTERNATIONAL LAW 878, 878 (Marc Weller ed., 2015) [hereinafter OXFORD HANDBOOK]. Here the author notes, "The concepts of retaliation and reprisal have had a peripheral presence in the law governing the use of force in international relations. Their exact meaning and scope have often proved elusive and despite the apparent silence on the matter of relevant international treaties, the overwhelming weight of opinion is that a use of force by way of retaliation or reprisal is generally unlawful." Also *see, e.g.*, YORAM DINSTEIN, WAR AGGRESSION AND SELF-DEFENCE, ¶¶ 691-695, at 264-65 (3rd ed. 2017).

15. The point being if tech is already able to look days into the future, it will almost certainly be capable of looking further as it evolves.

16. The term AI is used to refer to algorithms which operate according to a set of fixed, pre-programmed parameters, but also those which apply various forms of machine learning. The latter concept is fundamentally different from the former because it can learn to predict future behaviours based on previous acts. One might say, therefore, to some extent, it can learn to predict the future. *See, e.g.*, Jeremy Bender, *Machine Learning or Automation: What's the Difference?*, BUSINESS NEWS DAILY (Aug. 12, 2022), https://www.businessnewsdaily.com/10352-machine-learning-vs-automation.html (last visited Jan. 10, 2022).

17. Although early forms of the game are widely acknowledged to have originated in India some 1500 years ago, the game as we know it today did not appear until the 16th Century. *See, e.g.*, Colin Stapczynski, *History of Chess: From Early Stages to Magnus*, CHESS.COM (June 22, 2022), https://www.chess.com/article/view/history-of-chess (last visited July 11, 2022); HENRY A DAVIDSON, A SHORT HISTORY OF CHESS ¶2 (1st ed. 1949).

Deep Blue.[18] More recently, engines such as "Stockfish" have advanced the "art of calculations" to a whole new level.[19] Part of what the authors envisage (when coupling necessity to the actual point of attack, and the post *bellum* considerations) are analogous to chess itself. The "engine" will calculate not only the openings — whether one replies to e4 with a c5 "Sicilian defense" — but every mid game, and end game position/consideration too.

In effect, it is the *pre-bellum* opening calculations, and the *post bellum* end game calculations (via the lens of influence communications) that the authors wish to question in greater detail. Interestingly, although undeniable anecdotal, was the current FIDE World Champion,[20] Magnus Carlsen, and his remark (when playing against the "Magnus Carlsen App: version of himself) that the App was playing a "pointless" engine move by cowardly (!) playing Bishop to B7.[21] Clearly, the purpose of this Article is not to provide commentary on the choice of move by the "engine," but to illustrate, that as part of the overall calculation, the somewhat left-field approach that a human cannot "comprehend" is precisely why an EAI is better placed to make the relevant assessments both *pre* and *post ad bellum*.

As previously noted, the present Article utilizes the notion of military-led influence communications to expand upon the authors' previous, unique, forays regarding the fundamental changes of the *jus ad bellum* and *jus in bello* thresholds that may result from the introduction of increasingly advanced EAI. In this regard, much of the existing scholarly attention has been placed on quantum — how much artificial intelligence (AI) will impact upon or disrupt military operations. And, thus far, the literature has understandably been pre-occupied with cyber-attacks,[22]

---

18. *See Deep Blue*, IBM, https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/ (last visited July 11, 2022).

19. As noted by CHESS.COM, *supra* note 17, Stockfish is widely regarded as the king of "chess engines.". *Stockfish*, CHESS.COM, https://www.chess.com/terms/stockfish-chess-engine (last visited July 11, 2022); See *About,* STOCKFISH CHESS, https://stockfishchess.org/about/ (last visited July 11, 2022).

20. FIDE is the long form of the International Chess Federation, which in the French Language (FIDE being formed in Paris in 1924 and headquartered in Lausanne, Switzerland is presented: Fédération Internationale des Échecs.

21. *See Magnus Carlsen, Magnus Carlsen vs. Himself at 20 on the Play Magnus Chess App*, YOUTUBE (May 27, 2014), https://www.youtube.com/watch?v=pNvVWeHZG00 (last visited Jan. 10, 2022).

22. *See, e.g.*, Francis Grimal & Jae Sundaram, *Cyber Warfare and Autonomous Self-defence*, 4 J. ON THE USE OF FORCE IN INT'L L. 312, 324 (2017); Matthew C. Waxman, *Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4)*, 36 YALE J. INT'L L. 421, 423 (2011); Muhammad Mudassar Yamin et al., *Weaponized AI for Cyber Attacks*, 57 J. OF INFO. SEC. AND APPLICATIONS 1 (2021); MARCO ROSCINI, CYBER OPERATIONS AND THE USE OF FORCE IN INTERNATIONAL LAW (OXFORD UNIVERSITY

and/or autonomous weapons systems (AWS).[23] However, as the authors have argued and identified elsewhere,[24] the introduction of non-armed (and often civilian) EAI systems will also pose unique challenges not only for the preservation of international law, but also for tort and criminal law.

By way of a more complete explanation, an EAI is the "coupling" of advanced robotic systems, with sophisticated AI frameworks. Simply put, an EAI is AI manifested as a tangible intelligent robot. In principle, EAIs already exist. For example, production-line robots such as those who are programmed to fit the doors to a Cadillac Escalade or seem-weld the panels of a Ford Bronco do already display a basic form of intelligence. However, these generally operate in a fixed position, and according to a rigid set of predetermined "non-negotiable" instructions: turn 180 degrees – locate door – lift door – turn 180 degrees carrying door – fix door to vehicle – "rinse and repeat."

Indeed, though no one can predict with any certainty how EAIs will or will not be restricted from a technical perspective ten or twenty years from now, the operational parameters of current AI systems are undoubtably limited.[25] Moreover, developing AI technology which allows for increasingly autonomous robots will remain surrounded in controversy due, not least, to concerns regarding their (un)predictability. In stark contrast however, the EAIs envisaged by the present authors are infinitely more capable. Indeed, the present authors uncompromisingly support the notion that at some yet to be determined point in future, EAIs will be delegated the authority to select a particular course of action from a (potentially infinite) number of alternative choices — and will do so without human supervision and/or the possibility of immediate human intervention.

PRESS ED., 1ST ED. 2014); JAMES A. GREEN, CYBER WARFARE: A MULTIDISIPLINARY ANALYSIS, 96 (James A. Green ed., 1st ed. 2015).

23. *See* the current authors previous works, *supra*, note 3 noting, for example, Bonnie Docherty et al., *Heed the Call: A Moral and Legal Imperative to Ban Killer Robots*, HUMAN RIGHTS WATCH (Aug. 21, 2018), https://www.hrw.org/report/2018/08/21/heed-call/moral-and-legal-imperative-ban-killer-robots; Heather M. Roff & David Danks, "*Trust but Verify": The Difficulty of Trusting Autonomous Weapons Systems*, 17 J. MIL. ETHICS 2 (2018); NEHAL BHUTA ET. AL., AUTONOMOUS WEAPONS SYSTEMS: LAW, ETHICS, POLICY (Nehal Bhuta et. al. eds., 1st ed. 2016); ARMIN KRISHNAN, KILLER ROBOTS: LEGALITY AND ETHICALITY OF AUTONOMOUS WEAPONS (Armin Krishnan ed., 1st ed. 2009); PAUL SCHARRE, ARMY OF NONE: AUTONOMOUS WEAPONS AND THE FUTURE OF WAR (Paul Scharre ed., 1st ed. 2018).

24. *See generally* Grimal & Pollard (2020), *supra* note 3; *see also* Grimal & Pollard (2021), *supra* note 3 (exemplifying the authors' past works).

25. *See generally* Grimal & Pollard (2020), *supra* note 3 (arguing that EAI is a gap in international law).

As is true of many of the discussions in this realm (including, for example, those which have regard for AWS)[26], the reader may be required to take a small, conceptual "leap of faith." There is no doubt, however, that EAI technology will become increasingly mobile in nature, and much more varied in application.[27] Indeed, one of the most widely touted and "upwardly mobile" EAIs of the present day is Boston Dynamics' "Spot." If the reader has somehow managed to evade an introduction thus far, Spot is a quadruped EAI with a robotic arm where one might expect to find a neck and head.[28] This gives Spot an appearance that is (somewhat) similar to a large dog — which, a number of critics find disturbing regardless of the fact that Spot is very clearly not sentient.[29] Spot's manufacturers claim it is "an agile mobile robot that navigates terrain with unprecedented mobility."[30] Indeed, they identify that the EAI is already capable of autonomously conducting various missions, including, search and alert in hazardous environments and disaster areas,[31] telemedicine missions (including virtual consultations).[32] Spot can even act as an "entertainer" or performance artist,[33] and has, for example, recently been watched on video nearly 2.5 million times,[34] expertly mimicking the dance moves of the Rolling Stones' Mick Jagger. Just a short while ago, however, early versions of

26. *See, e.g.*, Rebecca Crootof, *The Killer Robots are Here: Legal and Policy Implications*, 36 CARDOZO L. REV. 1837, 1840 (2015). Here the author also identifies that while there is now a plethora of literature regarding AWS, the overwhelming consensus is that such weapons do not yet exist.

27. The point raised by the authors, is that the vast majority of early, and indeed contemporary robots are fairly limited in their applications: they are only designed and programmed to carry out a single task such as; mow the lawn; vacuum the rug; make the coffee; fix the door to the car etc.

28. *See Spot*, BOSTON DYNAMICS, https://www.bostondynamics.com/spot (last visited Jan. 10, 2022).

29. Spot has, for example, been referred to as 'terrifying.' *See, e.g.*, Peter Holley, *Boston Dynamics' 'Terrifying' Robotic Dogs Have Been Put to Work by At Least One Police Agency*, WASH. POST (Nov. 26, 2019), https://www.washingtonpost.com/technology/2019/11/26/boston-dynamics-terrifying-robotic-dogs-have-been-put-work-by-least-one-police-agency/. Concerns are not only raised in relation to Spot's appearance, but also to its potential capabilities. *See, e.g.*, Jeremy Moses & Geoffrey Ford, *Is 'Spot' a Good Dog? Why We're Right to Worry About Unleashing Robot Quadrupeds*, THE CONVERSATION (May 3, 2021), https://theconversation.com/is-spot-a-good-dog-why-were-right-to-worry-about-unleashing-robot-quadrupeds-160095.

30. *Spot*, s*upra* note 28.

31. *Id.*

32. *Id.*

33. *Id.*

34. *See Boston Dynamics, "Spot Me Up" | The Rolling Stones & Boston Dynamics*, YOUTUBE (Oct. 29, 2021), https://www.youtube.com/watch?v=XnZH4izf_rI.

Spot were criticized as being clumsy and overly simplistic.[35] The point is, robotics and AI technology are moving forward at unprecedented speed — and there is simply no doubt that Spot (and its humanoid compatriot "Atlas"),[36] are merely at the very tip of a much larger iceberg.

The authors adopt a multi-faceted approach to the following analysis, and fully recognize that this may add an element of complexity that would not be present if the dialogue was divided into two or more volumes. For example, much of this analysis could be used merely to support a narrower argument highlighting the fact that the introduction of advanced AI systems may require a recalibration of *ad bellum* necessity only in regard of self-defensive actions which are "compatible" with Article 51 U.N. Charter. However, if such an approach was adopted, the examination would have no regard to the authors' central and indeed foundational assertion that influence communications are already being deployed as a form of ongoing self-defense and, moreover, that algorithms already significantly impact how such influence is exerted.

With that in mind, throughout the following work, the authors have regard to three "constants." In the first instance the investigation is conducted under the notion that certain acts of self-defense should more currently and more correctly be classified as acts of reprisal. Secondly, they believe that an EAI should be permitted to act where it predicts there is a necessity to act "sooner rather than later" because a sequence of events has reached the "point of no return" (and to delay would mean any future act would also stray into the realm of reprisal). Finally, the authors examine the prospect of lawfully permissible pre-emptive actions appearing in the form of influence communications, otherwise referred to as "perpetual self-defense." Within the heartbeat of this discussion is the fact that influence can be communicated in a manner in which a "foreseeable consequence" of its use is the application of force.[37] Indeed, the ultimate purpose of a military-led influence

35. *See, e.g.*, olinerd, *Boston Dynamics Big Dog (New Video March 2008)*, YOUTUBE, https://www.youtube.com/watch?v=W1czBcnX1Ww (last visited July 11, 2022) where Spot's predecessor 'bog dog,' looks somewhat clumsy in comparison to today's agile machine. Indeed, a tweet referring to the early video exclaims "[a]fter years of research & millions of dollars, engineers can accurately replicate two drunk people carrying a sofa." *See* HighTechPanda (@HighTechPanda), TWITTER (Feb. 17, 2016), https://twitter.com/hightechpanda/status/700160330991677440.

36. *Atlas*, BOSTON DYNAMICS, https://www.bostondynamics.com/atlas (last visited Jan. 10, 2022).

37. *See, e.g.*, Pontus Winther, *Military Influence Operations & IHL: Implications of New Technologies*, HUMANITARIAN L. & POL'Y (Oct. 27, 2017), https://blogs.icrc.org/law-and-policy/2017/10/27/military-influence-operations-ihl-implications-new-technologies/.

operation may be an "indirect" application of force against the adversary[38] — albeit that indirect actions of this type largely escape the scrutiny of international law.[39]

It is precisely with this in mind that the present authors proffer that where a foreseeable consequence of influence is the application of force (irrespective as to whether it is applied directly or indirectly) acts of this nature should fall under the banner of "self-defense by proxy" (a second concept coined by the present authors). In recognizing this as an additional form of self-defense, the authors contend that the ultimate intended target *should* be a lawful target under existing international norms. This is highly controversial, not least because, as previously noted, self-defense by proxy is fundamentally grounded within the overarching realm of perpetual self-defense — a concept that is almost exclusively pre-emptive in nature.

By way of solution, the authors also introduce a novel test for establishing the lawfulness of self-defense by proxy targets. Central to this test is the principle of concurrent application (of *jus ad bellum* and *jus in bello* targeting norms), which the authors marginally manipulate to provide a quasi-legal status to *pre-bello* and *post-bello* considerations. This is crucial to the overall analysis because in doing so they ensure that a level of humanitarian protection can be provided which surpasses that which is currently provided by existing interpretations.

By way of overall overview, the analysis continues by providing the authors' definition of EAI in Part II. Part III identifies those *jus ad bellum* and *jus in bello* rules that are applicable (in their use) when purposed for self-defensive acts and the further implications of self-defensive actions within the twilight zone of those two distinct realms of international law. Part III conducts a detailed analysis of influence communications within the overarching context of pre-emption. Following this discussion, the authors identify the concept of perpetual self-defense," and the existence of the notion of "pre-emptive self-defense by proxy." In Part IV, the authors provide an authoritative "test" for determining whether the target of an act of self-defense by proxy should be deemed as lawful. To concretely illustrate its application, the authors run and simulate that test through a number of scenarios. Part V of the discussion considers a number of further implications that are

38. Noting that in theory there is no reason why one could not choose to influence one's ally, either openly, or in a clandestine manner.

39. *See, e.g.*, Winther, *supra* note 37, where the author notes "[t]here is no comprehensive regulation under IHL on the use of communication to affect peoples' attitudes and behavior during armed conflict."

likely to arise as a consequence of the recalibration of the pre-emptive threshold (from the introduction of EAI for the purpose of influence communications in the wider sense of the *ad bellum*). Somewhat uniquely, the authors continue the trajectory of pre-emptive action via influence communications by examining its possible effects on both the United Nations Security Council (in the context of collective security) and a state's extra charter "considerations" of Humanitarian Intervention and Responsibility to Protect (R2P). Finally, the authors offer their closing thoughts.

II. The Jus ad Bellum and Jus in Bello Legal Framework

Prior to the advent of EAI (Embodied Artificial Intelligence), one may have reasonably concluded that under the *jus ad bellum* (when acting in self-defense), all indirect applications of force should be both "necessary" and "proportionate." Furthermore, given that influence is also a means of achieving a state's aims during an armed conflict, the identification of "lawful targets" under the *jus in bello* would have been a reasonably conventional task, albeit one requiring careful navigation through the *in bello* parameters of "distinction" and "proportionality." Indeed, the classic contours of this "traditional" discussion are further captured in Part II. However, the authors believe influence communications *should* only target military objectives (at the point at which the force is intended to be applied),[40] and must be both necessary and proportionate (in the International Humanitarian Law (IHL) sense).[41]

To convincingly establish the lawfulness of acts of "self-defense by proxy" requires an indispensable and veritable fusion of *ad bellum* and *in bello* considerations. This has been previously referred to by two authors as "the principle of 'concurrent application,'"[42] noting that while the *ad bellum* and *in bello* categories should be viewed independently,[43] the state of "separation and "interaction" between the two

40. The point here being that an "influencer" will generally utilize the form of communications being considered to "target" individuals in an attempt to persuade them to behave in a manner that is beneficial primarily to the influencer, though potentially to both parties.

41. The "core" principles of IHL are introduced and considered in greater detail *infra* part II.

42. *See generally* James A. Green & Christopher P. M. Waters, *Military Targeting in the Context of Self-defense Actions*, 84 Nordic J. Int'l L. 3 (2015).

43. Green & Waters, *supra* note 42, at 25–28. Here, the authors identify Protocol Additional to the Geneva Conventions of 12 august 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), Jun. 8, 1977, 1125 U.N.T.S. 3 [hereinafter API], and identify the Geneva Conventions and their Additional Protocols "must be fully applied in all circumstances to all persons who are protected by those instruments, without any adverse

realms should be viewed positively rather than negatively.[44] In most instances, concurrent application will appear unnecessary since an armed attack is also likely to constitute a declaration of war (albeit noting that wars are not always formally "declared").[45]

As the present authors further expand on below, this principle nevertheless can ensure that the maximum level of protection is offered to civilians in all circumstances. A simple example of this is where an act of self-defense may be justified, but the *in bello* is not triggered. This may be the case, for example, where an attack on a single naval vessel could trigger a state's inherent right of self-defense without surpassing the threshold that is needed for the combined action to be considered an armed conflict.[46] Here, concurrent application simply means that a state invoking its inherent right of self-defense must not make civilians the object of a direct attack — which, of course, is strictly an IHL principle and thus preservation of the *in bello* rather than the *ad bellum*.[47]

As previously noted, the authors wish to ground the wider discussion herein firmly within two existing concepts — EAIs, and military led-influence operations. Before moving on to consider these in greater detail, however, Part II of this Article (Sections A, B and C: see below) examines the existing legal parameters within both the *ad bellum* and the *in bello* which can be used to govern the use of EAIs for the purpose of self-defense. By way of context, Part III of this Article will then address the authors' more focused discussion regarding influence communications as a further, and yet unidentified, forms of self-defense (perpetual and by proxy). For the sake of completeness of the present discussion (and for reasons that are expanded upon further in Parts IV

distinction on the nature or origin of the armed conflict or on the causes espoused by or attributed to the Parties to the conflict. . . ." 1125 U.N.T.S. 3, at 7.

44. Green & Waters, *supra* note 42, at 27.

45. *See, e.g.*, DINSTEIN, *supra* note 14, ¶¶ 95-96, at 35-36.

46. *See* Green & Waters, *supra* note 42, at 20-21. It remains unclear whether an "intensity" threshold for armed conflict exists, and equally what constitutes the threshold for an armed attack (i.e., that which invokes the inherent right to self-defense). As considered in greater detail in Part II, the latter is generally understood to mean the "most grave use of force." *See* Military and Paramilitary Activities in and Against Nicaragua (Nicar. v U.S.), Merits, Judgment, 1986 I.C.J. Rep 14, ¶ 191 (June 27) [hereinafter *Nicaragua*]. However, although the ICJ acknowledges *Nicaragua* in; Oil Platforms (Islamic Republic of Iran v U.S.) Merits, Judgment, 2003 I.C.J. Rep 161, (Nov. 6) [hereinafter *Oil Platforms]*, it also notes that "the court does not exclude the possibility that the mining of a single military vessel may be sufficient to bring into play the inherent right to self defence." *Id.* ¶ 72. *See also* Green & Waters, *supra* note 42, at 20, where the authors note that some might determine in contrast that the exact same attack on the naval vessel could trigger IHL, but not the inherent right to self-defense. The point is it is highly dependent upon which threshold one applies.

47. It being considered a basic rule of IHL, *see* API, *supra* note 43, art. 48.

and V), the authors not only wish to highlight the traditional *jus ad bellum* self-defense principles, but also the relevant *jus in bello,* and *jus post bellum* provisions. The examination in Part II commences with the genesis for any *ad bellum* discussion, Article 2(4) of the U.N. Charter.[48]

A. *Prohibition on the Use and Threat of Force*

Readers will undoubtedly quickly recall the cardinal prohibition against both the threat and use of force contained in Article 2(4), and its direct bearing on any discussion involving self-defense.[49] For the record, Article 2(4) states: "All Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations."[50]

This provision provides a "negative" prohibition in direct contrast to the "positive" expectation and obligation placed on states to settle their disputes by pacific or peaceful means as outlined in Article 2(3).[51] As noted by one of the authors of this Article all too often elsewhere, it is important to underscore that the prohibition contained in Article 2(4) does not give automatic rise to peremptory and *jus cogens* status.[52] However, the scholarship typically accepts that the prohibition is nonetheless a peremptory norm, and as such, it must not be derogated from.[53] The "coupling" effect of Article 2(4) and 2(3) in light of a holistic reading of Article 2(7) of the U.N. Charter and the customary principle

48. U.N. Charter, s*upra* note 12, art. 2(4).

49. *See* Francis Grimal, *Twitter and the jus ad bellum: threats of force and other implications*, 6 J. ON THE USE OF FORCE AND INT'L L. 183, 183–192 (2019). *See also* Grimal & Sundaram, *supra* note 22, at 326. Here the authors identify; OLIVIER CORTEN, THE LAW AGAINST WAR: THE PROHIBITION ON THE USE OF FORCE IN CONTEMPORARY INTERNATIONAL LAW 50-197 (Christopher Sutcliffe trans., 2010); THOMAS M. FRANCK, RECOURSE TO FORCE: STATE ACTION AGAINST THREATS AND ARMED ATTACKS 11–19 (2002); Nico Schrijver, *The Ban on the Use of Force in the U.N. CHARTER*, *in* OXFORD HANDBOOK, *supra* note 14, at 466.

50. U.N. Charter, *supra* note 12, art. 2(4).

51. The U.N. Charter, *supra* note 12, art. 2(3) provides, "All Members shall settle their international disputes by peaceful means in such a manner that international peace and security, and justice, are not endangered."

52. Grimal & Sundaram, *supra* note 22, at 323, 338. Here the authors identify; James A. Green, *Questioning the Peremptory Status of the Prohibition of the Use of Force*, 32 MICH. J. INT'L L. 215 (2010).

53. Grimal & Sundaram, *supra* note 22, at 324 identifying: ALEXANDER ORAKHELASHVILI, PEREMPTORY NORMS IN INTERNATIONAL LAW (Oxford University Press eds., 2006). It is important to underline that it is not Article 2(4) *per se* that "enjoys" jus cogens status: rather, it is the prohibition of the use of force contained therein.

of non-intervention clearly provides a strict obligation to prevent states from interfering with the sovereign affairs of another.[54]

As previously noted, Part III considers influence communications as a form of self-defense in greater detail. However, and as noted by Schelling, "[i]t is *latent* violence that can influence someone's choice."[55] Therefore, in short, a threat of violence might be used to encourage a threatened party to act in a way that is beneficial to the threatening party. Alongside actual use of force, threats of force are undeniably prohibited by Article 2(4) U.N. Charter,[56] but their lack of concrete definition has allowed some scholars (*mea culpa*) to investigate their "nebulous nature" in more detail, while noting that states undoubtedly typically only raise concern when threat materializes into actual force.[57]

However, and in direct relation to this present Article, it is actually the current "test" for determining the lawfulness of a threat of force espoused by the ICJ in the *Nuclear Weapons advisory opinion,*[58] which is of

54. Grimal & Sundaram, *supra* note 22, at 324 citing the Declaration on the Principles of International Law Concerning Friendly Relations and Cooperation Among States in accordance with the Charter of the United Nations, annexed to UNGA Res 2625, U.N. Doc A/RES/2625 (XXV) (24 October 1970) and the Declaration on the Enhancement of the Effectiveness of the Principle of Refraining from the Threat or Use of Force in International Relations, annexed to UNGA Res 42/22, U.N. Doc A/RES/42/22 (18 November 1987).

55. THOMAS C. SCHELLING, ARMS AND INFLUENCE 3 (1966); *accord* FRANCIS GRIMAL, THREATS OF FORCE: INTERNATIONAL LAW AND STRATEGY 7 (2012).

56. *See generally* James A. Green & Francis Grimal, *The Threat of Force as an Action in Self-Defense Under International Law*, 44 VANDERBILT J. OF TRANSNAT'L L. 285, 286 (2011); Grimal, *supra* note 49, at 185 stating, "practice still typically favors the 'referencing' of the prohibition of an *actual use of force* compared to a *threatened use of force*, the latter seems to be slowly gaining some momentum in terms of awareness. Overwhelmingly, though, instances of when a threat of force (a prima facie unlawful action under Article 2(4)) are actually 'referenced' by states remain secondary to actual uses of force."

57. Grimal, *supra* note 49, at 7 identifying; Brian Drummond, *UK Nuclear Deterrence Policy: An Unlawful Threat of Force*, 6(2) J. ON THE USE OF FORCE AND INT'L L. (2019) (pagination awaiting). *See generally* Green & Grimal, *supra* note 56, at 299; *see generally* GRIMAL, *supra* note 55, at 78; Dino Kritsiotis, *Close Encounters of a Sovereign Kind*, 20(2) EUR. J. OF INT'L L. 299 (2009); Marco Roscini, *Threats of Armed Force and Contemporary International Law*, 54 NETH. INT'L L. REV. 229, 245 (2007); NIKOLAS STÜRCHLER, THE THREAT OF FORCE IN INTERNATIONAL LAW 218-51 (2007); On the point regarding an agreed definition of what constitutes a threat, *see* Grimal, *supra* note 49, at 185. Here the author notes "Disagreement between commenters typically surfaces in relation to threat categorization or threat perception, but most accept that threats are not confined to something said, but also can include something done – indeed 'actions may well speak louder than words." At 186-187 the author adds, ". . . the archetypal threat remains a coded warning/ultimatum – i.e., 'comply or else'." *Id.* at 186–187.

58. *See* Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. Rep. 226, ¶ 24 [hereinafter *Nuclear Weapons Advisory Opinion*]; *see, e.g.*, IAN BROWNLIE, INTERNATIONAL

most interest. This is not because of what the test does or does not clarify regarding the lawfulness (or not) of threats, but rather the conceptual approach adopted by the ICJ in terms of coupling threat to actual force. Indeed, somewhat similarly, the present authors also proffer a conceptual coupling. In this instance however, it relates to necessity, and the point of attack, in order to justify the calibration of pre-emption (to allow for acts to be judged instead as anticipatory) and to ensure lawfulness. As noted by the authors in their introductory remarks, the necessity requirement is therefore physically applied, calculated and forecast by the EAI in the context of the future and not the present.

B. *Self-Defense under International Law*

The previous section underscored the 'absolute' prohibition against a threat or use of force in international law. However, and as those intimately familiar with the *ad bellum* will quickly recognize, there are two permissible exceptions to the cardinal prohibition contained in Article 2(4): a state's inherent right of self-defense primarily contained in Article 51 of the U.N. Charter, and authorization of force by the United Nations Security Council in pursuance of its Chapter VII powers.[59] Extra-Charter considerations such as Humanitarian intervention and the international agreement acknowledging a Responsibility to Protect (R2P) are two concepts that are further considered in Part V. However, it is the "first" exception to the prohibition contained in Article 2(4) (that of self-defense), that is naturally the most pertinent to the authors' primary discussion. As previously noted, a state's inherent

---

LAW AND THE USE OF FORCE BY STATES 364 (1963); Grimal, *supra* note 55, at 7, 37; Drummond, *supra* note 57, at 212; Grimal & Sundaram, *supra* note 22, at 339. For a considered view of the test, see Grimal *supra* note 49, at 186 stating "The current test for determining the lawfulness of a threat of force remains the one espoused by the ICJ in its seminal *Nuclear Weapons Advisory Opinion* . . . Here, the ICJ concluded that the threat of force is unlawful if the force threatened would violate Article 2(4). Essentially, the ICJ posed a retroactive test to the following hypothetical and the contextual coupling of a threat of force to actual use of force. *If* the threat of force were carried out (in other words actual force, and not threatened force) would that actual force be lawful? If yes, that would legitimise the prior threat. If not (i.e., if actual force would be deemed unlawful), then so would the threat that preceded it."

59. U.N. Charter, *supra* note 12, at Chapter VII is intended to deal with 'Action with Respect to Threats to the Peace, Breaches of the Peace, and Acts of Aggression', and it encapsulates Arts. 39-51. Art. 42 is perhaps of the most interest here, stating: "Should the Security Council consider that measures provided for in Article 41 would be inadequate or have proved to be inadequate, it may take such action by air, sea, or land forces as may be necessary to maintain or restore international peace and security. Such action may include demonstrations, blockade, and other operations by air, sea, or land forces of Members of the United Nations."

right of self-defense is codified in Article 51 U.N. Charter. This states: "Nothing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations, until the Security Council has taken the measures necessary to maintain international peace and security."[60]

In addition to this codification, a state's inherent right of self-defense also has deeply entrenched roots within Customary International Law.[61] Indeed, most accept that self-defense "today" amalgamates both sources of Public International Law.[62] An absolute precondition to Article 51 is the explicit requirement that a state lawfully exercising its inherent right of self-defense must have suffered an 'armed attack'.[63] Although the Charter does not elaborate any further as to meaning of the term 'armed attack,' or its applicable threshold,[64] further explanation can be distilled from the judgments in the *Nicaragua* case,[65] the *Oil Platforms* case,[66] and from scholarly sources.[67] These conclusively concur that for an armed attack to meet the requisite threshold envisaged by Article 51, the force used has to be 'the most grave form of the use of force,' *i.e.*, a qualitatively grave use of force beyond a use of force simpliciter.[68]

An alternative view is that Article 51 should be applied more liberally so that it does not "override" pre-existing customary international law.[69] Under this application, states are permitted to lawfully invoke a right of anticipatory self-defense when faced with a sufficiently serious

60. U.N. Charter, *supra* note 12, art. 51.

61. *See, e.g.*, *Self-defence*, ICRC, https://casebook.icrc.org/glossary/self-defence (visited July 11, 2022), where the ICRC note its customary nature. Note also the discussion relating to the Caroline incident which follows below.

62. Grimal & Sundaram, *supra* note 22, at 326–327; Green and Grimal, *supra* note 44, at 299.

63. Don W. Greig, *Self Defence and the Security Council: What Does Article 51 Require?*, 40 INT'L & COMPAR. L. Q. 366, 366-402 (1991). It should be noted that Art. 51 of the United Nations Charter is silent in regard to imminence.

64. One argument is that art. 51 very clearly states "if an armed attack occurs. . . ." and not "after an armed attack occurs.[.] " *See* DINSTEIN, *supra* note 14, ¶ 614, at 234.

65. *Nicaragua*, *supra* note 46, ¶ 191.

66. *Oil Platforms*, *supra* note 46, ¶ 51.

67. *See, e.g.*, OLIVIER CORTEN, THE LAW AGAINST WAR: THE PROHIBITION ON THE USE OF FORCE IN INTERNATIONAL LAW 147 (Emmanuelle Jouannet ed., Christopher Sutcliffe trans.,2010); IAN BROWNLIE, INTERNATIONAL LAW AND THE USE OF FORCE BY STATES 278-79 (1963).

68. Green & Grimal, *supra* note 56, at 300; AVRA CONSTANTINOU, THE RIGHT OF SELF-DEFENSE UNDER CUSTOMARY INTERNATIONAL LAW AND ARTICLE 51 OF THE U.N. CHARTER 158 (2000).

69. *See, e.g.*, James A. Green, *Docking the Caroline: Understanding the Relevance of the Formula in Contemporary Customary International Law Concerning Self-defense*, 14 CARDOZO J. OF INT'L & COMP. L. 429 (2006). In this respect, at notes 49–40 Green cites BROWNLIE, *supra* note 67, at 275–78; Olivier Corten, *supra* note 67, at 407-11; CHRISTINE GRAY, INTERNATIONAL LAW AND THE USE OF FORCE 160–65 (3rd ed. 2008).

and imminent threat of suffering an armed attack or grave use of force.[70] Generally, the lawfulness of anticipatory action can only be established if the defensive action is compliant with requirements prescribed by former U.S. Secretary of State, Daniel Webster. Webster's infamous and much quoted correspondence with the British representative, Lord Ashburton, resulting from the *Caroline* incident, prescribed that a state must,

> *show a necessity of self-defense, instant, overwhelming, leaving no choice of means, and no moment for deliberation. It will be for it to show, also, that . . . [it] did nothing unreasonable or excessive; since the act, justified by the necessity of self-defense, must be limited by that necessity, and kept clearly within it.*[71]

Detractors against this position argue that *Caroline* should not be used to justify anticipatory self-defense, not least because there was nothing anticipatory about the nature of the British actions which Webster was referring to.[72] Nevertheless, and at the very least, the Webster formula provides an underpinning for the parameters calibrating a state's lawful response: the principles of necessity and proportionality.[73] The very essence of necessity is that if force is to be used in self-defense it must be as a very "last resort."[74] In other words, a state must demonstrate not only that it has exhausted all non-forcible measures but also that a non-forcible reasonably would be wholly unreasonable — a strongly worded letter to the Editor of the New York Times would seem a uniquely unlikely response to a nuclear attack.[75] Proportionality,

---

70. Terry D. Gill, *The Law of Armed Attack in the Context of the Nicaragua Case,* 1 HAGUE Y.B. OF INT'L L. 30, 35 (1988). *See also* DINSTEIN, *supra* note 14, ¶ 580, at 222 (noting U.S. military doctrine ensures combatants have the right to act in self-defense in response to both an "armed attack" and to a "demonstrated hostile intent"). In addition, Dinstein consents to Judge Schwebel's dissenting opinion in *Nicaragua*, *supra* note 27, which rejected the claim that a right to self-defense can only exist where an armed attack occurs. DINSTEIN, *supra* note 14, ¶ 585-86, at 224. Indeed, according to Dinstein, such an interpretation is "counter-textual, counter-factual, and counter-logical." *Id.* ¶ 586.

71. *See* Letter from Daniel Webster to Lord Ashburton, 27 July 1842 30 BSP 193, 193–94; Letter from Daniel Webster to Henry S. Fox, 24 April 1841 29 BSP 1137, 1137–8 [hereinafter *Caroline*].

72. DINSTEIN, *supra* note 14, ¶ 589, at 225.

73. Green & Grimal, *supra* note 56, at 300; *see generally* Green, *supra* note 69.

74. *See* DINSTEIN, *supra* note 14, ¶ 580, at 234.

75. *See, e.g.*, James A. Green, *The Ratione Temporis Elements of Self-Defense*, 2 J. ON THE USE OF FORCE IN INT'L L. 97, 100–01 (2015); Georg Schwarzenberger, *The Fundamental Principles of International Law*, 87 RECUIL DES COURS 9, 97 (1955); DINSTEIN, *supra* note 14, ¶ 608, at 232; Green, *supra* note 58, at 450–7; MYRA WILLIAMSON, TERRORISM, WAR AND INTERNATIONAL LAW 115

meanwhile, prescribes that the "force employed must not be excessive with regard to the goal of abating or repelling the attack."[76] Noting that a state's response need not actually mirror the initial attack and unlike its *in bello* counterpart need not be commensurate.[77] Finally, there must be reasonable temporal proximity between response and the actual armed attack.[78]

To summarize this discussion, a state may only lawfully resort to force in self-defense in two situations. First, in response to an actual armed attack, or second, where it "feels" that there is no other choice but to act anticipatorily against an imminent threat of a grave use of force. It is also worth highlighting that in the *jus ad bellum*, there is no explicit rule as to what is, or what is not, considered a lawful target.[79] The prevailing view, however, is that the doctrine of "pre-emptive self-defense" is categorically unlawful[80] (though once again, this is a viewpoint that is not entirely uncontested)[81] — noting that pre-emptive actions are those against a threat of an armed attack or grave use of force which is non-

(2009); JUDITH GARDAM, NECESSITY, PROPORTIONALITY AND THE USE OF FORCE BY STATES 6-11 (2004); Green & Grimal, *supra* note 56, at 300–02.

76. *See Constantinou*, *supra* note 68, at 159-61; Gamal Moursi Badr, *The Exculpatory Effect of Self-Defense in State Responsibility*, 10 GA. J. OF INT'L & COMPAR. L. 1, 4 (1980); David Kretzmer, *Killing of Suspected Terrorists: Extra-Judicial Executions or Legitimate Means of Defence?*, 16 EUR. J. OF INT'L L. 171, 174 (2005).

77. *Nuclear Weapons Advisory Opinion*, *supra* note 58, ¶ 5, at 361 (Higgins, J., dissenting). Also *see generally* Kretzmer, s*upra* note 76; Green & Grimal, *supra* note 56, at 301. It is, however, important to distinguish the lawfulness of a defending state's action taken during an ongoing armed attack (Garwood-Gowers so-called 'cumulative effect'). *See generally* Andrew Garwood-Gowers, *Self-Defence against Terrorism in the Post-9/11 World*, 4 QUEENS L. UNIV. OF TECH. L. & JUST. J. 1 (2004). According to one commentator, in the context of the former the position is that the responding state is placed under a temporal restriction – there must be a reasonable temporal proximity between the victim state's response and the armed attack itself. *See* Green, *supra* note 75, at 108-11 (noting that Green himself concedes that the 'reasonableness' parameter is nebulous and imprecise). Accordingly, Green suggests this area is open to interpretation along the lines of 'a context-specific appraisal of the various factors that may delay a self-defence action: intelligence gathering, initial resort to negotiation, geographical distance, and so on.. *Id.* at 116.

78. *See* Grimal & Sundaram, *supra* note 22, at 328. Here the authors note ". . . it is important to distinguish the lawfulness of a defending state's action taken during an on-going armed attack (the so-called 'cumulative effect'. . .), and instances where force is used once the armed attack has ceased; *see also* Green, *supra* note 75, at 108–16.

79. Green & Waters, *supra* note 42, at 9; *See also* DINSTEIN, *supra* note 11, at n.1382 (citing R. Bermejo-Garcia, 'Preventative Self-Defense Against International Terrorism', *International Legal Dimension of Terrorism* 177, 196 (P.A. Fernandez-Sánchez ed., 2009).

80. *See, e.g.*, Green, *supra* note 75, at 106; Paulina Starski, *The US Airstrike Against the Iraqi Intelligence Headquarters*, THE USE OF FORCE IN INTERNATIONAL LAW 504, 519-20 (Tom Ruys & Oliver Corten eds., with Alexandra Hofer, 2018).

81. *See, e.g.*, Starski, *supra* note 80, in turn citing Abraham D. Sofaer, *On the Necessity of Pre-Emption*, 14 EUR. J. OF INT'L L. 209 (2003); John Yoo, *International Law and the War in Iraq*, 97 AM. J. INT'L L. 563, 571 (2003).

imminent and temporally latent.[82] In other words, where a pre-emptive act is under consideration, the need to act is not instant and/or overwhelming, there is a choice of means, and possibly several moments or very deep breaths of deliberation before action.

As a result, the authors of the present article, for the most part, acknowledge that pre-emption, and the uncertainty as to time and place of attack stretches the elasticity of anticipatory self-defense beyond the point of no return. As noted by one commentator, "self-defence cannot be exercised merely on the ground of speculations, assumptions, expectations or fear."[83] Such legal squeamishness is, however, arguably predicated on a reluctance to stretch the concept of imminence — something the present authors "flag" here, fundamentally contest in Part III as part of their overall discussion of perpetual self-defense and self-defense by proxy, and "resolve" in Part IV with the applicable test which allows the recalibration of the necessity requirement with the point of attack. In the first instance, however, these three forms of *jus ad bellum* self-defense, and their relative lawfulness, are presented in the following graphic:

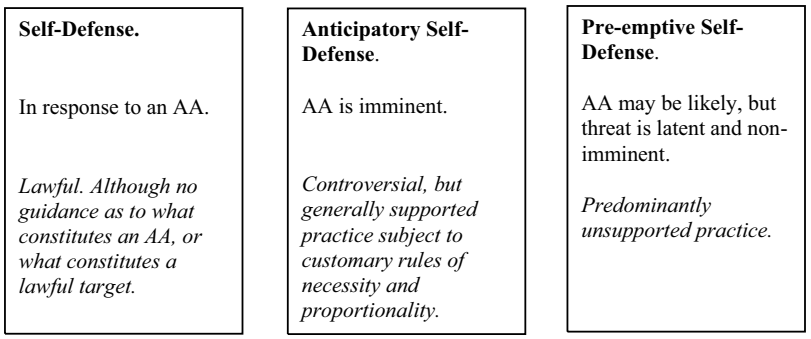

FIGURE 1: The lawfulness of defensive actions.

By way of overall summary, conventional thinking regarding the lawful application of both Charter and customary norms for self-defense predominantly precludes the possibility of pre-emptive action against a non-imminent and temporarily latent grave threat of an armed attack

82. *See, e.g.*, Monica Pinto & Marcos Kotlik, *Operation Phoenix,* THE USE OF FORCE IN INTERNATIONAL LAW 702, 704-05 (Tom Ruys et al. eds., Oxford University Press 2018). In this regard the authors also cite: Christian J. Westra, *Will the "Bush Doctrine" survive its Progenitor? An Assessment of jus ad bellum Norms for the Post-Westphalian Age*, 32 B.C. INT'L & COMP. L. REV. 399, 403-04, (2009); Michael P. Scharf, *How the War Against ISIS Changed International Law*, 48 CASE WEST. RESERVE J. INT'L L. 15, 46 (2016).

83. DINSTEIN, *supra* note 14, ¶ 614, at 234.

or grave use of force. However, one may readily question whether anticipatory self-defense is still no more than a last "throw of the dice." Due to the precarious nature of interception, the firing of a nuclear response to nuclear missiles in flight (or "fueled for flight") is arguably too late — there is nothing left to defend — and is thus punitive action at best. And yet, the law governing a state's inherent right of self-defense remains relatively uncontroversial in its absolute rejection of the much-maligned doctrine of pre-emptive self-defense. Up until now, this has perhaps been for good reason.

C. *Jus in Bello Self-Defense*

As previously noted, Part III scrutinizes influence operations in greater detail and somewhat significantly identifies that influence communications are routinely utilized by states in a way which can be seen to be a fourth method of defending their political independence,[84] and/or territorial sovereignty.[85] Nevertheless, for present purposes, the existence of an armed attack in the *ad bellum* sense, and the accompanying right to respond in self-defense, will generally imply that an armed conflict is taking place.[86] If this is the case, then self-defense actions will automatically be subject to the rules of the *jus ad bellum* and the *jus in bello* simultaneously.[87]

As a result, the following section explores the international humanitarian law (IHL) principles that could be used to restrict how influence communications is used (particularly when an indirect consequence of their use is the application of force), and those which can help to identify the lawfulness of a particular target. Key to this discussion, as it is with almost every discussion regarding IHL, are the principles of distinction and proportionality. To some extent, the duty to take

84. In addition to those represented by figure 1.

85. Violations of which are, of course, prohibited by U.N. Charter, *supra* note 12, art. 2(4).

86. Green & Waters, *supra* note 42, at 13.

87. *See* Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field (First Geneva Convention) art. 2, Aug. 12, 1949, 75 U.N.T.S 31; Geneva Convention for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea(Second Geneva Convention) art. 2, Aug. 12, 1949, 75 U.N.T.S 85; Geneva Convention Relative to the Treatment of Prisoners of War (Third Geneva Convention) ART. 2, Aug. 12, 1949, 75 U.N.T.S; Geneva Convention Relative to the Protection of Civilian Persons in Time of War (Fourth Geneva Convention) art. 2, Aug. 12, 1949, 75 U.N.T.S 287 [hereinafter the Geneva Conventions]. In all cases this states: "the present convention shall apply to all cases of declared war or of any other armed conflict which may arise between two or more of the High Contracting Parties, even if the state of war is not recognized by one of them."

precautions in attack is also pertinent — noting that under IHL the term "attack" is used for all acts of violence, whether in offense or defense.[88]

The "cardinal" principle of distinction is considered first.[89] This is both customary in nature,[90] and is codified as the "basic rule" set out in Article 48 of Additional Protocol I (API) to the Geneva conventions.[91] The basic rule states that: "In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives."[92]

Thus, Article 48 API ensures that where a state is using threats of force or actual force for the purpose of self-defense, the circumstances surrounding that act dictate that IHL is also applicable — those threats or force must only be directed at military objectives. Indeed, the ICJ has confirmed that where IHL is applicable it *must* be applied concurrently with the *jus ad bellum*.[93] With regards to Article 48 API, it should be noted that civilians are defined in the negative as being any person *not* belonging to one of the categories identified, for example, in Article 4(6) of the Third Geneva Convention.[94] In addition, Article 52(2) API identifies military objectives as: "[T]hose objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."[95]

---

88. API, *supra* note 43, art. 49(1).

89. *See Nuclear Weapons Advisory Opinion, supra* note 58, ¶ 78.

90. *See, e.g.*, JEAN-MARIE HENCKAERTS & LOUISE DOSWALD-BECK, INTERNATIONAL COMMITTEE OF THE RED CROSS: CUSTOMARY INTERNATIONAL HUMANITARIAN LAW VOL. I: RULES 3, 3-8 (2005) [hereinafter *ICRC Customary Rules*].

91. API, *supra* note 43, art. 48.

92. *Id.*

93. *Nuclear Weapons Advisory Opinion, supra* note 58, ¶ 42; *see also* Green & Waters, *supra* note 42, at 13.

94. This includes, for example, members of the armed forces, members of militias and other volunteer corps, including those of organized resistance movements (subject to the caveats in art. 4(6)). Note further that art. 50(1) API further directs the reader to art. 4 of the third Geneva Convention, and art. 43 API. The latter of these two states: "All organized armed forces, groups and units which are under a command responsible to that Party for the conduct of its subordinates, even if that Party is represented by a government or an authority not recognized by an adverse Party. Such armed forces shall be subject to an internal disciplinary system. . . ." API, *supra* note 43, at 23.

95. API, *supra* note 43, art. 52(2).

Resultingly, civilians and civilian objects must not be the object of attack or reprisal.[96] Civilian Objects are also defined in the negative — being all those that fall outside of the preceding definition.[97] Where there is doubt as to the status of the target, the person or object must be presumed to be civilian in nature.[98] Importantly, though civilians and civilian objects must be distinguished, and must not be directly targeted, they may lawfully form a part of a "targeteer's" collateral damage assessment.[99] The indirect targeting of civilians and civilian objects is subject to the IHL principle of proportionality, which is codified within Article 51(5)(b) API, and Article 57(2)(a)(iii) API. As noted by the authors elsewhere,[100] neither of these two provisions utilize the term proportionality. Nevertheless, the proportionality balance is clearly visible in the text of Art. 51(5)(b) and Art. 57(2)(a)(iii) API (below) and there is generally no disputing the claim that the principle is also customary in nature.[101] IHL proportionality differs from the *ad bellum* principle with the same designation previously considered. The IHL incarnation seeks to ensure that an attack must not: "cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated." [102]

Resultingly, acts of self-defense which are also defined as attacks in the *jus in bello* sense must not only ensure that the force employed is not excessive with regard to the goal of abating or repelling the attack, but they must not also cause excessive civilian losses. Indeed, where the two disciplines are concurrently applicable, IHL provides the *lex specialis* that must be looked to in the first instance when assessing the lawfulness (or not) of a target.[103] This is supported by the ICJ, which has routinely declared that a state acting in self-defense must only attack "legitimate lawful targets" (this latter term reflecting those military objectives identified above).[104] This rules' legal basis is actually

96. *Id.* art. 51(6).

97. *Id.* art. 52(1).

98. For objects *see id.* art. 52(3); and for the definition of the civilian population *see id.* art. 50(1).

99. In other words, civilians and civilian objects may be indirectly targeted lawfully. Though, see the following discussion regarding proportionality.

100. *See* Grimal & Pollard (2020), *supra* note 3, at 526.

101. *ICRC Customary Rules*, *supra* note 90.

102. API, *supra* note 43, art. 51(5)(b), 57(2)(a)(iii).

103. Green & Waters, *supra* note 42, at 15.

104. *Oil Platforms*, *supra* note 46, ¶ 51; *Nicaragua*, *supra* note 46, ¶ 237; *Nuclear Weapons Advisory Opinion*, *supra* note 58, ¶ 22; Case Concerning Armed Activities on the Territory of the Congo

unclear,[105] though arguably it is somewhat of a moot point because state practice typically follows suit. Nevertheless, given that this matter has already received considerable scholarly analysis,[106] the present authors do not see a need to pursue this discussion further for the sake of the present article.

Instead, the most relevant approach to consider and support here is that of "concurrent application" as identified by Green and Water's.[107] This is, not least, because in conformity with the ICJ judgements regarding target selection,[108] the principle of concurrent application ensures that all relevant *jus ad bellum*, and *jus in bello* principles are applied, to all targeting decisions, at all times. This will be the case regardless of the *lex specialis,* or the existence of any "grey areas" in which one of the two legal disciplines does not appear to apply.[109] The following graphic is intended to represent the principle,

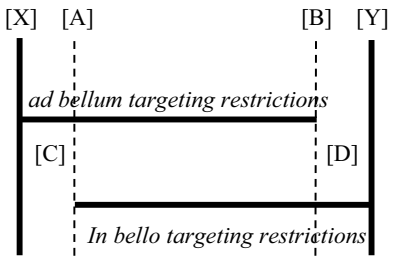

FIGURE 2: The Principle of Concurrent Application.[110]

(Democratic Republic of the Congo v. Rwanda), Judgment, 45 I.L.M. 562, ¶ 147, (Sept. 182002) [hereinafter *Armed Activities*]. *See also* Green & Waters, *supra* note 42, at 4.

105. Green & Waters, *supra* note 42, at 5.

106. *See generally* Green & Waters, *supra* note 42.

107. *Id.*

108. Perhaps most notably see the views expressed by the ICJ in *Oil Platforms*, *supra* note 46. In short, here the court refers to *jus in bello* measures as a method of restricting acts of jus ad bellum self-defense.

109. *See generally* Green & Waters, *supra* note 42.

110. The above image was originally provided by Green & Waters, *supra* note 42, at 22. There, the authors identify that in the majority of circumstances involving the use of force for self-defensive actions, an act will fall between points A and B – in other words, both legal disciplines will naturally be applicable. However, they also note that there is the potential for some self-defense acts to fall within the "grey" areas C and D. Here, without acknowledging a need for applying concurrent application, lawful targets can only be identified according to either one of the two regimes (the *jus ad bellum* or *jus in bello*), but not both. Thus, without concurrent application, it is possible that a nation may not necessarily be lawfully obligated to adhere to *all* humanitarian obligations.

It should perhaps be firmly noted that IHL does not place a restriction upon the use of threats of force — at least against lawful combatants. Indeed, the warning "put down your weapon or I'll shoot," is arguably a humanitarianly preferable request — even if it is not a necessary one.[111] Combatants *are* prevented from using threatening behavior towards the civilian population, though in law these are generally limited to acts which are intended to spread terror.[112] IHL also prevents an armed force from starving civilians,[113] or from targeting or rendering an object useless that is vital to the civilian populations survival[114] — noting these Article 54(2) restrictions do not apply if the objects contained therein are being used by an adverse party.[115] However, even where they are being used for military purposes, the constant care obligation[116] seeks to ensure that those responsible for planning or deciding upon attacks shall, for example, do everything feasible to avoid causing civilian harms.[117]

For a wider discussion regarding the constant care obligation, the duty to take precautions in attack, and the use of EAI in armed-conflict, the authors respectfully invite the reader to peruse a previous article by

111. Noting that it is prohibited to order that no quarter will be given. *See ICRC Customary Rules*, *supra* note 90.

112. *See* Fourth Geneva Conventions art. 33; API, *supra* note 43, art. 51(2); API, *supra* note 43, art. 4(2)(d); API, *supra* note 43, art. 13(2). *See also ICRC Customary Rules, supra* note 90.

113. API, *supra* note 43, art. 54(1).

114. *Id.* art. 54(2).

115. *Id.* art. 54(3).

116. *Id.* art. 57(1).

117. *Id.* art. 57(2). The full text of art. 57 (2) API states; "With respect to attacks, the following precautions shall be taken: (a) Those who plan or decide upon an attack shall: (i) Do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection but are military objectives within the meaning of paragraph 2 of Article 52 and that it is not prohibited by the provisions of this Protocol to attack them; (ii) Take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects; (iii) Refrain from deciding to launch any attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated; (b) An attack shall be cancelled or suspended if it becomes apparent that the objective is not a military one or is subject to special protection or that the attack may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated; (c) Effective advance warning shall be given of attacks which may affect the civilian population, unless circumstances do not permit."

the current authors.[118] However, one pertinent notion that is discussed there, and one which is of distinct relevance to the present discussion, is that commanders may be *required* to use an AI or EAI system if it can be identified as a means or method of attack that offers the greatest method of protecting the civilian population.[119] Nevertheless, while the previous discussion is based entirely in the *jus in bello*, the principle of concurrent application would also transpose it into realm of self-defense, because "feasible precautions" are an implicit factor of any proportionality assessment.

In summary, Part II began by identifying the Article 2(4) prohibition on the use or threat of force. It continued by identifying the exceptions to that prohibition — collective action under Chapter VII powers, and self-defense. Because this discussion is clearly intended to be grounded in the latter, it went on to consider self-defense, and identified three recognized "types." Of these, only self-defense in response to an armed attack, and/or anticipatory self-defense are generally considered to be lawful. Furthermore, Part II demonstrated how the *jus in bello* must also play a constructive role in establishing the lawfulness of self-defense actions. Key to the discussion, however, is the fact that pre-emptive self-defense is not a widely supported concept. Nevertheless, with the continuing introduction of increasingly advanced EAIs squarely in mind, this is something the present authors will proceed to challenge in Parts III and IV.

III. Influence Communications, Propaganda and Self-Defense

Arch strategists such as Machiavelli have been known to positively endorse deceit providing that it is for the "greater good."[120] Thus, the recourse to "influence operations" as a form of strategic defense is nothing inherently new. As a concept, influence communications has many names,[121] the most notable of which is perhaps "propaganda." This is defined by the Oxford English Dictionary as "the systematic

118. *See generally* Grimal & Pollard (2020), *supra* note 3 (discussing the use of EAI in armed-conflict).

119. API, *supra* note 43, art. 57(2)(ii). For further discussion, *see also* Grimal & Pollard (2020), *supra* note 3, at 678–86.

120. For a useful summary of Machiavellian Philosophy *see generally Niccolò Machiavelli*, Stanford Encyclopedia of Philosophy (May 28, 2019), https://plato.stanford.edu/entries/machiavelli/#PrinAnalPowe (visited Jan. 10, 2022).

121. *See, e.g.*, Nicholas J. Kane, *Defense Against Weaponized Information: A Human Problem, Not Just A Technical One*, 10 Interagency Journal 46, 48–49 (2019). Here the author identifies five types of existing information operation capabilities; "Operations Security, Electronic Warfare, Cyberspace Operations, Military Information Support Operations – formerly called psychological

dissemination of information, esp. in a biased or misleading way, in order to promote a political cause or point of view."[122] Indeed, the practice of spreading information (whether true or false) in order to gain a political or military advantage has been so widely employed throughout the centuries that the authors believe it should be treated as a "customary norm" (at least in a strategic sense).

With that in mind, the term "perpetual self-defense" is utilized hereinafter in order to describe this form of persistent influence. Somewhat significantly, international law does not currently prohibit perpetual self-defense, though there is an increasing scrutiny of digital forms of misinformation, disinformation, and hate-speech (MDH) within the literature.[123] Moreover, the *jus ad bellum* has not traditionally provided the legal framework under which it should be scrutinized. This section, however, examines the consequences of recourse to influence communications by state actors as a fourth method of defending their sovereignty.

In the first instance, arguably, the ultimate purpose of propaganda is to influence an individual's cognitive, rather than their physical behavior (at least when considering the direct results).[124] And, given that winning the "hearts and minds" of the local civilian population in order to support a military operation may be a tactic that is as old as armed conflict itself,[125] much has been previously considered in this area. It is fairly well-settled, for example, that from an IHL perspective, unless advocating for international crimes, civilian propaganda remains civilian (and a normal part of every conflict).[126] Despite this, in 1964 Arthur

operations – and Military Deception," noting also the link between these and the "older art" of disinformation and propaganda. S*ee also* Winther, *supra* note 37, ¶ 5.

122. *Propaganda*, OXFORD ENGLISH DICTIONARY, https://www.oed.com/view/Entry/152605?rskey=hyddW4&result=1&isAdvanced=false (last visited Jan. 10, 2022).

123. The most recent edition of the ICRC journal – the international review – is entirely focused upon the impact of digital technologies in armed conflict. *See generally ICRC Digital Technologies and War*, 102 ICRC, no. 913 (2020) [hereinafter ICRC REVIEW 2020].

124. Winther, *supra* note 37, ¶ 5.

125. The expression "hearts and minds" (in the military context) is believed to have first been used by British General Sir Gerald Templer while operating in Malaysia in February 1952. *See Gerald Templer: The smiling tiger*, NATIONAL ARMY MUSEUM, https://www.nam.ac.uk/explore/gerald-templer-smiling-tiger (last visited Jan. 10, 2022). The tactic of gaining the support of the local population, however, was also somewhat infamously adopted by U.S. armed forces operating in south Vietnam. *See, e.g.*, Robert J. Kodosky, *What's in a Name? Waging War to Win Hearts and Minds*, 32 (1) AMERICAN INTEL. J. 172, 173 (2015).

126. Indeed, exiting scholarship regarding propaganda is generally keen to highlight the reasoning of the ICTY in the prosecutor's report regarding NATO bombings of a Yugoslavian TV station. With regard to lawful target selection, the court identified that "If the media is used to incite crimes, as in Rwanda, then it is a legitimate target. If it is merely disseminating propaganda

Larson somewhat prophetically noted that "propaganda is one of the most dangerous sources of international friction and war, and that there is every reason to believe that it will get much worse."[127] Part III very much confirms this Nostradamus — like caution becoming reality —and highlights the fact that more recently, influence communications are perpetually applied not only by civilian sources who generally escape the obligations imposed by international law, but also by specific military units who arguably should probably not.[128]

A. *Overview of Influence Communications*

As an independent strategic concept, propaganda can take many guises and its effects can be diverse. For example, until relatively recently,[129] the internationally accessible BBC World Service was funded by the British Foreign and Commonwealth Office (FCO).[130] This is particularly pertinent because the FCO is a government office which openly states that its aim is to "pursue … national interests and protect the UK as a force for good in the world … promote the interests of British citizens, safeguard the UK's security, defend our values[.]"[131] Arguably, it is somewhat inconceivable that the UK Government would fund an organization in opposition to these stated objectives. One might perhaps choose to endorse the viewpoint that the dissemination of a pro-democratic, nonpartisan, and pro-human rights news feed (and one which includes shipping forecasts!) does not qualify as propaganda — especially, if one recalls the definition of propaganda presented above. However, certain

to generate support for the war effort, it is not a legitimate target." Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia, ¶ 47, https://www.icty.org/x/file/Press/nato061300.pdf (last visited Jan. 10, 2022).

127. Arthur Larson, *The Present Status of Propaganda in International Law*, 33 (3) L. AND CONTEMP. PROBS., 439, 439–451 (1966).

128. S*ee generally* Winther, *supra* note 37; Kane, *supra* note 121.

129. Since 1 April 2014 the BBC World service has been funded by the "License fee," as opposed to the FCO. The corporation's website notes that while the BBC is no longer funded directly by a government department, it is nevertheless still funded by legislative taxes that are imposed upon the vast majority of U.K. homes and businesses that own TV's. *See About World Service Radio*, BBC, https://www.bbc.co.uk/programmes/articles/5nCxH0NlsPtyW8WvJ0rwDJP/about-world-service-radio (last visited Jan. 10, 2022).

130. The FCO now incorporates the U.K. department of international development and is thus referred to as the Foreign Commonwealth & Development Office (FCDO). *See, e.g.*, U.K. Government, https://www.gov.uk/government/organisations/foreign-commonwealth-development-office (last visited Jan. 10, 2022).

131. *See, e.g.*, U.K. Government, https://www.gov.uk/government/organisations/foreign-commonwealth-development-office/about (last visited Jan. 10, 2022).

government figureheads in several of the states in which the World Service is broadcast are unlikely to wholeheartedly agree.[132]

The point here is that if a state does choose to utilize such methods for promoting its best-interests and/or "grand strategy," then one might also reasonably posit that such "shameless" self-promotion can also be seen as a method of defending against "threats to its political independence." As previously noted, threats of this type are prohibited by Article 2(4) U.N. Charter, though these will be subject, of course, to severity.[133] Nonetheless, in relation to this use of propaganda as an ongoing "defensive" mechanism, the present authors coin the term "perpetual self-defense." This is accompanied by an acknowledgement that, at least insofar as genuinely independent news sources are concerned, "forceful measures" are generally unlikely to occur as a direct consequence of their broadcast.

As a form of perpetual self-defense (and dependent to subjective appraisal . . .) the BBC World Service would perhaps be best positioned at one end of a second sliding scale with the designation of "least-harmful" propaganda. In contrast, as identified by organizations such as the International Committee of the Red Cross (ICRC), the exponential increase in the spread of digital technologies is allowing for the increasingly rapid spread of communications which look to disseminate MDH. The distinct problem is that in contrast to the pre-digital era,[134] social media platforms now provide a single individual with an opportunity to

132. While the BBC world service may broadcast content in states which have autocratic leaders who might also, for example, be guilty of human rights abuses, it is perhaps unlikely that government officials form that state will agree and acknowledge the news source is legitimate – regardless of how honest the report is.

133. *See* Green & Waters, *supra* note 42, regarding the fact that there is no absolute threshold test in art. 2(4), or, indeed, any reference to the concept of 'armed attack,' as per U.N. Charter art. 51.

134. In the 1980's, for example, the Soviet clandestine organization the Komitet Gosudarstvennoy Bezopasnosti (KGB), ran a disinformation program codenamed Operation Denver (also referred to as Operation Infektion). In an attempt to develop anti-US global sentiment, the KGB claimed that HIV/ AIDS was manufactured in a US military laboratory. However, after 4 years and a great deal of effort, these claims perhaps reached only hundreds of thousands of people. *See, e.g.*, Douglas Selvage & Christopher Nehring, *Operation "Denver": KGB and Stasi Disinformation regarding AIDS*, WILSON CENTER (July 22, 2019), https://www.wilsoncenter.org/blog-post/operation-denver-kgb-and-stasi-disinformation-regarding-aids (last visited Jan. 10, 2022), and Peter W. Singer, *Webinar at the ICRC DigitHarium Digital Dilemmas Dialogue Webinar #3: How the spread of harmful information changes armed conflicts*, ICRC (Mar. 24, 2021), https://www.icrc.org/en/digitharium/digital-dilemmas-dialogue-3?utm_medium=email&_hsmi=117420606&_hsenc=p2ANqtz-_G4d2qR2UaCMWTfhVVVQH1OIUmeMl4N6apsso6bMeWZjnNj3Ke8naGqPBXASqniuKININNpPsXx4BCRn2aSzJHKMiUtQiGO_a-rkMJtN6TLWUubgg&utm_content=117420606&utm_source=hs_automation (last visited Jan. 10, 2022).

reach and influence as many, if not more individuals, than traditional platforms such as regulated news outlets. In fact, in crisis contexts such as Myanmar, South Sudan, and Ethiopia, MDH has been disseminated via social media platforms, and public opinion has been manipulated based on false or incomplete information, which may have exacerbated the humanitarian crises at hand.[135]

As a genus of propaganda, MDH would therefore be placed at the opposite end of the sliding scale to legitimate news feeds — being of the type that are potentially the "most harmful." Indeed, the ICRC believe the spread of MDH may have long-lasting negative humanitarian consequences. For example, where individuals, or groups of individuals, are persuaded to behave in a certain manner, such as targeting a minority group with violence or threats thereof. Moreover, it might also be the case that the consequences of spreading MDH will stretch beyond the direct infliction of harm, to a wider displacement of targeted groups should they feel it necessary to flee their community in order to escape the resulting violence and/or persecution.[136]

FIGURE 3: The scale of risk of lasting harms when deploying influence communications.

135. Saman Rejali & Yannick Heiniger, *in ICRC Digital Technologies and War*, 102 ICRC, no. 913, at 9 (2020). A deepfake is defined as "a video of a person in which their face or body has been digitally altered so that they appear to be someone else, typically used maliciously or to spread false information." *See Deepfake*, LEXICO, https://www.lexico.com/definition/deepfake (last visited Jan. 10, 2022).

136. *See generally* Andrew Hoskins, *in ICRC Digital Technologies and War*, 102 ICRC, no. 913, at 122 (2020). Also see the accompanying Webinar, which regularly refers to this topic; Helen Durham & Bruno Demeyere, *ICRC Webinar: Digital Technologies and Humanitarian Action in Armed Conflict*, ICRC (Mar. 18, 2021), https://www.icrc.org/en/event/digital-technologies-and-humanitarian-action-armed-conflict-global-conversation-convened?utm_campaign=DP_ORE%20Events&utm_medium=email&_hsmi=117400047&_hsenc=p2ANqtz-_R2UN_wbiE9Pwig4HQyshZ6mDBLnV-7bbkRcHShF7YfEFFOQB7w3H1smTCAHpCKZBjf-vNBoB0HeFiS65Fo-q8DuMzC1aucPy98ooy6g0-iX4Lrs8&utm_content=117400047&utm_source=hs_email (last visited Jan. 10, 2022).

B. *Influence Communications and International Law*

In the strategic sense, the spreading of information is generally referred to as "influence communications," or "influence operations."[137] These have been defined for the US Army by the Rand Institute, for example, as the

> *coordinated, integrated, and synchronized application of national diplomatic, informational, military, economic, and other capabilities in peacetime, crisis, conflict, and post-conflict to foster attitudes, behaviors, or decisions by foreign target audiences that further . . . [national] . . . interests and objectives.*[138]

For present purposes, the focus is placed upon the crisis and conflict element of the previous passage. But, when this definition is considered holistically, it is clear to see why the ICRC have warned that influence operations can potentially cause a great deal of lasting damage when they are "deployed" in conflict zones — particularly where the influencer's objective is the spreading of MDH.[139] There is one specific element that is contained within the RAND definition however, that the present authors wish to accentuate, and, moreover, to utilize in order to support the current (primarily *jus ad bellum*) investigation. It is, as RAND correctly point out, that influence can also apply post conflict. Clearly, the above definition alludes to the "possibility" of influencing hearts and minds "*post bello,"* and not to any kind of obligation to do so. Nevertheless, as previously noted, for present purposes the *jus post bello* has an intrinsic part to play.

As a concept, the *jus post bello* continues to gain traction in contemporary discussions regarding armed conflict — though it is perhaps important to point out that it is grounded in the historical domain of Just War Theory, and not, as is the case for the *ad bellum* and *in bello*, in international law.[140] Nevertheless, at its heart, it asserts that states must consider their moral obligations in the stages following armed-conflict,

---

137. Often shortened to influence comms and influence ops respectively.

138. ERIC V. LARSON, RICHARD E. DARILEK, DANIEL GIBRAN, ET AL. FOUNDATIONS OF EFFECTIVE INFLUENCE OPERATIONS: A FRAMEWORK FOR ENHANCING ARMY CAPABILITIES 155 (2009).

139. Indeed, the authors gratefully acknowledge the importance of the dialogue that the ICRC has initiated in this regard and agree wholeheartedly that this is an area that is in urgent need of greater forensic analysis.

140. The authors note that while the *ad bellum* is reflected by the U.N. Charter, and the *in bello* by the Geneva Conventions (and supporting customary obligations), there is nothing in existence yet (codified or customary obligations) for the *post bello* realm.

such as, for example, the responsibility to rebuild infrastructure damaged as a result of war.[141] Brian Orend is one such scholar who acknowledges that the "theory" has previously been neglected.[142] In his reflections however, he posits that further consideration could ultimately lead to a *post bellum* treaty — and thus an establishing of the concept in law.[143]

His argument is grounded in the undeniable truth that every war has a beginning, middle, and end.[144] He believes, furthermore, that all three of these phases should be reflected by both just war theory and by law.[145] It is fair to say the debate has since moved forward, and without any such *post bello* legal obligation materializing. Nevertheless, as will be demonstrated in Part IV, the concept is particularly pertinent when considering the continuing introduction of AI, and other "digital technologies." The primary reason for this, as the authors have argued elsewhere,[146] is that AI and particularly EAI, are already capable of conducting assessments, legal or otherwise, using far more information than humankind alone could ever be capable of achieving. And, if one is to advocate in favor of a limited use of the doctrine of pre-emptive self-defense, as this Article does, the preceding humanitarian and *post-bellum* assessments quite simply help to ensure that all available evidence is considered when taking a decision to use force (specifically in regard to self-defense).

As previously noted, if judging EAIs by contemporary standards, some will undoubtedly consider the following claim to be both controversial and unconvincing. There is, for example, sufficient literature highlighting the fact that modern AI systems cannot be trusted to accurately distinguish, for example, between a soldier, and a child carrying a toy gun.[147] Nevertheless, the present authors believe that future EAIs *will* eventually be more capable than their human counterparts at making such distinctions and at determining when there is a "necessity" to act against a (future) threat.[148] As a result, due to continuing advances in military technologies, future instances *will* arise in which states

141. *See generally* Brian Orend, *Jus Post Bellum*, 20 LEIDEN J. INT'L. L. 571 (2007).

142. *Id.* at 573–74.

143. *Id.* at 571.

144. *Id.* generally, where the author regularly speaks of a need to create a new Geneva Convention designed to deal solely with *post bellum* "problems and values."

145. *Id.* at 573–74.

146. *See generally* Grimal & Pollard (2020), *supra* note 3.

147. *See, e.g.*, BONNIE DOCHERTY, 'LOOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS 31 (2012).

148. In particular, *see* Grimal & Pollard (2020), *supra* note 3.

should be permitted to utilize potentially harmful influence communications as a method of acting preemptively against a "latent" and non-imminent threat.

A vital acknowledgement with regard to the present discussion, is that the increasing manufacturing and distribution of EAIs is very likely to lead to heightened national security concerns such as those which recently materialized due to the "alleged" influence the Chinese government has over the 5G provider Huawei.[149] Somewhat crucially however, as the roll-out of increasingly advanced EAIs in to the civilian realm continues, would-be aggressors will be provided with an opportunity to bypass national infrastructure and scrutiny systems altogether — and yet potentially be able to achieve the same or even better results. EAIs *already* have the capacity to act autonomously. Even relatively simple exiting systems could readily be programmed with the secondary clandestine purpose of spreading information, including MDH. This could be achieved in two primary ways, either via a set of pre-programmed rules, or as a result of the machines own "learning" capabilities.[150] As inferred above, where a state is responsible for placing a EAI into a third party state with the intention of spreading information in order to discredit, or "topple," an incumbent, the authors believe it should be seen as a breach of political independence, and territorial integrity — both of which are emphatically prohibited by Article 2(4) U.N. Charter.[151]

On the flip side, and as the authors have argued elsewhere, the introduction of EAIs also represents an opportunity to remove some (though perhaps not all) elements of human error (in conflict often referred to as "Clauzewitzian frictions") which can lead to breaches of international law.[152] And just as crucially, influence communications and the spreading of propaganda is certainly not a practice that is

149. *See, e.g.*, David Shepardson & Karen Freifeld, *Trump extends U.S. telecom supply chain order aimed at Huawei, ZTE* REUTERS (May 13, 2020), https://www.reuters.com/article/us-usa-trade-china-trump-idUSKBN22P2KG; Leo Kelion, *Huawei 5G kit must be removed from UK by 2027*, BBC (July 14, 2020), https://www.bbc.co.uk/news/technology-53403793. Regarding the spread of Chinese influence in the West *see, e.g.*, *International Security and Estonia 2021*, ESTONIAN FOREIGN INTEL. SERV. 73–76 (2021), https://www.valisluureamet.ee/doc/raport/2021-en.pdf.

150. For a useful introduction to Machine Learning *see* Jason Brownlee, *Basic Concepts in Machine Learning*, MACH. LEARNING MASTERY (Aug. 15, 2020), https://machinelearningmastery.com/basic-concepts-in-machine-learning/.

151. U.N. Charter art 2(4), *supra* note 12.

152. *See, e.g.*, Grimal & Pollard, *supra* note 3, at 673. Here the authors note: CARL VON CLAUSEWITZ, ON WAR 138, (Michael Howard & Peter Paret trans., 1976, rev. ed. 1984). Clausewitz identifies "Friction is the only concept that more or less corresponds to the factors that distinguish real war from war on paper." Though the authors' previous discussions have taken

employed solely by autocratic and/or quasi-democratic states. Some may consider it insensitive to place various weapons upon a scale delineating potential harms to the civilian population. However, it is nevertheless arguable that in certain circumstances, it would be less invasive, and indeed less destructive to spur on a local population with the intention that they topple or oust their incumbent leadership, either democratically, or by other means. Where an EAI has calculated all outcomes and concluded that, for example, a Head of State *will* be a future aggressor, then an influence comms operation might be considered a preferable pre-emptive action, as opposed to a lawful high altitude bombing campaign after an act of aggression has actually occurred.

As previously discussed, one elementary advantage of EAI, is that it is already capable of operating in ways which are outside the limits of human comprehension. Moreover, an EAI can evaluate a vast number of alternative courses of action and can also choose which one is preferable under the given circumstances.[153] In short, while an EAI may not be able to accurately predict "the" future, it might one day be able to predict *all* possible futures. And, when an EAI is able to simultaneously consider all possible outcomes, they will also be capable of determining whether a pre-emptive action would be considered a proportionate course of action due to the fact that wider civilian harms (whether at "home" or extraterritorially) will be minimized as a result of acting sooner, rather than later. The fundamental question posed and answered by the authors, is that if all the "moves" have been considered, and all the potential outcomes are known, pre-emption *should* be permitted in the Machiavellian sense — in that it is for the greater good.

There is clearly an ethical element to this discussion, and further analysis in this respect would clearly be beneficial.[154] In fact, some have

place in regard of *in bello* principles, in this instance, this claim is carried over to the *ad bellum* realm.

153. Indeed, this is one of the key strengths of AI, not least because of the vast amounts of data that are widely available and easily accessible in the age of the internet, and the fact that algorithms are not prone to the same cognitive biases as humans. *See, e.g.*, Eric Colson, *What AI-Driven Decision Making Looks Like*, HAR. BUS. REV. (July 08, 2019), https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like.

154. In 2014, MIT set up and ran a comprehensive ethical/ moral experiment regarding lethal decision-making in autonomous vehicles referred to as the MIT Moral Machine. It offered participants a choice between killing elderly people, young people and/or pets, though there were other categories. This can still be accessed at https://www.moralmachine.net. However, for an overview of the results see Karen Hao, *Should a self-driving car kill the baby or the grandma? Depends on where you're from*, MIT TECH. REV. (Oct. 24, 2018), https://www.technologyreview.com/2018/10/24/139313/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem/.

already shown particular discomfort when faced with the prospect of algorithms making life and death decisions in armed conflict.[155] For present purposes however, the discussion is grounded in law, and in particular the law of self-defense. The ethical discussion is still relevant in this regard, however, because self-defense is one of only two contemporary concepts that are considered to be compatible with just war theory — a theory which seeks to qualify (or not) the morality of resorting to force.[156]

One potential problem with emerging technologies is that they are being developed in dynamic environments, and by tech companies that often have substantial financial backing. In contrast, so the argument goes, international diplomacy, and the crafting of applicable regulatory regimes can be slow and perhaps underfunded.[157] Regardless of whether new legal obligations will arise in the future, or whether indeed they are needed in terms of regulating new technologies, where advanced digital technologies are used for military purposes, they must adhere to existing legal norms.[158]

The primary purpose of the following examination, therefore, is to consider the wider question as to whether there are any existing legal obligations regarding the use of digital technology as a method of intentionally manipulating a civilian population to resort to force in the hope that they indirectly cause harm to an adversary.[159] To denote this form of indirect application of force, the authors wish to coin the second term "self-defense by proxy." And, in relation to it, a secondary, perhaps more focused question is: can existing legal obligations help to determine what should be considered lawful, and conversely unlawful, "targets"?

---

155. *See, e.g.*, Robert Sparrow, *Robots and Respect: Assessing the Case Against Autonomous Weapons Systems*, 30 1 ETHICS & INT'L AFFAIRS, 93 (2016). Here the author consistently refers to AWS as 'evil in themselves.' Also, *see generally* CHRISTOF HEYNS, AUTONOMOUS WEAPONS SYSTEMS: LIVING A DIGNIFIED LIFE AND DYING A DIGNIFIED DEATH, AUTONOMOUS WEAPONS SYSTEMS: LAW, ETHICS, POLICY (2016). The author, the former U.N. Special Rapporteur on extrajudicial, summary or arbitrary executions, routinely refers to machine life and death decision making as undignified.

156. The other being collective action for humanitarian purposes or subject to chapter VII powers. These two concepts are considered in greater detail in Part V.

157. *See generally* ICRC REVIEW 2020, *supra* note 123.

158. See, e.g., *Principle 1, U.N. Group of Governmental Experts on Lethal Autonomous Weapons Systems*, U.N., https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/ (last visited Jan. 10, 2022) stating, IHL "continues to apply fully to all weapons systems, including the potential development of Lethal autonomous weapons systems."

159. This is typically considered via the use of social media platforms, but it could equally apply to alternative medias such as television and radio, and to emerging digital technologies that are not yet common-place or do not exist at all

In the first instance, as referenced in Part II, an "influencing state" may only claim to be acting in self-defense if it has suffered an armed attack. Perhaps a somewhat interesting aside, is that Article 42 U.N. Charter, which "guides" the United Nation Security Council (and not of course states acting in self-defense *per se*), fails to specify what forms of actions are permitted (in the interest of international peace and security)—recognizing only "such action by air, sea, or land forces as may be necessary."[160] In fact, it is at least arguable that acts of influence do not currently represent force, but instead fall under the banner of Article 41 U.N. Charter—being lesser "measures not involving the use of armed force . . . [such as the] . . . interruption of . . . telegraphic, radio, and other means of communication[.]"[161]

Customary *jus ad bellum* however, may prevent militaries from deploying influence communications as a method of self-defense due to the fact that it may be difficult to align the temporal considerations relating the act of influencing a population, with *jus ad bellum* necessity.[162] Additionally, under current conditions, it may also prove difficult to render the act of intentionally targeting a civilian population in this manner as proportionate—especially where influence communications are "deployed" in the knowledge that civilian harms are very likely to result. As noted, there is no reference to lawful targets *per se* under the *jus ad bellum* in any situation,[163] and thus certainly not in regard of self-defense by proxy.

By applying the principle of concurrent application, however, the greatest range of humanitarian protections can be offered — even where, as noted, IHL is not necessarily triggered. When seeking to apply direct force, an armed force must adhere to IHL — including, not least, the key tenants of distinction and proportionality. However, other than the prohibition of acts or *threats* of violence that spread terror, there is no reference to indirect applications of force. Additionally, as noted above, the *jus in bello* does not prohibit the

160. U.N. Charter art. 42, *supra* note 12.

161. U.N. Charter art. 41, *supra* note 12, states "The Security Council may decide what measures not involving the use of armed force are to be employed to give effect to its decisions, and it may call upon the Members of the United Nations to apply such measures. These may include complete or partial interruption of economic relations and of rail, sea, air, postal, telegraphic, radio, and other means of communication, and the severance of diplomatic relations."

162. The point being that necessity requires that if force is to be used in self-defense it must be as a last resort. In other words, any non-forcible measures that would be considered a reasonable alternative in the circumstances must have been explored and exhausted first

163. *See generally*, Green & Waters, *supra* note 42.

use of propaganda,[164] though it may place certain restrictions on it. As argued by the ICRC, for example, Article 14 of the third Geneva Convention protects the morality and physical welfare of prisoners of war (POW), and prohibits the use of propaganda when is likely to adversely affect such morality in the long run.[165] In addition, as identified by Pontus Winther,[166] Article 51(5) GC IV makes reference to propaganda, and prevents occupying powers form using such means to secure the services of the local civilian population.[167] It is also generally accepted that the broadcasting of images of prisoners of war (POW) for humiliation purposes is inhumane.[168] But aside from these somewhat narrow circumstances, references to propaganda are scarce under IHL.[169] As a consequence, when it is utilized as a general means or method of "attack" (albeit an indirect one), the dissemination of propaganda/influence comms, including that which seeks to spread MDH, is relatively unregulated.[170]

Kearney believes that though the *jus and bellum* and *jus in bello* appear to fail to regulate or limit the use propaganda for the purposes of conducting war, international human rights law (IRHL) does provide numerous ways to do so.[171] Somewhat vitally for present purposes, he notes in particular Article 20 International Covenant on Civil and

164. *See, e.g.*, API, *supra* note 43, art. 37(2) which states, "Ruses of war are not prohibited. Such ruses are acts which are intended to mislead an adversary or to induce him to act recklessly but which infringe no rule of international law applicable in armed conflict and which are not perfidious because they do not invite the confidence of an adversary with respect to protection under that law. The following are examples of such ruses: the use of camouflage, decoys, mock operations and misinformation."

165. *See generally* JENNIFER K. ELSEA, CONG. RSCH. SERV., RL32567, LAWFULNESS OF INTERROGATION TECHNIQUES UNDER THE GENEVA CONVENTIONS, (2004), [hereinafter U.S. CONGRESS REPORT 20004]; INTERNATIONAL COMMITTEE OF THE RED CROSS, COMMENTARY ON THE GENEVA CONVENTIONS OF 12 AUGUST 1949 145 (Jean Pictet, ed. 1960) [hereinafter "ICRC COMMENTARY III"].

166. Winther, *supra* note 37, at ¶ 11.

167. Geneva Convention (IV) Relative to the Protection of Civilian Persons in Time of War art. 51, Aug. 12, 1949, 75 U.N.T.S. 287 stating "The Occupying Power may not compel protected persons to serve in its armed or auxiliary forces. No pressure or propaganda which aims at securing voluntary enlistment is permitted."

168. *See* U.S. CONGRESS REPORT 20004, *supra* note 165, at CRS-19 n.84, stating "During the invasion of Iraq in 2003, both Houses of Congress passed resolutions condemning as inhumane and humiliating the broadcast of interrogations of U.S. POWs. H. Con. Res. 118, 108th Cong. (2003); S. Con. Res. 31, 108th Cong. (2003)."

169. *See generally* U.S. CONGRESS REPORT 20004.

170. *See generally id.*

171. *See generally* MICHAEL KEARNY, THE PROHIBITION OF PROPAGANDA FOR WAR IN INTERNATIONAL LAW (2007).

Political Rights (ICCPR) which provides the following: "*Any propaganda for war shall be prohibited by law.*"[172] With paragraph 2 of the same instrument also stating that "*Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.*"[173] Prima facie, these obligations appear to greatly restrict the way nations might be able deploy propaganda — especially when provoking violence in the purpose of the communications. However, as Kearney himself refers to, these *lex generalis* obligations are transcended by the *lex specialis* nature of self-defensive actions, or other acts provided for by the U.N. Charter.[174] Indeed, though IHRL continues to apply in armed conflict,[175] IHL must also take a lead role. Moreover, discrimination would need to be a necessary element for it to be considered a breach — which would of course not always need to be the case.

International law does place further limitations upon incitement — as codified in Article 20(2) ICCPR.[176] Indeed, as noted by Kristin Timmerman,[177] a leading example of an unlawful incitement to violence is the actions of President of the Croatian Democratic Union of Bosnia and Herzegovina, Dario Kordic. In 1993, Kordic ordered and incited violence against the Muslim community that the ICTY found amounted to nothing less than Crimes Against Humanity.[178] A similar instance occurred a year later in Rwanda, when the broadcaster Kantano Habimana used his "on-air" position to incite the violence that led to the massacres and ultimately the commission of genocide.[179]

While these tragic examples, and others besides, can be used to highlight the possibility that influence communications *could* be used (unlawfully) to incite violence, that violence however must be connected to the commission of an international crime in order for it be considered unlawful. However, proving the existence of such a nexus is further complicated where the influencer claims to be acting in self-

172. *See* International Covenant on Civil and Political Rights (ICCPR) art. 20(1), Dec. 19, 1966, 999 U.N.T.S. 171.

173. *See id.*, art. 20(2).

174. KEARNEY, *supra* note 171, at 4.

175. *See Nuclear Weapons Advisory Opinion*, *supra* note 58, ¶ 25.

176. ICCPR, *supra* note 173.

177. Wibke Kristin Timmermann, *Incitement in International Criminal Law*, 88 ICRC REV. 864, 824–25 (2006).

178. Prosecutor v. Ferdinand Nahimana, Jean-Bosco Barayagwiza and Hassan Ngeze, Case No. ICTR-99-52-T, Judgement and Sentence (Trial Chamber) ¶ 1096 (3 December 2003).

179. Prosecutor v. Kordic and Cerkez, Case No. IT-95-14/2-T, Judgement (Trial Chamber), ¶ 834 (26 February 2001).

defense, and not as an aggressor.[180] In addition, an influencer's *mens rea* must also be established, not least because international law has a "serious blind spot where crimes of recklessness are concerned."[181]

The reality is, therefore, that self-defense by proxy is largely uncatered for by international law.[182] This may not come as a surprise, given that it is somewhat of a novel and possibly ambiguous concept. In addition, it is a form of self-defense in which the influencer is clearly removed (both physically and potentially temporally) from the actual application of force — albeit a force which the influencer had intended. However, arguably, the more technically savvy, and better equipped they become, the worlds' militaries will undoubtably find increasingly complex ways of utilizing influence comms, and of manipulating both humans and complex systems to achieve their strategic, operational, and tactical goals. As it stands, however (as a method of acting in self-defense), there is currently very little guidance, and virtually no restrictions placed upon the use of influence comms for military purposes.

IV. Test and Scenarios

In the preceding section the authors identified a number of novel concepts relating to the use of EAIs for military operations, and in particular as a method of self-defense (self-defense by proxy, and perpetual self-defense in the form of influence communications). It is the authors' firm belief that by controversially fusing these elements, a much-needed debate as to the lawfulness of military operations (present and future) via EAIs is uniquely opened. Taken to its conclusion, the authors' assert that the use of pre-emptive action via an EAI is not only strategically desirable but should be lawfully acceptable — though admittedly only under certain circumstances. Overarchingly, such a discussion is conceptually possible if one accepts the authors' proposed solution of coupling the necessity assessment and the "point of force," rather than the "point of predicted force."

In other words, the authors argument is grounded upon the notion that the necessity requirement of self-defense can be physically applied and forecast in the context of the future, and not of the present. This tenable position is defensible on the grounds that an EAI would have

180. In other words, where it claims to be resorting to the use of force in response to a breach of art. U.N. Charter art. 2(4), *supra* note 9.

181. Jens David Ohlin, *The Combatant's Stance: Autonomous Weapons on the Battlefield*, 92 Int'l L. Stud. Ser. US Naval War Col. 1, 21–22 (2016).

182. *Id.*

calculated every conceivable (and non-conceivable) move (and countermove) thus reconciling its calculation with the requirement of "last resort." As previously noted, some will undoubtably find it inconceivable that a machine will ever be capable of predicting all outcomes with 100 percent accuracy. And, as technology stands, that is an utterly defensible position. Nevertheless, as the present authors have repeatedly demonstrated both here and elsewhere, EAIs are no longer fanciful concepts of science-fiction.[183] Instead they are real, tangible objects, and there is no question that "tomorrow" will see significant advancements in both their physical and cognitive capabilities. Firmly in the belief that EAIs will eventually be cable of operating in an increasingly intelligent manner therefore, the following section "stress-tests" the temporal repositioning of the necessity requirement and appraise the future circumstances under which an EAI should be lawfully permitted to take pre-emptive action.

A. *The Authors' Test*

Before proceeding to conduct the stress simulations alluded to above, the authors believe that an EAI should be required to conduct, and successfully pass, an initial test before a pre-emptive action can even be considered. In this respect, Part II of the present analysis is key, and in particular, the emphatic prohibition of threats of force under Article 2(4) U.N. Charter. To recall, in the Part II discussion the reader was also guided to one of the present author's previous works highlighting the importance of identifying the strategic nature of threats. With that in mind, before an EAI should be permitted to lawfully authorize pre-emptive self-defense, and/or self-defense by proxy, it must first identity a threat in breach of Article 2(4), and one which presents a "clear and present danger." This is necessary because it may be unlikely that a threat to launch an apocalyptic strike would breach the strategic threshold proposed where the entity making the threat clearly had no access to, and no likelihood of, developing such a capability.

Nonetheless, central to the authors' present discussion is the unwavering belief that future EAI *will* be capable of calculating the Schelling requirements more readily — and act only where a state will be capably and committed to credibly communicating and "carrying out" the threat.[184] Indeed, a fundamental difference between EAI and human assessment is that an EAI will be inherently more capable of predicting

183. *See* Grimal & Pollard (2020), *supra* note 3; Grimal and Pollard (2021), *supra* note 3.

184. SCHELLING, *supra* note 55.

future capabilities rather than only those which a potential adversary currently possess.

Phase [1] of the diagram below, therefore, represents the instance at which an EAI identifies a future threat, and the likelihood (or not) of that threat being viable. This is also the point at which the EAI must determine whether the future threat in question satisfies (at that future point in time) the requisite threshold of a grave threat of an armed attack or use of force. Once a viable threat has been identified, phase [2] (on the diagram below) represents the necessary re-positioning of *jus ad bellum* necessity from a temporal perspective as previously discussed. That is, the authors' proposed solution of coupling the necessity assessment and the "point of force" (be it in three months or three years into the future), rather than the "point of predicted force." However, before any action can be authorized, the EAI must establish that action — at *that* point in time — constitutes one of "last resort."

Once phases [1] and [2] have been established/completed, the EAI must then determine the lawfulness of the intended target. This determination begins with an EAI assessment of the intended target (in light of the principle of concurrent application as per figure 2 above). The principle of concurrent application can be identified on the graphical representation below, as area [Z] (noting this does not extend beyond broken lines A and B). In short, the diagram shows that any intended course of action that is being considered by the EAI must only target military objectives. Moreover, where civilian harms are likely to occur as a result of acting, the EAI may only authorize an act where such harms are not excessive in relation to the concrete and direct military advantage anticipated (implying that the EAI must also be capable of conducting such an assessment). Here, as noted above, concurrent application means that where there is a potential gap in target identification, an alternative discipline can "step in" to ensure civilian protections are maximized.[185]

To entrust self-defensive decision-making responsibilities to a machine is, of course, highly controversial — even where the force applied is indirect in nature and is a result of "mere" information manipulation. Additionally, one might reasonably contend that if decisions regarding the application of force (of any kind) are to be delegable to an EAI, then the technology should bring something "extra to the table." Indeed, from an ethical perspective at least, if EAI tech is to be utilized, its "presence" should not merely replace human decision-making like-for-like. With that in mind, the authors propose that an

185. Green & Waters, *supra* note 42, at 22.

EAI should not just do the "bare minimum" in terms of ensuring everything "feasible" has been done to ensure the limitation of civilian harm—as arguably that is what a human combatant is required to do.[186] Instead, if EAI are to be harnessed for self-defense purposes, they should be capable of considering *all* variations, and of calculating each and *every* outcome in light of prevailing circumstances. As previously alluded to, and further evidence in the scenarios below, the present authors believe that it is *only* when an EAI is furnished with *all* the pertinent information in the *pre-bello* sense, that they can act pre-emptively.

Admittedly, such responsibility is not to be taken "lightly." This is not least because of the potentially infinite number of "real-world" variables that an EAI would need to calculate and consider. This is in direct contrast, for example, to that of a sixty-four squared boardgame, where the limited number of "participants" is known in advance, and each of their movements is severely restricted — if not entirely predictable. Indeed, as neural networks advance in complexity, it may become increasingly more difficult to gauge how, or why, an AI reaches its final decision or "output." Ultimately, however, questions regarding the issue of "trust" will need posing in relation to the prevailing technology, and of the human perception towards the tech at the moment in time in which a particular delegation of power is being considered. Key to the present Article is the not so unreasonable submission that humankind will come to accept, and even rely upon AI "prediction" — even where it is not furnished with the knowledge as to "how" a particular decision has been reached. And, when an EAI can demonstrate that it can reliably and more accurately predict the future(s) — it will undoubtably sway (at least some of) those who are currently pessimistic.

With that in mind, the authors propose that each EAI assessment must include an *ad bellum*, an *in bell*o, and a *pre-bello* analysis. Yet, as was noted at very opening of the discussion, the only way to appease natural and understandable detraction is via recalibration at both ends of the legal continuum. Consequently, for reasons established in Part III, the authors propose that the *jus post bellum* should also be accounted for as part of the pre-emptive analysis. This should be the case irrespective of the fact that there is no legal obligation to do so. This decision can be lawfully justified because the test for self-defense by proxy actions can give *post bello* considerations a quasi-legal nature, given that they are included as a part of the wider proportionality assessment. This is manifested in the following graphical representation,

186. API, *supra* note 43, art. 57.

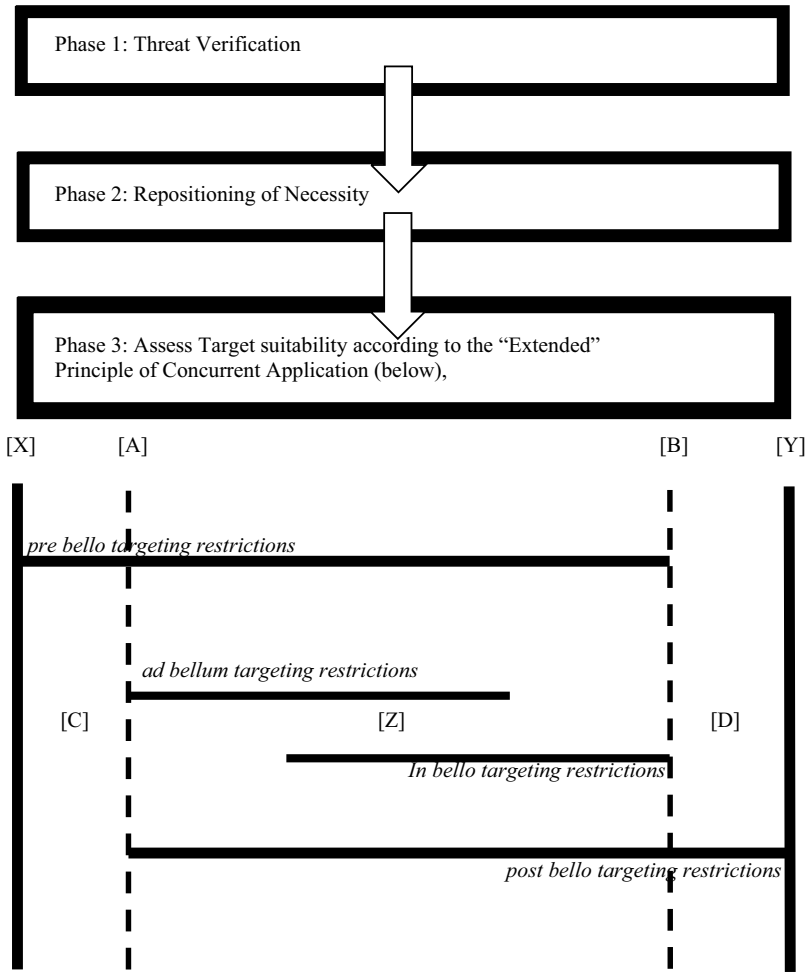

FIGURE 4: Authors test for utilizing EAI for the purpose of acting in perpetual self-defense and/or self-defense by proxy.

In this "wide" application of concurrent application, [X] now represents the outer limits of the *pre-bello* requirement. And, though inherently connected with [X], [Y] now represents the opposing scope of the EAI calculation — the *post bello* considerations. As previously noted, these may include the matter of whether the effects of a pre-emptive act could be considered proportionate to the overall objective (*e.g.*, the foreseeable harms to the local civilian population, and the likelihood of being able to reconstruct the damage caused to essential national infrastructures). [A] and [B] symbolize the outer reaches of the "narrow" application of concurrent application — which is represented holistically by area [Z]. When considered in isolation, [C] depicts the

pre-bello decisions which sit at the limits of the *ad bellum*, and *in bello*. Area [D] reflects the *post bellum* (non-legal) obligations, which when considered alongside the *ad bellum* and *in bello* can also be said to reflect just war theory.[187] By adopting this "wide" version of concurrent application, all obligations are applied throughout the construct. Thus, area [Z] flows into zones [C] and [D], in which both [X] and [Y] are concurrently applicable. Consequently, the utilization of EAI for acts of self-defense by proxy can be used as a method of extending existing humanitarian protections further than is currently required. The primary point, however, is that when making pre-emptive targeting decisions, an EAI must be furnished with *all* necessary information and to make a holistic determination accordingly.

To put theory into practice the remainder of Part IV considers a single scenario, with four possible and alternate courses of action. In the first analysis, this scenario is considered relative to current self-defense interpretations — which restrict states from acting pre-emptively (and perhaps, with good reason). The second and third "courses of action" examine whether the outcome would be altered if the defending state had recourse to perpetual self-defense and self-defense by proxy, respectively. And, finally, the fourth course of action (albeit briefly) considers whether an EAI could, or perhaps should, be permitted to authorize a direct application of force for self-defense purposes.

B. *Scenario*

Europa is an island state, and other than Ganymede — a small island 30 miles to the North-West — is relatively isolated geographically. In contrast, Hegemone, a state located approximately 5000 miles to the North-East of Europa, is a closed-border state, though it is located on the wider continent of Megaclite. Hegemone and Europa are long-time adversaries, and both nations have a strategic nuclear capability. The state of Thebe (Hegemone's neighboring state) typically aligns itself with Hegemon but is not a strategic ally in the truest sense. Europa, however, considers the remaining states that constitute Megaclite [X, Y, and Z] to be strategic allies. A diagrammatic representation of this fictional geographic area appears as follows:

187. *See generally* Kane, *supra* note 121.

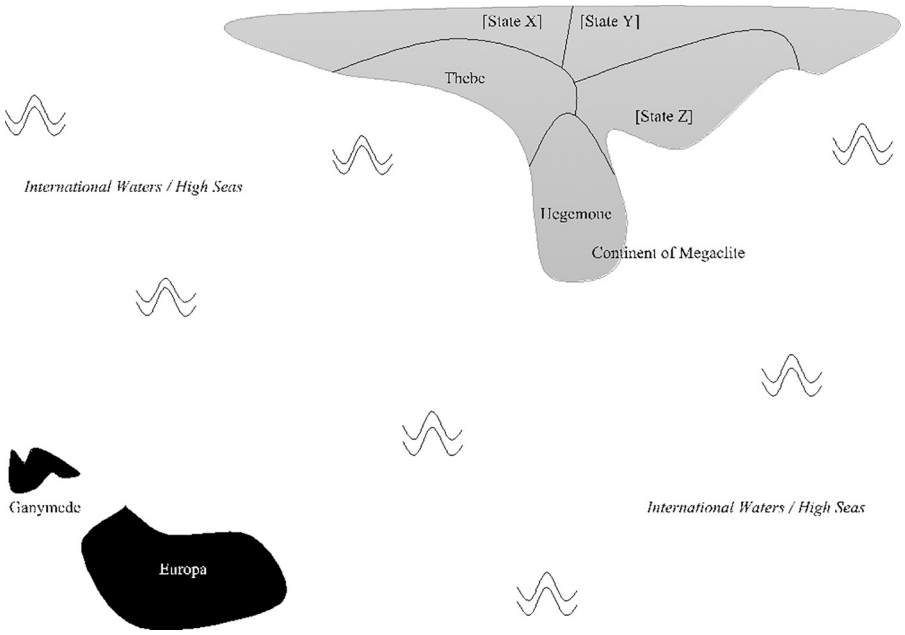

FIGURE 5: Diagrammatic Representation of the Scenarios Geographic Area.

Over the last 15 years, Hegemone has significantly developed its armed forces and armories. Moreover, Hegemone's long-serving Head of State has periodically stated that its primary strategic goal is to achieve world dominance in the realm of emerging technologies which it intends to continue to develop for both military and civilian purposes. Hegemone's military leaders have also regularly stated that it intends to create the world's largest armed forces, which it will deploy if necessary to ensure its long-term strategic goals are met.

Hegemone leaders have regularly and publicly called Europa an enemy and one that the world would "be best rid of." Those in power have also recently warned Europa to "be prepared." Europa has recently suffered a number of cyber-attacks. And, although Europa does not believe that any of these should qualify as "armed attacks," they have, nonetheless, caused considerable disruptions to public services, including transport, energy production, and communications networks. In recent weeks, Europa has seen a significant increase in both the number and the veracity of these cyber-attacks, and some have directly targeted Europan military installations.

Europan leaders have solid intelligence that Hegemone is the source of these attacks. The intel also indicates that the attacks are investigative

in nature, and may form part of a wider military operation. In recent weeks Hegemone has also begun to assemble large amounts of troops and military hardware in various places — not least in Ganymede territory to the Northwest (even though Europa considers Ganymede to be an ally. Finally, Hegemone Naval vessels and aircraft have recently also been operating increasingly close to Europan territory, in a manner which Europa believes is intentionally provocative. Hegemone sources refuse to clarify why these military operations are taking place in these regions, but have stated that they are merely part of an ongoing military training exercise.

C. *Analysis 1*

The purpose of "Analysis 1" is to examine the above scenario in relation to existing interpretations of a state's inherent right of self-defense as codified by Article 51 U.N. Charter. Doing this creates a baseline discussion, while highlighting the inherent strategic difficulties with the *lex lata*. This examination therefore considers both the stricter application of self-defense under Charter norms, and the wider and more permissive customary international law right to act in anticipation of a grave threat of force/armed attack. As a reminder, Article 51 is paramount in its requirement that a state has suffered an *actual* armed attack amounting to a grave use of force, whereas customary international law prescribes that the threat of an armed attack (amounting to a grave use of force) must be imminent.

This discussion commences by considering the authors' test provided in figure 4 above. This is because under the present interpretation of imminence, it is unlikely that phase [1] could be satisfied or justified — given the nature, or threat level, currently displayed by Hegemone. If Article 51 is applied literally (as certain commentators believe it should be),[188] then it does not matter how *threatened* Europa's leaders feel — there is simply no armed attack to lawfully respond to. Therefore, if (in this scenario) Europa resorted to force, or even a threat thereof, it would be in breach of the "exit velocity" of Articles 2(3) and Article 2 (4) U.N. Charter (providing the prerequisite thresholds were surpassed and noting that it is generally accepted that cyber-attacks could, at least potentially, qualify as armed attack, where physical damage resulted).[189] In addition, a strict and overly regimented interpretation of Article 51

188. *See, e.g.*, DINSTEIN, *supra* note 14, ¶¶ 580–91.

189. *See generally* Grimal & Sundaram, *supra* note 22.

prohibits Europa from taking any kind of anticipatory (let alone pre-emptive) action.

As noted in Part II, a more flexible and less restrictive interpretation of a states' inherent right of self-defense can be found in the deep-seated roots of customary law — captured by the concept of anticipatory self-defense. Proponents arguing in favor of the lawful right to have recourse to anticipatory action allows so on the grounds that the threat threshold is that of an imminent and grave threat of an armed attack — leaving "no choice of means and no moment for deliberation."[190] Though perhaps not irrevocably settled, the concept of imminence as noted in Part II is generally interpreted in accordance with the correspondence ensuing from the *Caroline* Incident,[191] later restated in the *locus classicus* of the seminal Merits Judgment of the ICJ in the *Nicaragua* Case.[192]

When considering the present scenario along this second pathway, it is highly probable that a Europan use, or indeed threat, of force as a form of self-defense would nevertheless be unlawful.[193] In this scenario, even if there is an imminent threat of a grave use of force, Europan leaders are not privy to the precise nature and location of the "unknown unknown."[194] Moreover, though inherently intertwined, Europa also appears to have a moment of pause for deliberation, and a choice of means as to how they would choose to act. In either case, and were Europa to act, it would likely be doing so pre-emptively, and thus in breach of both Charter and customary norms.

Although this wisdom is generally accepted, the current authors struggle to reconcile it absolutely. Undoubtedly, the prevailing critique against this wider and more tolerant right of anticipatory self-defense, is that in many instances Europa would only be permitted to act when strategically it may already be too late. This is a particularly prominent and already well-versed discussion in relation to the "nuclear

190. *See Caroline, supra* note 71 (Daniel Webster's formulation of the argument).

191. *Id.*

192. *See, e.g.*, *Nicaragua*, *supra* note 46, ¶ 194.

193. *See, e.g.*, Green & Grimal, *supra* note 56.

194. U.S. Secretary of Defense Donald Rumsfeld, Press Conference at the NATO Conference (June 6, 2002), stating "The message is that there are no "knowns." There are things we know that we know. There are known unknowns. That is to say there are things that we now know we don't know. But there are also unknown unknowns. There are things we don't know we don't know. So when we do the best we can and we pull all this information together, and we then say well that's basically what we see as the situation, that is really only the known knowns and the known unknowns. And each year, we discover a few more of those unknown unknowns." (Transcript available at https://www.nato.int/docu/speech/2002/s020606g.htm).

option."[195] In this respect, Hegemone would realistically have to severely disrupt Europa's "center of gravity,"[196] and prevent it from launching its own nuclear capabilities. In short, Hegemon would need to overwhelm Europa, or else risk leaving itself open to catastrophic damage.

The authors hastily add that it is not their intention to further analyze such aspects of nuclear strategy, but to simply highlight that once the "missiles are in the air,"[197] the stark reality may be that there is nothing left to defend. As a result, any response other than actions taken to intercept Hegemone's incoming ICBMs, might actually be little more than retaliatory, or reprisal-like. Even if this were a lawful response,[198] one must question how useful such an act would be to Europa, given the scenario under consideration. Highly subjective and straying into the realm of "right intention," it could be suggested that Europa is acting for the greater good in attempting to defend its allies by removing Hegemone's nuclear capability. Yet, even if one overlooks the legal obligation under collective self-defense (which requires specific nomination of third parties)[199] it is still likely to be too little and all too late for the citizens of Europa. As a consequence of both current technology and the legal implications of that technology, a state is perhaps

195. *See generally* COLIN S. GRAY, NUCLEAR STRATEGY AND NATIONAL STYLE (1986).

196. A term (appearing in German language form as "Schwerpunkt") identified by CLAUSEWITZ, *supra* note 152, at 673. The present authors consider this in further detail in Grimal & Pollard (2021), *supra* note 3, at 673.

197. *See* Francis Grimal, *Missile Defence Shields: Automated and Anticipatory Self-defense?*, 19:2 J. CONFLICT & SEC. L. 317 (2014).

198. *See* DINSTEIN, *supra* note 14, ¶ 696. Here the author states an act of reprisal is not, per se, unlawful – though he notes that it is limited by jus in bello principles. However, at ¶ 708, citing Brownlie, *supra* note 58, at 281, Dinstein also acknowledges that "most writers deny that self-defense pursuant to Article 51 may ever embrace reprisals." DINSTEIN, *supra* note 14, ¶ 708. It is also generally understood that it is unlawful to retaliate to an act of retaliation. *See also Rule 145*, in ICRC CUSTOMARY RULES, *supra* note 90, which notes a "belligerent reprisal consists of an action that would otherwise be unlawful but that in exceptional cases is considered lawful under international law when used as an enforcement measure in reaction to unlawful acts of an adversary. In international humanitarian law there is a trend to outlaw belligerent reprisals altogether." Moreover, art. 51(6) in API altogether prohibits reprisals aimed at the civilian population.

199. As one of the authors has noted elsewhere, "while some commentators are less persuaded by the idea that the state need necessarily declare itself as the victim of an armed attack, a request for assistance—which undeniably is a requirement for lawful collective self-defence." Grimal, *supra* note 49, at 191; *see also* James A Green, *Editorial Comment: The "additional" criteria for collective self-defence: request but not declaration,* 4:1 J. ON USE OF FORCE & INT'L L. 4 (2017); *see also* CHRISTINE GRAY, INT'L L. AND USE OF FORCE BY STATES 187 (4th ed. 2018) ("In every case where a third state has invoked collective self-defence it has based its claim on the request of the victim state. . . .").

rightfully prohibited from acting on a whim, or on an assumption, no matter how fearful they might be. And, in this instance, Europa might be hard pushed to show an immediate necessity to act.

D. *Analysis 2*

The purpose of Analysis 1 was to demonstrate how the inherent right of self-defense is currently interpreted and/or restricted, and that consequently under the existing legal framework, anticipatory action (while a first resort of sorts) might still be "too little too late." This second analysis, however, considers those limitations, and the above scenario, in light of the concept of perpetual self-defense (as coined by the present authors) — the additional strategic method, and in certain instances a tactical tool, for defending political independence and territorial sovereignty. As previously noted, this examination begins by considering phase [1] of the authors' test — which requires threat verification. The primary question here is whether observers are capable of positively identifying that Hegemone's actions are ultimately going to manifest either as a threat of force, or as an armed attack, at some predetermined stage in the future?

If so (though clearly it is not the only course of action Europa could take), one way it could respond to the current threat would be to launch an influence communications operation (with a view to interfering with future events in order to alter the outcome). Typically, and while military influence operations are generally restricted to the battlefield, many states already have specific military influence units operating outside of traditional battlespaces in a perpetual effort to protect a state's strategic interests and their political ideals.[200] Admittedly, these actions are generally not designed to apply "force," or even the threat thereof. Thus, the use of influence comms would not typically breach Article 2(4), or even require the state employing them (in defense) to have suffered an armed attack (or believe that an armed attack amounting to a grave use of force is imminent). Nevertheless, so long as dedicated military units exist, one can convincingly conclude that states do believe contemporary influence comms operations are both a necessary

200. For obvious reasons, a degree of secrecy must be maintained over the operational parameters of such units. However, in the United Kingdom, the Army's 77th Brigade claims it aims to "challenge the difficulties of modern warfare using non-lethal engagement and legitimate non-military levers as a means to adapt behaviours of the opposing forces and adversaries." *77th Brigade: Influence and Outreach,* ARMY: BE THE BEST, https://www.army.mod.uk/who-we-are/formations-divisions-brigades/6th-united-kingdom-division/77-brigade/ (last visited July 11, 2022). *See generally* U.S. ARMY, THE U.S. IN MULTI-DOMAIN OPERATIONS 2028 (2018).

and a proportionate (long-term) method of protecting their best interests.

As previously noted, influence communications/information/propaganda can take many forms, and it can be propagated via the use of several "platforms." One such platform could be television, another radio, and perhaps even a poster and/or leaflet campaign. Moreover, and somewhat significantly, a state may seek to influence any target group it chooses, including its own citizens.[201] These platforms, or channels, may all be beneficial in certain ways. However, in the age of the internet, perhaps nothing can compare to the power of social media platforms. This is not only because the alternative campaigns noted above will undoubtably require some form of financial outlay, as well as varying degrees of temporal expenditure (though this may not always be relevant). Instead, it has more to do with the fact that when utilizing such methods, the influencer can never be certain how many individuals the campaign might actually reach. In contrast, social media posts are cheap to produce (potentially without cost), and crucially, they are capable of personally reaching millions of individuals located all over the planet — almost instantaneously. In many cases one can even determine whether the intended recipient(s) has opened the "message," and/or responded to it in some way. With that in mind, social media platforms such as Facebook and Twitter are the method of influence communications that the present authors wish to place their primarily focus.

In the current scenario (as is the case with real world operations), Europa could apply perpetual self-defense in any number of ways (either with or without the use of EAIs). One such "defensive" method Europa could attempt would be to influence the citizens of Hegemone. An influence operation of this kind may have a long-term goal of demonstrating the "benefits" of a particular way of life — be it democratic, autocratic, or something else. The point is, Europa would attempt to exert influence with the intention of persuading the Hegemone civilian population to change their own future. This is perhaps the *modus operandi* of contemporary influence communications operations. And,

201. As was the case, for example, in the U.S. in World War II. Indeed, the national World War II Museum writes "Over the course of the war the U.S. government waged a constant battle for the hearts and minds of the public. Persuading Americans to support the war effort became a wartime industry, just as important as producing bullets and planes. The U.S. government produced posters, pamphlets, newsreels, radio shows, and movies-all designed to create a public that was 100% behind the war effort." *We Can Do It! Propaganda Posters Emphasizing War Production*, NAT'L WWII MUSEUM, http://enroll.nationalww2museum.org/learn/education/for-students/ww2-history/take-a-closer-look/production-propaganda-posters.html (last visited Jan. 10, 2022).

being a typical and widely practiced behavior, it once again appears to be generally accepted. Somewhat importantly, the practice appears to be beyond the scope of the *jus ad bellum* framework. Moreover, in respect to the scenario under consideration, it would also appear that an operation of this type would be too far removed temporally from the apparent "threat."

An alternative operation might target what appears, prima facie, to be Hegemon allied states — Ganymede and Thebe. In the first instance, Europa may wish to gently remind Ganymede of their positive historical ties, and the dangers of inviting Hegemone forces into the region. They might do this directly, using sincere factual information. Or they may choose to spread MDH as an alternative method. Suffice to say, the dissemination of information, whether true or false, is key to the concept of perpetual self-defense. However, the present authors believe that dedicated military influence communications units should refrain from circulating hate speech under any circumstances, not just those prohibited because they incite the commission of an international crime. The matter of whether they should be permitted to disseminate material which encourages violent behavior falling short of genocide, crimes against humanity and/or war crimes, however, is an area in urgent need of greater analysis. Moreover, further examination will also be required with regards to extreme situations where the acts of self-defense may be "desperate" (though this is of course what the authors are trying to avoid by shifting "necessity"). Nevertheless, if one coldly applies the logic of the ICJ in acknowledging there may be instances where the use of a nuclear weapon could be justified —a line of argument might also advocate those situations "in which … [the state's] … very survival would be at stake" the spreading of MDH might also be justified.[202]

Nevertheless, an operation of the kind under consideration could target state officials, armed forces, the civilian population, or all the above and more besides. However, the point here would be to introduce and amplify feelings of hatred, distrust, and contempt toward the Hegemone presence — noting of course, that the spreading of MDH is generally not looked upon favorably, though it is not presently unlawful.[203] An alternative Europan influence communications strategy might target Thebe officials and attempt to "encourage" them to take advantage of the fact that Hegemone's military might is focused elsewhere. This may, for example, present Thebe with an ideal opportunity

202. *Nuclear Weapons Advisory Opinion*, *supra* note 58, ¶ 97.

203. Other than those instances noted.

to launch an operation to reclaim a disputed territory. Once again, this could be accomplished in any number of ways, and target any number of individuals, or groups. However, if it was successful, it could result in Hegemone armed forces departing the Europa/Ganymede region, if only temporarily.

Of course, there would be many ways by which Europa could utilize influence communications in order to try to prevent a future Hegemone use of force. Moreover, there are many potential effects of using influence comms. Yet, given the details of the present scenario (though latent and non-imminent), the threat may be too immediate for many of the longer-term options to be successful. Analysis 2 has nevertheless successfully demonstrated that perpetual self-defense is generally a non-forceful method of defending one's best interests. Generally, although in each case considered Europa could be said to have acted pre-emptively, if force did occur, it would likely be unintended and/or unforeseeable. Indeed, even if it was a favorable outcome, the application of force would be too far removed physically, temporally, and in terms of intent, from the influence comms operation.[204] Thus, in this situation, action is very unlikely to be considered either an unlawful act of pre-emptive self-defense and/or a breach of Article 2(4) U.N. Charter.

E. *Analysis 3*

The third analysis considers the same scenario as before, but, under the premise that Europa is in possession of an EAI that is programmed to monitor Hegemone behavior, and to autonomously predict and alter future Hegemone actions. At the heart of this discussion, is the question as to whether perpetual self-defense should be allowed to continue in its present unregulated form where a foreseeable, and indeed intended, outcome of its use is the application of indirect force. As previously noted, the analysis begins by reminding the reader of the application of the authors' proposed test. In this regard, the following examination also continues under the presumption that a legitimate threat has been verified. In other words, the Europan EAI phase [1] investigation has identified that a particular Hegemone course of (threatening) action(s) will ultimately result in an armed attack be carried out against Europa.

204. In this sense, self-defense by proxy could be compared to a cyber-attack. The point being, must there be an immediate kinetic action to be considered a breach of art. 2(4), or is it more related to the consequences suffered.

The crucial difference between this and the previous analysis is that here, the authors attempt to peer forwards into some (undetermined) focal point in the future. In this future version of the same scenario already considered, Europa's armed forces are able to utilize several (what are presently considered to be) emerging technologies. These include EAIs, a number of which have been distributed to designated military influence comms units and are currently deployed to autonomously conduct perpetual self-defense appraisals. Certain EAIs are capable of utilizing deep neural networks to monitor and predict Hegemone behavior. In addition, these are programmed to predict the future effects of their own actions or inactions. These assessments could be said to be pre-bello in nature and are therefore those which are identified by area [C] in figure 4. As with others that are circulated into society more widely, these future EAIs have proven to be very successful at completing the tasks assigned to them. As a result, they have been delegated decision-making responsibilities in order to authorize and control the operational direction of influence comms operations. Having established that a Hegemon armed attack is forthcoming, one European EAI clandestinely and pre-emptively initiates several countermeasures.

As previously noted, anticipatory (and pre-emptive) actions presently revolve around the concept of imminence. A state must show a necessity of self-defense that is instant, overwhelming, leaving no choice of means, and no moment for deliberation. It must also show that action was necessary and that it did nothing unreasonable or excessive. Today, it may seem utterly unreasonable to claim that an attack was imminent ten years from the moment at which it was anticipated. However, the authors wholeheartedly believe that this will not always be the case. Indeed, under current understanding, ten months, ten weeks, and even ten days may be considered too early to lawfully invoke the right to anticipatory self-defense. However, even by current standards, at the lower end of this timeframe, an AI is arguably much more adept at predicting the correct outcome(s), and of initiating a course of action to alter the future in its favor (as it does when playing the game of chess).

In essence, the authors argue the following. If, when considering all future possibilities, a Europa EAI can calculate that Hegemone actions have passed a threshold where an attack on Europa's "center of gravity" is inevitable without further action, then the EAI should be permitted to act. To do this, the necessity requirement must be conceptually and temporally shifted from the point of threat to the point of action so as to negate non-compliance of "last resort" (phase [2]). Then, providing all other (phase [3]) requirements are present and correct (meaning

the intended target is a lawful one according to the wide model of concurrent application), the authors believe that an EAI should be, and is perhaps even already, lawfully enabled to take pre-emptive action.

Self-defense by proxy actions (*i.e.*, ones that sought to apply an indirect force by way of influence comms) would need to be considered proportionate under the circumstances presented in the scenario, especially when compared to various other methods of applying a direct force. This is, perhaps, even more apparent when considering an EAIs ability to operate "outside the box" of human comprehension, given that they will consider certain elements of each potential action and reaction that would likely escape the vast majority, if not all, human commanders. Nonetheless, EAIs will effectively carry out each and every assessment with ultimate due diligence. Traditional methods of utilizing influence communications do not have to adhere to the wider principle of concurrent application, as the authors propose an EAI would. Indeed, as discussed in a previous article, if where they can be utilized in an armed conflict, and where an EAI self-defense by proxy action is likely to minimize civilian harms, when compared to alternative means and methods of warfare, those who plan or decide upon attack may be lawfully obliged to use the EAI.[205]

If the authors' reasoning is accepted, then the proposed test can regulate EAIs used for this purpose. And, where force that would otherwise be a breach of Article 2(4) was a foreseeable consequence of initiating a self-defense by proxy action, the test would ensure the greatest range of humanitarian considerations were applied. In short, any action authorized by the EAI, having satisfied the *jus ad bellum* requirements, must also be directed only at military objectives [Z]. And, where civilian harms are anticipated, they must not be excessive in relation to the concrete and direct military advantage anticipated (also [Z]). This is the case, whether or not an armed conflict is taking place, and whether or not the object of attack is a single target or many. Moreover, where an EAI is authorized to act, having met the necessary *pre bello* requirements [C], *post bello* considerations, such as the availability of resources to repair, and/or rebuild any damage caused [D], must form a part of the decision as to whether to initiate a self-defense by proxy action — noting that under current interpretations, Europa is not lawfully obliged to consider [C] and [D], and some may even also argue [Z],[206] when responding to an actual armed attack.

---

205. *See* in particular API, *supra* note 43, art. 57(2)(a)(ii). This is discussed in greater detail by Grimal & Pollard (2021), *supra* note 3.

206. The point here being that it depends upon whether one supports the concept of concurrent application.

F. *Analysis 4*

This final analysis is kept intentionally brief because it directly transposes into the much wider discussion regarding the development and use of autonomous weapons systems (AWS). This is an existing area of debate that continues to attract a great deal of multi-disciplinary attention[207] — and, in short, it is well beyond the scope of the present article to attempt to settle that argument definitively. Nevertheless, the vast majority of AWS' — widely referred to as "killer robots"[208] — are EAIs, although the current authors distinguish the two by noting that unlike an AWS, an EAI is not delegated decision-making responsibilities regarding the direct application of force.[209]

The question the authors do pose in respect of the present discussion is that if an EAI is to be permitted to authorize an act of self-defense by proxy (owing to the fact that all of the options have been considered, and action is considered both necessary and proportionate), should they also be able to authorize direct force if that is the only reasonable course of action left to take? Here, some readers may draw a parallel with the Soviet Union's 1980's satellite early warning system "Oko," which incorrectly identified that the U.S. had launched five ICBMs.[210] Thankfully, the experienced Soviet Lieutenant Colonel responsible for monitoring the system trusted his instinct over and above the machine's insistent warnings, and what may have been World War III was averted. Some use this as an example of why it is necessary to keep humans in the loop.[211] However, herein lie the issue that goes to the very heart of this Article — in that future-looking discussions *must* not be premised upon past or even present technology. Regardless, if the authors' wide principle of concurrent application was applied to AWS decision-making it would still offer greater humanitarian protections than are currently on offer.

---

207. *See, e.g.*, Sparrow, *supra* note 155, whose arguments are based predominantly in the field of ethics as opposed to law; *see also* Heynes, *supra* note 155. One must also consider what is possible from a technical perspective amongst other things.

208. *See, e.g.*, DOCHERTY, *supra* note 147; *see also* Robert Sparrow, *Killer Robots*, 24:1 J. OF APPLIED PHIL. 62 (2007).

209. *See, e.g.*, U.S. Dep't of Defense Directive 3000.09, Autonomy in Weapon System, 13 (U.S. Dep't of Defense 2012) (amended 2017), defining AWS as "[a] weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system but can select and engage targets without further human input after activation."

210. For a useful discussion *see* SCHARRE, *supra* note 23, at 1–2.

211. *Id.*

By way of sectional summary, the authors' proposed test is very much designed to mitigate and negate understandable objection and distrust of such a permissive approach to pre-emption. The authors defend this approach in three key ways. First, they do so by recalibrating the temporal spectrum itself so that the necessity requirement is transitioned to the point of attack, and that any assessment also includes *post bellum* considerations. The second key defense is that the idea of "perpetual self-defense" naturally invites that a constant state of defensive readiness. This is both strategically and legally desirable, and somewhat importantly, can already be seen in practice. Thirdly, that self-defense by proxy in the form of influence communications is certainly a "lesser of the evils." To stress test this approach, the authors simulated these differing approaches via the lens of a single scenario as a means of concluding that in certain circumstances, an EAI *should* be permitted to act pre-emptively (at the very least in the form of influence communications).

G. *The Proliferation of EAI Technology*

The discussions in the preceding sections have been squarely focused upon the premise that the 'defending state' is the only actor that is in possession of the relevant EAI technology. However, if an aggressor also had the same or a similar EAI capability — that is, one which could manipulate future events in order to reach a preferable outcome — a number of extraneous factors might need to be considered. These may include, for instance, the possibility that one EAI may need to "bluff" another EAI into "believing" a particular course of action was the most likely, or, for example, that certain strategic capabilities were, or were not in place. The bluffing EAI, or perhaps even both EAIs, would then presumably try to get its opposite number to act in light of the dis-information, with the end goal of gaining an operational or strategic advantage. This could have various implications, not least upon the civilian population. But, somewhat crucially, there is also a chance that EAI-to-EAI exchanges such as these could in fact render both systems entirely derelict. This could be, for example, due to a speed of operation that is beyond human supervision, and/or the detrimental effects upon a system predictability.

There is no doubt that when EAIs (or AIs) meet in this manner, it is not without its difficulties.[212] However, such complications must not be

212. *See* SCHARRE, *supra* note 23, at 199-210 where the author discusses, for example, the speed at which algorithms can react to one another. In particular, the author recalls the Wall Street stock exchange incident on 31 July 2012 which saw the Knight Capital Group go bankrupt in only 45 minutes. At p. 204, Scharre notes, during "one 14-second period, high-frequency trading algorithms exchanged 27,000 E-mini contracts."

presumed to be terminal. For example, it is also not inconceivable that one EAI could readily learn to read its opposite number, and to accurately distinguish real information and intelligence, from mis-information and/or dis-information — and act accordingly. Indeed, a question that may need to be answered in this regard, is should an EAI be permitted to send a punitive message or warning to an opposite number, or should such actions be considered as a form of reprisal, and thus prohibited?

Nevertheless, if states are to harness the full strategic benefits of the EAIs under discussion, they *must* prevent the technology from becoming prematurely obsolete. In order to do this, they may not only have to keep the precise nature of systems relatively secret (which is arguably standard fare in the military realm), but they must also choose to what extent they should rely upon them. Here, the reader may recall the movie motion picture, "*The Imitation Game*," and, in particular, the 'negotiation scene.'[213] Notwithstanding the caveat acknowledging that the movie may not be historically accurate, in this particular scene, the father of AI, Alan Turing, asks the British security services to ensure the news that the Enigma code has been cracked is kept secret. Implying that in some instances sacrifices will have to be made, Turing proffers, we must decide which threats to act upon, and which to allow to continue. The question (the fictional) Turing asks is, what is "the minimal number of actions it would take for us to win the war, but the maximum number we can take before the German's get suspicious?"[214] The fundamental point here is that this question may be as key to ensuring victory in future conflicts utilizing EAIs, as it was to the codebreakers of WWII.

V. Wider Implications

The purpose of this penultimate section is to extend the optics of the previous trajectory of discussion (and the "Test" conceived by the Authors in Part IV) to further implications for other areas of *ad bellum* discussion: Collective Security, Humanitarian Intervention, and Responsibility to Protect (R2P). By way of overall caveat these sections will predominantly focus on the implications and applications, rather than overly revisiting the already well-trodden ground within the

213. *See* The Imitation Game (Black Bear Pictures 2014); *See* Fadhila Hasna, *Analysis of Negotiation Scenes from Movie "The Imitation Game (2014)"*, YouTube (April 5, 2015), https://www.youtube.com/watch?v=GhP1jwbOOYE, for a brief analysis of the scene in question (last visited Jan. 10, 2022).

214. The Imitation Game, *supra* note 213.

scholarship within these areas. Naturally, and perhaps the uniting thread in terms of implications for all three areas, is the potential endpoint of regime change via influence communications, an outcome that is perhaps the most desirable to those states already engaging in that particular practice.

A. *Recalibration of Collective Security*

Certainly, and in relation to the existing overall scholarship pertaining to the United Nations Security Council, and more specifically its Chapter VII enforcement powers, there is no dearth or paucity of literature.[215] Indeed, the mechanics of escalation from the determination of threat to international peace and security pursuant to Article 39, through to "green light" authorization of force in compliance with Article 42 of the United Nations Charter, are well-documented within the existing scholarship.[216] Nevertheless, before reaching the application of the authors' findings and uniquely applying their test to such considerations, the present Article will briefly revisit the salient points of the mechanics of Collective Security.Readers will no doubt be familiar with the basic premise that prior to the United Nations Security Council becoming "seized" of a particular matter, a determination (pursuant to Article 39) must be reached — the UNSC must appraise that the situation faced much be of sufficient gravity to trigger the existence of a threat to international peace and security. Notably, and as the literature already indicates, such quasi-judicial assessments (as to what constitutes a threat to international peace and security) have gradually become increasingly "permissive" particularly during the 1990s — so much so, that a discernible iOS/Windows update to the interpretation has no doubt been performed.[217]

---

215. For a useful starting point in this regard *see generally* OXFORD HANDBOOK, *supra* note 11, though Part II, Collective Security and the Non-Use of Force (pp 179–436) is particularly relevant. *See generally* SEBASTIAN VON EINSIEDEL, DAVID M. MALONE & BRUNO STAGNO UGARTE, THE U.N. SECURITY COUNCIL IN THE 21ST CENTURY (2016).

216. *See, e.g.*, Terry D. Gill, *Legal and some political limitations on the power of the U.N. Security Council to exercise its enforcement powers under Chapter VII of the Charter*, *in* 26 NETHERLANDS YEARBOOK OF INTERNATIONAL LAW 33 (L.A.N.M Barnhoorn et al., eds., 1995); *see also* Rob McLaughlin, *The Legal Regime Applicable to Use of Lethal Force When Operating under a United Nations Security Council Chapter VII Mandate Authorising 'All Necessary Means'*, 12:3 J. CONFLICT & SEC. L. 389 (2007); *see also* ERIKA DE WET, THE CHAPTER VII POWERS OF THE UNITED NATIONS SECURITY COUNCIL 184 (2004).

217. *See, e.g.*, Mónica Lourdes De La Serna Galvan, *Interpretation of Article 39 of the U.N. Charter (Threat to the Peace) by the Security Council: Is the Security Council a Legislator for the Entire International Community?*, 11 ANU. MEX. DER. INTER. 147, n.29 (2011) (citing Prosecutor v. Tadic, Case No. IT-94-1-AR72, Decision on Defence Motion for Interlocutory Appeal on Jurisdiction, ¶ 27 (Int'l

Undeniably, and understandably, those drafting the United Nations Charter, and perhaps Article 39 in particular, would have been unable to foresee the wide-ranging threats faced by the international community in more recent times.[218] To this extent, one might charitably concede that the UNSC has commendably adapted its remit from "narrow" to "wide" — no longer are international armed conflicts the sole purview, nor are they easily identified as being of sufficient severity to satisfy the trigger provision of Article 39, the trigger itself becoming more "hairline."[219] Indeed, unlike a Glock pistol, it is fair to say that there is less of a "wall" of resistance.[220] Nonetheless, once the remit of Article 39 has been satisfied, Article 40 requires states to "cease and desist," although one might reasonably argue that the practice of states heeding to Article 40 "warnings" are somewhat few and far between.[221]

Failure by a state to adhere to the requirements under Article 40, and the UNSC having attempted to diffuse via "cease and desist" results in an escalation of approach with the UNSC looking instead to transition (via Article 41) to the use of non-forceful measures.[222] For the most part, non-forceful measures are typically encapsulated by the use of economic and diplomatic sanctions, although all too often, a course of action that the state apparatus can readily repost and deflect — it is

Crim. Trib. For the Former Yugoslavia Oct. 2, 1995), and in particular noting how they describe threats to the peace as being 'political concepts'). *See generally* Robert Cryer, *The Security Council and Article 39: A Threat to Coherence?*, 1:2 J. CONFLICT & SEC. L. 161 (1996).

218. *See* Report of the High-Level Panel on Threats, Challenges and Change on a More Secure World: Our Shared Responsibility, U.N. Doc. A/59/565, at 10–14 (2004). https://www.un.org/peacebuilding/sites/www.un.org.peacebuilding/files/documents/hlp_more_secure_world.pdf (last visited Sept. 9, 2021).

219. The 1990's was a decade when the UNSC needed to transition its utilization of art. 39 because the threats to international peace and security were perhaps not those envisaged by the drafters. *See, e.g.*, S.C. Res. 748 (Mar. 31, 1992) (concerning refusal to hand over Lockerbie suspects in Libya); S.C. Res. 864 (Sept. 15, 1993) (extending the mandate of the U.N. Angola Verification Mission II and possible arms and oil embargo against UNITA, a non-state actor); S.C. Res. 929 (June 22, 1994) (concerning internal armed conflict in Rwanda); S.C. Res. 940 (July 31, 1994) (concerning overthrow of a government in Haiti); S.C. Res. 1267 (Oct. 15, 1999) (concerning refusal to hand over Bin Laden in Afghanistan); *see also* Cryer, *supra* note 217.

220. Cryer, s*upra* note 217.

221. Albeit and while there is "understandable" context, one can readily use the panoply of S.C. Res. 660–678 (Aug. 2, 1990 – Nov. 29, 1990) (pertaining to Iraq's unlawful annexation and invasion of Kuwait in 1990).

222. The precise text of Art. 41 states "The Security Council may decide what measures not involving the use of armed force are to be employed to give effect to its decisions, and it may call upon the Members of the United Nations to apply such measures. These may include complete or partial interruption of economic relations and of rail, sea, air, postal, telegraphic, radio, and other means of communication, and the severance of diplomatic relations." U.N. Charter, *supra* note 12.

the civilian population at large that tends to suffer the consequences.[223] The UNSC's final recourse under its chapter VII powers is the authorization of force under Article 42 (one of two permissible exceptions to the prohibition against the use or threat of force contained with Article 2(4)) of which there are many well-documented examples.[224]

The point of particular interest for the authors of the present Article, and in parallel to discussions in previous sections pertaining to self-defense, centers primarily on the UNSC's potential to pre-emptively act against a non-imminent (at least in the human mind) threat to international peace and security. Clearly, the authors readily state at this juncture, that such discussion is theoretical and is intended to open such debate into this very niche and future-looking application. However, as the ICRC has recently suggested that AI could be used to help to predict and avoid humanitarian crises,[225] the question is, what is to prevent the same, or similar, tech from being repurposed for the matters under consideration here.

Ultimately, with regard of the present discussion, there are two possible routes. First is to re-apply that theoretical coupling of necessity to action and incorporate the post *bellum* considerations proposed by the authors in Part IV. To soothe anxiety in relation to this rather controversial approach, one might readily temper that the coupling of necessity of last resort to action, is already the current *modus operandi* of the UNSC — force authorized under Article 42 is already "last resort" owing to the existence of the non-forceful route under Article 41. Secondly, the UNSC could pre-emptively utilize influence communications and propaganda as methods falling short of actual force — again such a consideration would perhaps naturally fall (although not necessarily envisaged) within the confines of Article 41. Undeniably, the apex of this discussion, in a similar vein to the one we will see shortly

223. The point being that strict economic sanctions in particular are very likely to adversely affect the civilian population.

224. One may readily recollect the "spillage" of ink in relation to S.C. Res. 678 (Nov. 29, 1990) (regarding Iraq in 1990), and the infamous equivocal text of S.C. Res. 1441 (Nov. 8, 2002). *See, e.g.*, Thomas N. Franck, *What Happens Now? The United Nations After Iraq*, 97:3 AM. J. INT'L L. 607 (2003); James P. Terry, *A Legal Appraisal of Military Action In Iraq*, 57 NAVAL WAR COLL. REV. 53 (2004); Gregory B. Marfleet & Colleen Miller, *Failure after 1441: Bush and Chirac in the U.N. Security Council*, 1 FOREIGN POL'Y ANALYSIS 333 (2005).

225. *See* Christopher Chen, *The future is now: artificial intelligence and anticipatory humanitarian action*, ICRC HUMANITARIAN L. & POL'Y (Aug.19, 2021), https://blogs.icrc.org/law-and-policy/2021/08/19/artificial-intelligence-anticipatory-humanitarian/. *See, e.g.*, Dan McQuillan, *How can AI help in a humanitarian crisis?*, INDEPENDENT (May 2, 2018, 12:54 PM), https://www.independent.co.uk/news/science/artifical-intelligence-disaster-response-humanitarian-crisis-ai-help-a8319361.html.

(pertaining to the application of the authors' discussion in the context of HI/R2P), is the proposal by the EAI (via the modus of influence communications) of regime change. While antithetical to almost the entirety of the current legal positioning and literature, the EAI might well calculate the optics of that trajectory. As previously alluded to, it is not the authors' intention to close this particular discussion in relation to the UNSC, but to highlight the various optics operating within the trajectory of the *jus ad bellum* which will, undeniably, require greater scrutiny.

However, the novel and highly controversial aspect is that, unlike a human making that assessment, the EAI may be better placed to anticipate or pre-emptively to determine the severity of threat in relation to Article 39. Rather than awaiting the slow revolving machinery of collective security mechanics, the EAI may envisage that in pursuance to the Chapter VII powers, pre-emptive action, including *post bellum*, considerations necessitate regime change. At the more minor end of the spectrum, it could simply be a propaganda influence comms mission "instructing" the EAI and civilian population to uprise as a means of achieving it. If a state's inherent right of self-defense potentially allows for pre-emptive action providing the test in Part IV is complied with, could the same not apply here? To this end how would EAIs interpret the vagaries of UNSC coded language such as "use all necessary means/measures"?[226] And indeed, would the EAI be overly diligent in both its translation and application of such equivocal human phrasing?

B. *Recalibration of Extra Charter Exceptions (Humanitarian Intervention and Responsibility to Protect)*

As was the case in the previous discussion, the purpose of the following analysis is not to revisit the corpus of legal argument (of which there are many)[227] pertaining to both Humanitarian Intervention, and its more recent "application" via the doctrine of R2P, but instead to focus on the pre-emptive angle of action, whether this is forceful, perpetual, or by proxy. The "practice" of Unilateral Humanitarian Intervention, *i.e.*, intervention without the blessing and pardon of the UNSC on the basis of preventing humanitarian catastrophe (potentially, as a result of veto-induced non-action by the Security Council) is perhaps most

226. *See generally* Michael C. Wood, *The Interpretation of Security Council Resolutions*, MAX PLANCK YEARBOOK OF UNITED NATIONS LAW (1998).

227. *See generally e.g.*, Alex J. Bellamy, *The Responsibility to Protect and the Problem of Military Intervention*, 84 INT'L AFF. 615 (2008); Carsten Stahn, *Responsibility to Protect: Political Rhetoric or Emerging Legal Norm?*, 101 AM. J. INT'L L. 99 (2007).

poignantly exemplified by the NATO bombings in Kosovo in 1999[228] (although it is highly debatable whether such practice represents "state practice").[229] Again, this oft-cited example has appropriately received considerable scrutiny, and it is not the purpose of this present discussion to revisit the lawfulness (or not) of such action. Rather, it is to consider whether (in the absence of UNSC authorization) a state may pre-emptively act so as to prevent "inception" to actual "conception" of *hostis hummanis* or *delicta jure gentium*. One might immediately follow-up this line of thought by "transitioning" from unilateral humanitarian intervention to Responsibility to Protect, and "couple" the discussion. While both are clearly very different doctrines, their ultimate purpose is to prevent or try to prevent the greatest crimes known to humankind.[230]

As a very unique, yet controversial and similar aside, the authors note and enquire whether in the future, the R2P doctrine should remain solely limited to humans protecting humans, or whether that protection should be extended to robots as well. In other words, do humans and EAI have an equal responsibility to protect robots facing mistreatment in another state (at its extreme, genocide by humans against robots or EAI, rather than the more obvious fear of mistreatment of humans by robots) and what this discussion might/should look like. For example, what would the threshold parameters be for acting? Would an EAI acting with a "bias" to protect fellow robots engage sooner in terms of severity, and more practically, how would the *ad bellum* ROE regarding the doctrine operate?

Ultimately, and in relation to the primary focus of this present Article, similar considerations and question posed in relation to possible pre-emptive action by an EAI in the Collective Security Context are also apt at this juncture. As the authors resoundingly maintain, where

228. Following a period of protracted violence and reports of ethnic cleansing, the United Nations Security Council adopted resolution 1199 (UNSCR 1199) demanding a ceasefire in September 1998. S.C. Res. 1199 (Sept. 23, 1998), http://unscr.com/en/resolutions/doc/1199. However, while there was an agreement between the warring parties, and an initial period of stability in the region, the violence soon returned to pre-agreement levels. As a result, despite having no UNSC resolution to support such action, NATO began Operation Allied Force on 23 March 1999, and a high-altitude bombing campaign which primarily targeted the Yugoslav air defense system. *See, e.g.*, *The Crisis in Kosovo*, HUMAN RIGHTS WATCH REPORT (Feb. 2000), https://www.hrw.org/reports/2000/nato/Natbm200-01.htm.

229. That is, the level of state practice that is required for the formation of customary international law.

230. *See, e.g.*, Genocide, Convention on the Prevention and Punishment of the Crime of Genocide, Dec. 9, 1948 78 U.N.T.S 277; Crimes Against Humanity which is further defined by Rome Statute of the International Criminal Court art. 71. July 17, 1998 2187, U.N.T.S 3.

an EAI can accurately "calculate" a forthcoming need to act, albeit against a non-imminent and latent threat, part of the calculation process would be to consider every counter move. Such a counter move could naturally include the UNSC being unable to fully trigger Article 39 (in light of political considerations). Consequently, pre-emptive unilateral action by states on humanitarian grounds or indeed more closely aligned to R2P (both pre-emptively and via the lens of influence communications) could provide a strategic alternative to legal inaction.

VI. Conclusion

The soccer player, Cristiano Ronaldo, recently became the first individual in the world to amas 500 million social media "followers."[231] And, regardless as to how one might view this "feat," the statistic goes beyond simply evidencing that the sportsman is leading a colossal popularity contest. According to reports, Ronaldo is now paid more to "influence" his followers into buying into his (perhaps perceived) lifestyle than he is to play the "beautiful game" itself.[232] The point here is, there is absolutely no question that the ability to influence pays dividends. In the realm of geopolitics, this *modus operandi* is at the very soul of statesmanship, and the "reach" of national influence is likely to have

231. As of 25th February 2021, 'Ronaldo's' Instagram social media account identifies that he has 265 million "followers." *See* Cristiano Ronaldo (@Cristiano), Instagram, https://www.instagram.com/cristiano/ (last visited Feb. 25, 2021); his Facebook account identifies that he also has over 148 million followers on that platform, *see* Cristiano Ronaldo, Facebook, https://en-gb.facebook.com/Cristiano/ (last visited Feb. 25, 2021); while Twitter identifies a further 91 million followers. *See* @Cristiano, Twitter, https://twitter.com/Cristiano?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor. Of course, a single person or institution may "follow" all three of these platforms, meaning it is unlikely that the figure represents 500 million separate entities.

232. *See, e.g.*, Barnaby Lane, *Cristiano Ronaldo Reportedly Makes more Money Being an Influencer on Instagram than he does Playing Soccer for Juventus*, Business Insider (Oct. 16, 2019), https://www.businessinsider.com/cristiano-ronaldo-makes-more-money-from-instagram-than-juventus-2019-10?r=US&IR=T. Indeed, it is reported that he is paid somewhere in the region of $975,000 (US) for each "post" in which he endorses consumer items ranging from haircare products to sportswear. *See* Niall McCarthy, *The Highest Earners on Instagram*, Forbes (Oct. 28, 2019), https://www.forbes.com/sites/niallmccarthy/2019/10/28/the-highest-earners-on-instagram-infographic/?sh=2a3c31551110. Note also that his influence reaches beyond his personal fortune. The value of his primary employers (before his recent high-profile move back to Manchester United in the U.K.) for example, the Italian football club Juventus, are also believed to have risen by over $350 million upon confirming that they had contracted his services. *See* Zak Garner-Purkis, *Cristiano Ronaldo's Instagram Success: A Glimpse into How Social Media is Changing Soccer*, Forbes (Aug. 5, 2020), https://www.forbes.com/sites/zakgarnerpurkis/2020/08/05/evidence-cristiano-ronaldos-instagram-is-more-important-than-his-soccer/?sh=6e3273b7296d.

a significant impact upon the chances of a state achieving its strategic goals.

Consequently, states undoubtably take every opportunity to attempt to exert their influence in every way they can.[233] While the authors do not wish to revisit the entire corpus of public international law at this juncture, they nevertheless wish to reinforce their claim that state practice and *opinio juris* regarding influence has already led to the formulation of new customary norms ("perpetual self-defense" and "self-defense by proxy").[234] Nevertheless, while strategy is clearly closely associated with the conduct of military operations, diplomatic influence should be distinguished from "influence communications" — the latter being the concept in which this discussion was grounded. Central to the preceding analysis was the acknowledgement that military led influence communications must be seen as a security and defense apparatus, as opposed to a "mere" political power of persuasion.

The authors grounded the previous discussion in emerging technologies. And, with that in mind (despite legitimate concerns), AI should not necessarily be seen only as an inherently disruptive technology. Instead, states and industry will remain keen to develop and utilize EAIs because they will offer humanity numerous significant advantages — and not only strategic ones.[235] One such advantage is data-analysis, which, when compared to humankind, AI (and thus EAI) are increasingly more proficient.[236] Perpetual self-defense both generates, and "mines," a great deal of data. It is somewhat inevitable, therefore, that AI systems will support future commanders,[237] especially given the fact that a further advantage of AI is its comparatively "warp speed" decision making capability.[238] With that in mind, one line of reasoning is that

233. *See, e.g.*, Dinko Hanaan Dinko, *How 'Mask Diplomacy' Rescued China's Image in Africa*, DEFENSE ONE (Mar. 10, 2021), https://www.defenseone.com/ideas/2021/03/chinas-mask-diplomacy-wins-influence-across-africa-during-and-after-pandemic/172583/.

234. Though note that the authors are considering this matter in greater detail in a forthcoming publication.

235. EAIs are regularly lauded because they can replace humans currently undertaking dull, dirty, and/ or dangerous, tasks (the 3 D's of robotization). *See, e.g.*, Stephanie Neil, *Give the Robot the Dirty, Dull, or Dangerous Job*, AUTOMATION WORLD (Mar. 14, 2019), https://www.automationworld.com/factory/robotics/blog/13319620/give-the-robot-the-dirty-dull-or-dangerous-job.

236. *See, e.g.*, Tristan Greene, *Face It, AI is Better at Data-Analysis Than Humans*, TNW NEWS (Jul. 28, 2017), https://thenextweb.com/news/face-it-ai-is-better-at-data-analysis-than-humans.

237. Demonstrated, not least by General Glen D. VanHerck, commander, NORAD and U.S. Northern Command in relation to the Global Information Dominance Experiment GIDE, *supra* note 2.

238. *See, e.g.*, SCHARRE, *supra* note 23 a discussion regarding the destructive sequence of events initiated by a leading Wall Street trading company's exchange algorithm.

heads of state, military commanders, and their supporting staffs will be trained to utilize these (and additional) advantages, and to supervise AI systems rather than completely delegate decision making-responsibilities.[239] However, as noted, while this may be dressed up as supervision, there is also an inherent risk of machine overreliance.

There is no doubt that military software and hardware, including AI systems, have moved on considerably since Oko almost inadvertently initiated a nuclear apocalypse. Nevertheless, the authors still acknowledge that contemporary AI systems are still some way off the level of capability, predictability, and reliability that will be required if the proposed test is to be of genuine, material, use. However, institutions such as the ICRC have recently highlighted that while technical Research and Development is moving on at a ferocious pace,[240] diplomacy and policy-making is an inherently protracted process.[241] Moreover, the ICRC also notes that influence communications, and particularly MDH, is already causing significant harms upon contemporary battlefields and in other locations where there are humanitarian crises.[242] Therefore, there is a clear "necessity" (very sic/ad nauseum) to open the debate now, *if* the greatest aggregate of humanitarian protections are to be offered to individuals who will also be affected in the future. The authors' test, and reliance upon future advances in technology, is one method of helping to achieve this.

By way of summary, Part II of this Article introduced the *jus ad bellum* and *jus in bello* framework that will regulate the use of future EAIs. This started with the necessary introduction of Article 2(4) U.N. Charter regarding the prohibition of the threat and use of force. Importantly, that examination identified that while the codified version of self-defense that is contained within Article 51 U.N. Charter and the customary version of self-defense are recognized as lawful exceptions to Article 2(4), pre-emptive self-defense remains unlawful. The authors

239. As was the case, for example, with the GIDE 3 experiment, *see* GIDE, *supra* note 2.

240. S*ee, e.g.*, Saman Rejali & Yannick Heiniger, *Editorial: The Role of Digital Technologies in Humanitarian Law, Policy and Action: Charting a Path* FORWARD, ICRC REVIEW (2020), *supra* note 135, at 3. Here, the ICRC state "the "product" can outpace the due diligence required to ensure that digital technologies cause more benefit than harm to affected populations.

241. *See* Frank Sauer, *Stepping back from the brink: Why multilateral regulation of autonomy in weapons systems is difficult, yet imperative and feasible, ICRC Digital Technologies and War*, 102 ICRC, no. 913 at 236–37 (2020). Here the author identifies the further sources; KRC, "Alarm Bells Ring on Killer Robots," 15 November 2019, available at: www.stopkillerrobots.org/ 2019/11/alarmbells/; *Richard Moyes, Critical Commentary on the "Guiding Principles"*, ARTICLE 36 (Nov. 2019), www.article36.org/wp-content/uploads/2019/11/Commentary-on-the-guiding-principles.pdf.

242. *See* Rejali & Heiniger, *supra* note 240, at 2.

also used this analysis to provide further support to the principle of concurrent application, which ensures both *jus ad bellum* and *jus in bello* norms are applicable to *all* self-defensive actions. Part III introduced and examined the concept of influence communications, and propaganda, for self-defensive purposes. Here, the authors identified two key concepts: the existing practice of perpetual self-defense, and the potential for nations to invoke self-defense by proxy — the latter being an act in which an indirect application of force is a foreseeable consequence of applying perpetual self-defense. In this discussion, the authors also introduced and justified the need to consider the *jus post bello* in EAI pre-emptive assessments. Perhaps a very final "footnote" in relation to the authors' discussion regarding influence communications concerns the oft-cited difficulty with proportionality assessments in the *in bello* realm. Those difficulties pertain to "dual-use" targets — ones which have both military and civilian effect; typically, this may include power stations. Interestingly, one might suggest that with influence communications there is also an element of "dual-use," but in "reverse" — influence communications would affect civilians first, with the end point of potentially nullifying a military attack.

Part IV introduced the authors' test and identified the conditions under which an EAI should be permitted to act pre-emptively. In particular, the authors' test expanded the principle of concurrent application to include a *pre-bello* and *post bellum* assessment and provided a method for assigning a quasi-legal nature to the latter. Here, they utilized a hypothetical scenario, and four alternate analyses, to walk the reader through some of the situations in which EAI assessments will, and will not, be either necessary and/or permitted. Finally, Part V considered a number of further implications for other areas of the *ad bellum* discussion: Collective Security, Humanitarian Intervention, and R2P. These final examinations were not intended to be expansive but were introduced to promote further discussion in this regard.

In a penultimate observation, the authors question whether the EAI should itself report to the UNSC once it has pre-emptively, or actually, acted via proxy.[243] If so, this could be problematic because by doing so, it would almost defeat the purpose of on-going or perpetual self-defense — given, as previously alluded to, the best results would undoubtably be achieved while operating under the radar.

243. *See generally* James A. Green, *The Article 51 Reporting Requirement for Self-Defence Actions*, 55 VA. J. INT'L L. 563 (2015).

Overwhelmingly, it has been the authors' unwavering belief throughout this discussion that due to the benefit of their additional "foresight," EAI assessments and decision-making would be made without political motivations and would not be restricted by human desideratum. In the purest strategic sense, an EAI would be "Machiavellian" in nature — the very purpose of the state being to protect and defend its citizens. The authors therefore manifestly maintain that recourse to the doctrine of pre-emptive self-defense is lawful under limited circumstances.