

A Spoonful of Poison : Why the Martens Clause Is the Antidote to Pre-Conflict Cyber Mischief

Samuel White*
Daniel Harker**

ABSTRACT

Pre-conflict data poisoning, whereby States covertly corrupt the training data of artificial intelligence and machine learning systems before hostilities arise, exposes a profound gap in international law. Operating below recognized thresholds of armed conflict and use of force, such operations allow adversaries to quietly degrade future military capability, manipulate civilian infrastructure, and embed latent vulnerabilities across interconnected systems with minimal legal consequence. Unlike traditional sabotage, poisoned data does not destroy a single object; it contaminates entire informational ecosystems, producing cascading and potentially irreversible effects across military and civilian networks alike. Existing international humanitarian law remains tied to kinetic, event-based understandings of violence, whilst international human rights law struggles with attribution, jurisdiction, and covert cyber operations. This paper argues that the Martens Clause provides the only coherent normative restraint on such activity, supplying a humanitarian framework capable of governing emerging forms of anticipatory, non-kinetic, and technologically diffuse state competition.

I. INTRODUCTION

As the International Committee of the Red Cross' Commentary on Additional Protocol 1 notes,

[T]he history of weapons is above all dominated by the contest between projectile and armour, between mortar fire and fortification. The discovery of a new means of attack leads to the introduction of a new means of defence, which in turn provokes the introduction of an even more powerful projectile.¹

* Scientia Senior Lecturer in Military Law and War Studies, UNSW Canberra; Global Fellow, Centre for International Law – National University of Singapore.

** GDLP, LLB/BA (Adelaide University).

The views, opinions and errors are those of the authors alone, and do not reflect any of their respective employers.

¹ Yves Sandoz, Christophe Swinarski and Bruno Zimmermann (eds), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*, International Committee of the Red Cross and Martinus Nijhoff Publishers, Geneva, 1987, p. 401, para. 1412.

Put simply, a familiar dynamic governs the evolution of conflict: every strengthening of defence (big shield) provokes a corresponding intensification of offence (bigger sword).

In recent years, artificial intelligence (AI) and machine learning (ML) systems have been integrated across a broad spectrum of services and infrastructures, interacting with civilian, governmental, and defence sectors. They have driven bigger swords, and bigger shields. However, system security implementations such as those being used in military infrastructure and operations are developing at a rate which makes each direct attempt at infiltration become increasingly likely to fail.² The advancement of such technology and the evaluation of these new and continually evolving shields then poses the question not of whether these shields can be breached, but of how they might be circumvented. In operation, contemporary practice has responded to this question through three technical vectors:

- (1) **adversarial attacks** — carefully constructed inputs or camouflage designed to deceive or confuse machine-learning systems;³
- (2) **algorithmic or logical manipulation** — interventions that alter system logic or model behaviour at the code or architecture level;⁴ and
- (3) **data poisoning** — the insertion of corrupted training data to subvert a model's learning process.⁵

Each pathway entails distinct technical mechanisms and normative implications; demanding cross-disciplinary scrutiny from engineers, policymakers and legal scholars. Whilst all can occur *within* armed conflict, the latter two forms occur primarily in peacetime. Significantly, these actions do not fall under the ambit of espionage, as that term describes tolerated information collection.⁶ By contrast, the objective of these operations is not collection but corruption. This paper will focus primarily on the application and avenues of curtailment applicable to the third form, data poisoning.

² Mark A. Visger, "Garbage In, Garbage Out: Data Poisoning Attacks and Their Legal Implications" in Laura A. Dickinson and Edward W. Berg (eds), *Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold*, Oxford University Press, New York, 2024, p. 179.

³ Jonathan Kwik, "Is Wearing These Sunglasses an Attack? Obligations under IHL Related to Anti-AI Countermeasures", *International Review of the Red Cross*, Vol. 106, No. 926, 2024, p. 739.

⁴ Micah Goldblum et al., "Dataset Security for Machine Learning: Data Poisoning, Backdoor Attacks, and Defenses", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 2, 2023, p. 1563.

⁵ M. Visger, above note 2, p. 179.

⁶ Martin C. Libicki, "The Coming of Cyber Espionage Norms", in Henry Rōigas et al. (eds), *2017 9th International Conference on Cyber Conflict: Defending the Core*, NATO CCD COE Publications, Tallinn, 2017, p. 7.

Data poisoning is the malicious insertion or alteration of data in datasets used to train AI or ML systems, with the aim of altering their function or capabilities once operational, unbeknownst to the model's user.⁷ It is not concerned with changing the algorithm or code that regulates the dataset (logic attacks) but with the data itself, feeding doubt or erroneous associations into the system either covertly or overtly. Data poisoning is both the logical evolution of competition in cyberspace, and a revolution in the history of armed conflict. Its break between application and effect, intention and event, and operation and outcome blur the lines between peace and war, the binary paradigms of international law. The topic merits discussion, as the international legal considerations and consequences of data poisoning are becoming increasingly important due to the fact that no clearly defined legal framework currently regulates such operations. The misalignment between the growing strategic competition of States, the separation of peacetime and wartime international law and the development of technological capabilities foreign to the imagination and contemplations of those who formulated the vast majority of existing codified international law has created an area of inter-state action devoid of regulation. This unregulated area sees itself continually enlarged by the reliance of the international community upon a black-letter binary paradigm of conflict/non-conflict in a world of grey technology.

Attempts to find analogies outside the cyber domain are problematic. Take the theoretical example of poisoning the fuel system of a Main Battle Tank so that it explodes only after travelling 10,000 kilometres. Although civilian harm is possible, the effects remain comparatively narrow and containable. The sabotage is directed at a single military object, with consequences largely limited by geography, function, and physical proximity. The same cannot be said of adversarial attacks or data poisoning against AI and machine learning systems. A sabotaged tank destroys a platform; poisoned data corrupts an ecosystem. Adversarial inputs can deceive systems at the point of operation, whilst data poisoning embeds latent vulnerabilities directly into the system's training environment. The result is not confined to one object or one battlespace.⁸

A poisoned dataset may silently distort targeting systems, logistics networks, autonomous surveillance, medical triage, or civilian infrastructure simultaneously. Once introduced, corrupted data or adversarial vulnerabilities may be replicated across interconnected systems, reused in downstream models, and propagated at scale without detection. Unlike traditional sabotage, the effects are not linear, localized, or necessarily reversible. This is the critical distinction. Conventional sabotage attacks a machine. Adversarial attacks and data poisoning weaponize the informational environment itself. The danger lies not only in destruction, but in the

⁷ Aaron Conti, "Data Poisoning as a Covert Weapon: Securing U.S. Military Superiority in AI-Driven Warfare", 30 June 2025, *Articles of War*, available at: <https://lieber.westpoint.edu/data-poisoning-covert-weapon-securing-us-military-superiority-ai-driven-warfare/>.

⁸ *Ibid.*, paras. 1–4, 8–10.

erosion of predictability, trust, and meaningful human control across systems that modern societies and militaries increasingly depend upon.⁹

The growing reliance on AI and ML platforms, due to their increasing capabilities and States' pursuit of a competitive edge, logically makes them a target for military operations seeking to utilize their vulnerabilities to hamper an adversary.¹⁰ Taking pre-emptive steps to damage datasets, however, does not always occur in times of armed conflict but in the alternate legal paradigm of peacetime (even if, under military doctrine, it is 'campaigning in competition').¹¹ Such data poisoning operations allow States to curb the development of an AI or ML system, to create future exploitations, or to destabilize an adversary's capabilities, all prior to the commencement of any currently recognisable form of warfare.¹² Acting in this way prior to armed conflict commencing represents an international activity not foreseeable when the Hague Regulations or Geneva Conventions were drafted, even if the ongoing march of technological progress was acknowledged.¹³

To date, much of the related legal discussion assumes that IHL faces a primary challenge in the cyber domain.¹⁴ However, a framing such as this may miss the point, as in comparison to the above discussion of pre-emptive steps taken to damage datasets in the paradigm of peacetime, IHL remains relatively comfortable if not adept in its regulation of such actions once hostilities arise. As is expanded upon in this paper, it is clear that IHL can – and does – spill the banks of the peace/war paradigms of international law (such as the need to review weapons,¹⁵ or to ensure respect).¹⁶ The Martens Clause is no different. However, it remains oppressively and ominously lacking in State practice or State positions with the nature of its obligation. What is the merit, then, in discussing the topic? Is it merely an academic endeavour? In short, no. Papers presented at the Workshop in April 2026 at the Centre for International Law, which triggered this Special Issue, made

⁹ *Ibid.*, paras. 19–29.

¹⁰ *Ibid.*

¹¹ M. Visger, above note 2.

¹² *Ibid.*

¹³ International Court of Justice, *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, ICJ Reports 1996, para. 257 [hereinafter *Legality of the Threat or Use of Nuclear Weapons*].

¹⁴ Laurent Gisel, Tilman Rodenhäuser and Knut Dörmann, "Twenty Years On: International Humanitarian Law and the Protection of Civilians against the Effects of Cyber Operations during Armed Conflicts", *International Review of the Red Cross*, Vol. 102, No. 913, 2021; Cordula Droegge, "Get Off My Cloud: Cyber Warfare, International Humanitarian Law, and the Protection of Civilians", *International Review of the Red Cross*, Vol. 94, No. 886, 2012.

¹⁵ See Protocol Additional (I) to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, 1125 UNTS, 8 June 1977 (entered into force 7 December 1978), Art. 36.

¹⁶ See Article 1 common to the four Geneva Conventions; on it, see further Eve Massingham and Annabel McConnachie (eds), *Ensuring Respect for International Humanitarian Law*, Routledge, Abingdon, 2020.

clear that the Martens Clause was a consistent touch point in academic thinking. The primary author, as a military law practitioner and academic, noted the clear divide between those that *do* and those that *think* of international law in this space. The purpose of this paper is thus to try, comprehensively, discuss the Martens Clause as a limiting legal factor (what members of the Profession of Arms would call a left or right of arc) in this space, and navigate the strength of the Clause.

II. THE PROBLEM

By its very nature, pre-conflict data poisoning activities appear to operate in a legal grey zone blending actions above and below, or before and after, the threshold of armed conflict. Are these issues to be treated as regulated and separated by two separate regimes, artificially spliced from one another even if they constitute a single act? Can actions be unconstrained in times of peace, but resulting outcomes regulated in times of armed conflict? It seems paradoxical. It is a friction point within international law, and one that the Martens Clause clearly aimed to regulate. Additional Protocol I Article 1(2) states:

In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience.¹⁷

This passage takes its lineage from the sentiments of the Clause, first introduced by Fyodor Fyodorovich Martens in the preamble of the 1899 Hague Convention:

Until a more complete code of the laws of war is issued, the High Contracting Parties think it right to declare that in cases not included in the Regulations adopted by them, populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the laws of humanity, and the requirements of the public conscience.¹⁸

The inclusion of this passage in the preamble was justified as the Conventions adopted in 1899 left numerous questions unanswered, requiring a dictate on the necessity of humanity and decency in decision-making, which was yet to be formalized in international law.¹⁹ This means that when treaties and existing

¹⁷ AP I, Art. 1(2).

¹⁸ Hague Convention (II) with Respect to the Laws and Customs of War on Land and its Annex: Regulations Concerning the Laws and Customs of War on Land, 187 CTS 429, 29 July 1899 (entered into force 4 September 1900), Preamble.

¹⁹ Y. Sandoz, C. Swinarski and B. Zimmermann (eds), above note 1, p. 38, para. 54.

law are silent or unclear, protection remains as identified through recognition of “the principles of humanity and from the dictates of public conscience.”²⁰ The inclusion of such sentiment in Additional Protocol I was intended to “reaffirm and develop humanitarian law: to supplement the substantive rules, and to reinforce the measures ensuring their application.”²¹

The need for such a legal principle is evident now more than ever before, with the vast development of technology and methods of adversarial state action creating scenarios not contemplated by codified law. The relevance of the Clause in a contemporary setting is illustrated in its utilisation in the International Court of Justice’s (ICJ) *Nuclear Weapons Advisory Opinion*, where the Clause was described as “an effective means of addressing the rapid evolution of military technology.”²² The explanation of the advisory opinion states that the Clause can validly rely on the identification of reasonable behaviour because the vast corpus of existing humanitarian law and treaty codification indicates to States the conduct expected of them.²³ This positions the Clause as a functional bridging mechanism for non-codified technologies, as it does not require a body of law, research, or academic work on a particular technological development to determine its constraints. Rather, it imposes an obligation on States to recognize how existing humanitarian law interacts with and imposes obligations forward. Furthermore, the creation of technology vastly different from that which existed prior has occurred previously in recent history, with the development of nuclear weapons presenting a “qualitative as well as quantitative difference” from “all conventional arms”.²⁴ The development of nuclear weapons came about after most of the principles of humanitarian law applicable in armed conflict had already been recognized, and the topic was left aside in the Conferences of 1949 and 1974–1977.²⁵ However, even given this, the Court found that it could not be concluded that such principles of humanitarian law applicable in armed conflict did not apply to these new weapons.²⁶ and that, the entirety of the laws of armed conflict apply to weapons of the past, present and future.²⁷ Although this view was developed in relation to weaponry used in warfare, the illustration that the development of new capabilities is still governed by the principles of humanity and dictates of public conscience is extremely relevant to discerning a framework within which pre-conflict data poisoning is governed. Further to this, during the *Krupp* case at Nuremberg, the US Military Tribunal articulated that the Clause could be applied generally as the yardstick with which to

²⁰ AP I, Art. 1(2).

²¹ Y. Sandoz, C. Swinarski and B. Zimmermann (eds), above note 1, p. 28, para. 26.

²² *Legality of the Threat or Use of Nuclear Weapons*, above note 13, p. 257, para. 78.

²³ *Ibid.*, p. 258, para. 82.

²⁴ *Ibid.*, p. 259, para. 85.

²⁵ *Ibid.*, para. 84.

²⁶ *Ibid.*, p. 259, para. 86.

²⁷ *Ibid.*

assess conduct not specifically addressed by the Hague Convention and its Regulations, representing the existing law of the time.²⁸

This shows that, even though there is no treaty law or international consensus on data poisoning, and these activities are new and lack substantial similarity to actions already governed by international law, there is no clear reason for data poisoning operations before armed conflict not to be constrained by the elements of the Clause, nor is there evidence that these actions are *de facto* legal. Such implementations and considerations of the Clause, and its integration with concepts of IHRL and IHL, have also been evident in contemporary discussions regarding emerging technologies and weapons. At the 2024 Group of Governmental Experts meeting on Lethal Autonomous Weapons Systems (LAWS), delegations emphasized the need to incorporate a Clause provision in the rolling text discussing the governance of LAWS, with some also underlining the importance of considering IHRL alongside IHL in this regard.²⁹

Returning to the foundational clause, Fyodor Fyodorovich Martens best illustrates his intention as to the recognition of the ongoing development of the international system, which forms the foundation of the Clause in the following passages:

If we take a brief look at the history of the development of international life, we cannot fail to see how international law is gradually changing, how the scourge of war is being alleviated and the benefits of peaceful coexistence among peoples are increasing.³⁰

[W]e understand international life as being in constant progression, and the aggregate of the conditions of its development determines the international law of the given age; so a change in the conditions of international life inevitably requires a change in existing law.³¹

The second of these passages speaks directly to the view that the development of the sword and the shield is intrinsically linked, with the development of one producing the development of the other. Martens continued in this publication to describe the entanglement of humanity and law, stating that “much of what could previously have been requested in the name of humanity or universal human

²⁸ United Nations War Crimes Commission, *Law Reports of Trials of War Criminals*, English ed., Vol. X (The I.G. Farben and Krupp Trials), HMSO, London, 1949, p. 133.

²⁹ Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Chair’s Summary – Second 2024 Session of the GGE on LAWS, UN Doc. CCW/GGE.1/2024/WP.11, 2 October 2024, para. 17.

³⁰ Fyodor Fyodorovich Martens, *On the Law of Private Property in Time of War*, V. Golovin, St Petersburg, 1869, p. 7.

³¹ *Ibid.*

relations must, at another time and under different conditions of life, be demanded in fulfilment of the law.”³²

The Clause presents a promising avenue for curtailing unrestricted grey zone actions including data poisoning. However, any proposed restrictions over grey zone operations achieved through utilisation of the Clause would likely be hampered by two key criticisms. Firstly, there is a clear traditional consensus that the Clause remains a creature of IHL confined to scenarios of identifiable armed conflict and, secondly, there is a significant lack of certainty regarding the applicability and operation of the Clause that would need to be overcome for it to be meaningfully applied. The following sections of this paper will seek to address these foreseen objections in a way that makes self-evident the applicability of the Clause to grey zone operations and illustrates a foundational view of operation that would see the Clause function with capabilities worthy of its mandate.

As international legal regimes are fragmented in their ability to govern cyber activity, hostile actors can purposefully design operations that exploit the rigid architecture of current international law to evade classification, subverting any accountability under current branches of international law.³³ This underscores how pre-conflict data poisoning operations are not only tailored to avert interaction with existing law, but are also designed to exploit the opacity between IHL and IHRL through a conflict/non-conflict binary paradigm that switches entire bodies of international law on and off.³⁴ In explanation of the notion of sitting between legal frameworks, Michael N. Schmitt discusses how Russia has exploited international legal principles and rules in its conflicts with Ukraine and cyber interactions with the USA, allowing it to act on “legal margins” masking involvement and avoiding the implications of *jus ad bellum* and use-of-force prohibitions.³⁵ Schmitt also discusses the ICRC’s view that “an operation designed to disable an object – for example a computer or a computer network – constitutes an attack under the rules on the conduct of hostilities, whether or not the object is disabled through kinetic or cyber means.”³⁶ The ICRC also states, “in order to differentiate between operations that

³² *Ibid.*

³³ Hitoshi Nasu, “Hybrid Threats and Grey Zone Conflict Symposium – Challenges in the Twilight of International Law”, 30 October 2024, *Articles of War*, available at: <https://lieber.westpoint.edu/challenges-twilight-international-law/>.

³⁴ Advisory Council on International Affairs, *Hybrid Threats and Societal Resilience*, Advisory Report No. 126, The Hague, 4 June 2024, pp. 34–37, available at: <https://www.adviesraadinternationalelvraagstukken.nl/documenten/2024/06/04/hybride-dreigingen-en-maatschappelijke-weerbaarheid>.

³⁵ Michael N. Schmitt, “Grey Zones in the International Law of Cyberspace”, *Yale Journal of International Law Online*, Vol. 42, No. 2, 2017, p. 1.

³⁶ International Committee of the Red Cross, *International Humanitarian Law and the Challenges of Contemporary Armed Conflicts*, Report No. 32IC/15/11, Geneva, October 2015, p. 41 (hereinafter *IHL Challenges of Contemporary Armed Conflicts*); M. Schmitt, above note 35, p. 18.

amount to attacks and those that do not, it has been suggested that the criterion of ‘inconvenience’ should be relied upon.”³⁷ However, what constitutes such inconvenience is not defined, and no such terminology exists in current IHL.³⁸ This further illustrates the significant grey area within which cyber operations can be undertaken. Schmitt goes on to discuss that during actions of conflict, “it is well accepted that when the destruction or alteration of data results in injury or damage, the cyber operation in question is subject to the full gamut of attack rules.”³⁹ In this, he notes a difference in evaluations of whether data is perceived as an object under the law.⁴⁰

To explain this in relation to data poisoning, it is important to retain an understanding that it is not the ML or AI software or system that is directly interacted with but the data it consumes that is poisoned, leaving the system intact even if corrupted or no longer fit for purpose.⁴¹ However, this view opens new questions as to what an intact ML or AI system actually represents. If the system remains but cannot perform tasks to the standard or capability intended, is it rendered totally inoperative, or even merely inconsistent in its operation? Is the functional system itself indirectly altered through interaction with intentionally poisoned data and, by definition, damaged due to the data poisoning operation?

It seems logical to argue that, if poisoned data can simply be removed, then the affected system itself remains intact. Yet this misunderstands the nature of modern AI and machine learning systems. The problem is not merely the continued presence of corrupted data within a dataset, but the fact that the model itself has already been shaped by that poisoned training environment. Once the system has internalized the malicious associations, biases, or vulnerabilities introduced during training, removing the original data may do little to reverse the corruption. The influence of the poisoned data may persist within the model’s parameters, decision-making pathways, and outputs long after the dataset itself has been cleansed. In practice, meaningful remediation may require retraining the model entirely, or rebuilding a new model from uncontaminated data. This significantly complicates any argument that the underlying system remains functionally undamaged, particularly where operators cannot readily identify the corruption, isolate its effects, or restore the model’s integrity with confidence.⁴² If the poisoned data is not identified, and its effect on the system’s function not recognized, it seems the system would remain ‘damaged’ indefinitely. Whilst this discussion may be intrinsically tied

³⁷ *IHL Challenges of Contemporary Armed Conflicts*, above note 36, p. 42; M. Schmitt, above note 35, p. 36.

³⁸ *IHL Challenges of Contemporary Armed Conflicts*, above note 36, p. 42; M. Schmitt, above note 35, p. 36.

³⁹ M. Schmitt, above note 35, p. 18.

⁴⁰ *Ibid.*

⁴¹ Heather Harrison Dinniss, “The Nature of Objects: Targeting Networks and the Challenge of Defining Cyber Military Objectives”, *Israel Law Review*, Vol. 48, No. 1, 2015, p. 42.

⁴² *Ibid.*; Michael N. Schmitt (ed.), *Tallinn Manual on the International Law Applicable to Cyber Warfare*, Cambridge University Press, Cambridge, 2013, pp. 93–94.

to that of whether data represents an object, it is interesting to consider that, regardless of the object-based recognition of data, if the system that allows the object to function is damaged by its utilisation of intentionally corrupted data, is the object not damaged?

Although not a direct topic of this paper or one that will be explored in depth, a contributing factor to the failure to attribute substantial declarations of attacks to data poisoning operations is that it remains unclear whether data alone qualifies as an object under IHL.⁴³ If data is not considered an object, then cyber operations that alter or damage data without impacting the underlying IT systems or causing physical effects (if this is even possible as discussed above) would not be covered by the principle of distinction or other targeting rules.⁴⁴ As a result, such operations may be left to interact with data in any way deemed operationally beneficial.⁴⁵ It is evident that although there are many adapting and evolving positions, IHL's protections towards objects do not currently apply to computer data and information.⁴⁶ Changing perspectives and academic positions cannot be ignored, especially in relation to adapting State policy considerations; however, these are inadequate in constructing a firm determination of existing law.⁴⁷ This doctrinal issue remains secondary to the analysis of this paper as the issue of pre-conflict regulation and governance arises regardless of data's classification or lack thereof as an object. In investigating applicable regulatory structures prior to armed conflict, the question does not hinge upon whether the poisoned dataset amounts to an 'object' or the qualification of such poisoning as an attack, but rather whether the deliberate corruption of data in the form of data poisoning is compatible with fundamental humanitarian principles.

If left unchecked, anticipatory interference could cripple future military capabilities and state infrastructure, pre-empting conflict and forcing international disunion. The lack of accountability and deterrence merely proliferates the use of such tactics, which if utilized, can lead to an erosion of trust, escalation of tension and "threaten international peace and security."⁴⁸ The grey zone blurs what was a reliable and ever-present distinction between peace and war, which has long allowed the independent and coherent function of multiple branches of international law. The more that operations such as data poisoning are tolerated under a perception of

⁴³ H. Harrison Dinniss, above note 41, p. 42.

⁴⁴ Robin Geiss and Henning Lahmann, *Protection of Data in Armed Conflict*, Working Papers, Geneva Academy, Geneva, February 2021.

⁴⁵ *Ibid.*

⁴⁶ Ori Pomson, "'Objects'? The Legal Status of Computer Data under International Humanitarian Law", *Journal of Conflict and Security Law*, Vol. 28, No. 2, 2023, p. 387.

⁴⁷ *Ibid.*

⁴⁸ General Assembly, Report of the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security, 76th Session, Agenda Item 91, UN Doc. A/76/135, 14 July 2021, p. 7, para. 9.

peacetime activity, the less meaningful the term armed conflict will become in a further technologically dependent age. This destabilizes not just a framework of deterrence and punishment, but also the coherence of a concept of sovereignty and non-intervention. The existing system may have functioned honourably when the only way to release water from a dam, infiltrate a federal reserve or immobilize a state was through the force of destruction-based machinery and personnel-based military operations. However, where all of these actions can be undertaken without armed force and prepared years prior in anticipation of future conflict, traditional IHL and IHRL frameworks no longer provide a constraint on gross militaristic and adversarial action.

III. A SOLUTION: THE MARTENS CLAUSE?

As foreshadowed, it is clear that some aspects of IHL can spill the banks of the peace/war paradigm.⁴⁹ A wider literature highlights that most discussions around the Martens Clause currently regard it as a creature of IHL and thus confined to its application.⁵⁰ Under this view, once interactions fail to meet the threshold of armed conflict, the Clause in any of its forms would be inapplicable. As has also been discussed, armed conflict appears comparatively comfortable in its regulation of cyber actions such as data poisoning once conflict exists, meaning the real issue remains the pre-conflict regulation of the upstream actions. This paper's view of the Clause's applicability outside of IHL is founded on Additional Protocol I and its wording.

For the benefit of contextualising the position, Article 3 of Additional Protocol I states, "the Conventions and this Protocol shall apply from the beginning of any situation referred to in Article 1 of this Protocol".⁵¹ Article 1(3) states, "[t]his Protocol, which supplements the Geneva Conventions of 12 August 1949 for the

⁴⁹ Christopher Greenwood, "The Concept of War in Modern International Law", *International and Comparative Law Quarterly*, Vol. 36, No. 4, 1987, p. 283.

⁵⁰ Israel, "Comments from the State of Israel on the International Law Commission's Draft Principles on the Protection of the Environment in Relation to Armed Conflicts as Adopted by the Commission in 2019 on First Reading" (Comments Submitted to the International Law Commission), 2020, para. 56, available at: http://legal.un.org/docs/?path=../ilc/sessions/73/pdfs/english/poe_israel.pdf&lang=E; United Kingdom, "Comments and Observations on the Draft Principles on Protection of the Environment in Relation to Armed Conflicts" (Comments Submitted to the International Law Commission), 2021, paras. 21–23, available at: http://legal.un.org/docs/?path=../ilc/sessions/73/pdfs/english/poe_uk.pdf&lang=E; United States, "Comments of the United States on the International Law Commission's Draft Principles on the Protection of the Environment in Relation to Armed Conflicts" (Comments Submitted to the International Law Commission), 6 October 2021, pp. 10–11, available at: http://legal.un.org/docs/?path=../ilc/sessions/73/pdfs/english/poe_us.pdf&lang=E; see further Dieter Fleck, "The Martens Clause and Environmental Protection in Relation to Armed Conflicts", *Goettingen Journal of International Law*, Vol. 10, No. 1, 2020, 250.

⁵¹ AP I, Art. 3.

protection of war victims, shall apply in the situations referred to in Article 2 common to those Conventions.”⁵² Common Article 2 of the Geneva Conventions states, “[i]n addition to the provisions which shall be implemented in peacetime, 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.”⁵³

Here, this paper focuses on the statement “[i]n addition to the provisions which shall be implemented in peacetime”.⁵⁴ Whilst it is recognized this refers generally to articles on dissemination of information and restriction of the use of emblems, this paper argues that Article 1 of Additional Protocol I, read with Common Article 2 of the Geneva Conventions, includes itself in such peacetime application through its direct and intentionally broad identification of “cases not covered by this Protocol”,⁵⁵ as well as through Additional Protocol I’s requirement that “[t]he High Contracting Parties undertake to respect and to ensure respect for this Protocol in all circumstances.”⁵⁶ As such, the Clause may remain active in peacetime through a combination of this ambiguity and the explicit application to “cases not covered”.⁵⁷ Not only this, but it is clearly apparent that the contemporary fulfilment of the Clause’s purpose both requires and actively dictates its application externally to the confines of traditional IHL.

Additional Protocol I complements the Geneva Conventions and sits within their broader compliance architecture. Common Article 2 expressly acknowledges that some provisions operate “in peacetime”, which supports the idea that the law of armed conflict is partly operationalized in advance through standing duties and anticipatory compliance disciplines.⁵⁸ The sentiments of the Clause are further removed from the confines of general provisions and principles in IHL treaty architecture through the articles on denunciation in each of the four Geneva Conventions. Article 63 of Convention I, Article 62 of Convention II, Article 142 of Convention III and Article 158 of Convention IV all state:

The denunciation shall have effect only in respect of the denouncing Power. It shall in no way impair the obligations which the Parties to the conflict shall remain bound to fulfil by virtue of the principles of the law of nations, as they result from the usages established among civilized

⁵² *Ibid.*, Art. 1(3).

⁵³ Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field of 12 August 1949, 75 UNTS 31 (entered into force 21 October 1950), Art. 2.

⁵⁴ *Ibid.*

⁵⁵ AP I, Art. 1(2).

⁵⁶ *Ibid.*, Art. 1(1).

⁵⁷ *Ibid.*, Art. 1(2).

⁵⁸ GC I, Art. 2.

peoples, from the laws of humanity and the dictates of the public conscience.⁵⁹

Whilst this speaks directly to parties to a conflict, it shows that the obligation to the principles of humanity and public conscience continues to exist outside the parameters of the Conventions through their endurance of any denunciation. Furthermore, at no point do these articles limit those obligations to any conflict currently ongoing.

In *Corfu Channel (UK v. Albania)*, the ICJ when explaining Albania's legal duty to warn British ships of the mines in its waters, found that among other things, the duty arose from an obligation to "elementary considerations of humanity" which the Court described as "even more exacting in peace than in war".⁶⁰ This demonstrated international law's recognition of similar humanity-based obligations existing outside armed conflict.

Additionally, as discussed, the ICJ in the Nuclear Weapons Advisory Opinion illustrated, that Additional Protocol I when adopted, acted to represent pre-existing customary law including the Clause and that the absence of codified law in regard to a form of weaponry being detailed in existing "does not permit the drawing of any legal conclusions relating to the substantive issues which the use of such weapons would raise."⁶¹ In addition to this, and in illustration of the way in which capabilities such as those possessed by the Clause are both able and obliged to police the gaps in law, the Court also stated "humanitarian law, at a very early stage, prohibited certain types of weapons either because of their indiscriminate effect on combatants and civilians or because of the unnecessary suffering caused to combatants, that is to say, a harm greater than that unavoidable to achieve legitimate military objectives. If an envisaged use of weapons would not meet the requirements of humanitarian law, a threat to engage in such use would also be contrary to that law."⁶²

In explanation of the decision, President Bedjaoui stated that "[i]nternational humanitarian law is a particularly exacting corpus of rules, and these

⁵⁹ GC I, Art. 63; Geneva Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea of 12 August 1949, 75 UNTS 85 (entered into force 21 October 1950), Art. 62; Geneva Convention (III) relative to the Treatment of Prisoners of War of 12 August 1949, 75 UNTS 135 (entered into force 21 October 1950), Art. 142; Geneva Convention (IV) relative to the Protection of Civilian Persons in Time of War of 12 August 1949, 75 UNTS 287 (entered into force 21 October 1950), Art. 158.

⁶⁰ International Court of Justice, *Corfu Channel (United Kingdom v. Albania)*, Judgement (Merits), *ICJ Reports 1949*, p. 22.

⁶¹ *Legality of the Threat or Use of Nuclear Weapons*, above note 13, p. 259, para. 84.

⁶² *Legality of the Threat or Use of Nuclear Weapons*, above note 13, p. 257, para. 74.

rules are meant to be applied in all circumstances.”⁶³ Here the emphasis is that IHL’s obligations are not loose and cannot be suspended by a particular set of circumstances. This paper suggests that the same view applies in evaluation of peacetime. Just because armed conflict is not present does not excuse a state from compliance with the humanitarian obligations of IHL, specifically those achievable through utilisation of the Clause.

The next foreseeable opposition to the application of the Clause in the proposed format is the general uncertainty over its function and ability. In pre-emptive response, it is easiest to identify what the Clause does not apply to: anything already codified. This foundational principle of the Clause is uncontroversial and illustrates it as a backstop operating where codified coverage runs out. With this view, the Clause functions like a net that catches what slips through the confines of existing law. However, that mandate is not stationary in time. As echoed in the views of the ICJ, the Clause’s real significance and likely prominence in a world where treaty-making continues to slow is its ability to operate when technology and capabilities evolve faster than codified law can be developed.⁶⁴

A further step in sketching the application of the Clause is to explain that its wording and implied mandate do not only call it to action when the stakes are high; rather it, exists as an ongoing check on legality sitting at the boundary of what is yet to be codified.

If there are no codified rules as to a particular technology, action, or capability, then the existence of the Clause illustrates that this does not, by virtue of absence, make the technology, action, or capability legally permitted. Here, the Clause’s existence sits in direct opposition to the *Lotus* principle and *a contrario* thinking.⁶⁵ This principle, taking its name from *The Case of the S.S. ‘Lotus’*, decided by the Permanent Court of International Justice in 1927, is commonly used to describe the view that States may act unless international law expressly prohibits the conduct in question.⁶⁶ Accordingly, the Clause functions in diametric opposition, implying that conduct is not presumptively permitted through a mere lack of express prohibition.

Illegality without formal codified prohibition of conduct, is not a foreign concept. In the *K.W.* case,⁶⁷ the defendant was accused of intentionally injuring

⁶³ Ibid., Declaration of President Bedjaoui, p. 272, para. 19.

⁶⁴ Ibid., p. 257, para. 78.

⁶⁵ Michael Salter, “Reinterpreting Competing Interpretations of the Scope and Potential of the Martens Clause”, *Journal of Conflict and Security Law*, Vol. 17, No. 3, 2012, p. 403.

⁶⁶ Permanent Court of International Justice, *The Case of the S.S. ‘Lotus’ (France v. Turkey)*, Judgment, 7 September 1927, *PCIJ Reports*, Series A, No. 10.

⁶⁷ Conseil de guerre de Bruxelles (Military Court of Brussels), *K.W.*, Opinion and Judgment, 8 February 1950, *Revue de droit pénal et de criminologie*, Vol. 30, No. 5, 1949–50, p. 562.

civilians who were resistant to the German occupation of Belgium. During the proceedings, the Court noted the defendant's alleged actions were not expressly forbidden by the Hague Regulations,⁶⁸ instead finding that through utilisation of the Clause, principles of international law such as the *Universal Declaration of Human Rights* could still be applied to illustrate the illegality of the actions. Through this, the Court ruled the defendant had breached Art. 5 of the Universal Declaration and thus his actions constituted a violation of the customs of war.⁶⁹

For the Clause to operate in this way, it is the responsibility of the state to evaluate and justify the proposed action's compliance with the principles of humanity and the dictates of public conscience. This embodiment of the Clause does not require a baseline trigger beyond a case not being covered by existing international law, as it is supported by its wording and the mandate which Additional Protocol I gives it to remain active in evaluating everything untouched by codified law. The Clause creates a presumption in favour of protection. It is the duty of the state seeking to take action not yet governed by codified law to ensure and justify that such action does not breach these thresholds.

For the Clause to act as proposed above, it requires a foundation to be supplied for the baseline meanings or practical invocation of the principles of humanity and dictates of public conscience. For this, the best available illustrations come from international judicial decisions.

In *Kupreškić*, the Court did not attempt to claim that individually the principles of humanity and dictates of public conscience, had "been elevated to the rank of independent sources of international law".⁷⁰ Instead, it showed that the Clause requires, at minimum, recourse to the principles of humanity and dictates of public conscience where a rule of IHL is not sufficiently rigorous or precise, so that the scope of the rule can be defined by reference to them.

This demonstrated that whilst the Court treated the principles of humanity and dictates of public conscience as being anchored in pre-existing law, their operation was not confined to the strict limits of codified rules. Here, a customary or contemporary position may utilize the Clause and relevant considerations to constrain action that is *prima facie* non-restricted when viewed only in comparison to existing law. The Court continued to illustrate the Clause's capacity to interact with

⁶⁸ Antonio Cassese, "The Martens Clause: Half a Loaf or Simply Pie in the Sky?", *European Journal of International Law*, Vol. 11, No. 1, 2000, p. 207.

⁶⁹ Marcus Hickleton, "Shaky Foundations: Killer Robots and the Martens Clause", *Perth International Law Journal*, Vol. 4, 2019, p. 52.

⁷⁰ International Criminal Tribunal for the Former Yugoslavia, *Prosecutor v. Kupreškić et al.*, Case No. IT-95-16-T, Judgment (Trial Chamber), 14 January 2000, p. 206, para. 525 [hereinafter *Kupreškić*].

various forms of international law, showing how it can still define boundaries even when hard law or consistent state practice is lacking.⁷¹

The Court found that use of the Clause does not require evident state behaviour in the form of consistent state practice. Instead, identifying that States' belief about what the law requires may be more important than the actions they take, and that binding law can emerge and develop from such a process.⁷² Further to this, the court did not treat public conscience as an incalculable moral idea. It demonstrated how *opinio necessitatis* can be evidenced in practice through military manuals, ratification patterns, ICRC memos and UN General Assembly resolutions.⁷³ The court used these *opinio necessitatis* identifiers as indicators of public conscience to support its implementation of the Clause. The court did not survey or attempt to codify global opinions. Rather, it evaluated how law applies, and how it ought to apply. This showed that the principles of humanity and dictates of public conscience are not freestanding morality-based vetos but are instead indicative criteria used to gauge the acceptability of the effects and interactions with humanitarian protections of actions which are not governed by current law.

In understanding principles of humanity, the identification of these principles should not be confined to IHL or IHRL alone, as the conduct being assessed will not yet be codified and contained within any single international legal apparatus; it would be an absurd notion to restrict the application of humanitarian principles to their required identification with any arbitrary legal paradigm. From this, principles of humanity should be drawn from existing law, custom, and scholarship structured to protect humanitarian interests across a broad range of international legal structures. To limit the Clause's applicability in this respect to only certain applications or structures would again, negate the breadth of its wording and mandate. As Jonathan Kwik has noted:

One can easily imagine the backlash and media attention generated from a perimeter defence turret being briefly fooled into shooting defenceless villagers, because an adversary handed out innocuous bags at the market with adversarial patterns on them.⁷⁴

This is exactly the kind of "moral outrage" that makes the most sense to channel via the Martens Clause. Even if lawyers could construct technical arguments that such conduct falls outside existing prohibitions, the instinctive reaction of most societies would remain the same: the act is morally repugnant. That reaction matters. The Martens Clause exists precisely to capture this category of conduct, where emerging technologies outpace codified rules but nevertheless offend fundamental

⁷¹ *Ibid.*, p. 207, para. 527.

⁷² *Ibid.*

⁷³ *Ibid.*, p. 210, para. 532.

⁷⁴ J. Kwik, above note 3, p. 744.

expectations of humanity and conscience. In this respect, the Clause is not a vague moral appeal, but a mechanism designed to ensure that legal restraint survives even when adversaries weaponize ambiguity itself.

From this, a combination of existing IHL, IHRL and customary positions should facilitate the utilisation of the principles of humanity in the Clause's application. As such, in illustration of its guiding principles, to anchor any risk-based assessment, the International Covenant on Civil and Political Rights states that "[e]very human being has the inherent right to life. This right shall be protected by law. No one shall be arbitrarily deprived of his life."⁷⁵ The facilitation of this right expands to an obligation where the threshold of the requirement for a state to act is met by the possibility of a loss to life not just the eventuation of a loss of life.⁷⁶ Such an obligation to ensure facilitation of this right is also engaged by a threat to infrastructure, social requirements and civilian systems which when impeded or restricted would pose a foreseeable threat to survival.⁷⁷ Additional to this and in argument for the adequate protection of this right in the contemporary world, it would be paradoxical to confine the civilian's "general protection against dangers arising from military operations" provided by Additional Protocol I Article 51 to scenarios of armed conflict.⁷⁸

IV. CONCLUSION

Pre-conflict data poisoning exposes a structural weakness in international law. Existing legal frameworks remain heavily dependent upon a binary distinction between peace and war, yet contemporary cyber operations are deliberately designed to exploit the space between them. Data poisoning does not require kinetic force, immediate destruction, or even visible disruption to produce strategic effect. Instead, it enables adversaries to quietly manipulate the informational environment years before conflict emerges, embedding vulnerabilities into the systems upon which militaries, governments, and civilian societies increasingly depend. In doing so, these operations undermine the coherence of legal regimes built around temporally identifiable moments of armed conflict.

It is within a gap of strict interpretation of the two paradigms of international law, that the Martens Clause becomes legally and strategically significant. Its historical function has been to ensure that technological innovation does not outpace legal restraint, as illustrated in the ICJ's Nuclear Weapons Advisory Opinion and in

⁷⁵ International Covenant on Civil and Political Rights, 999 UNTS 171, 16 December 1966 (entered into force 23 March 1976), Art. 6(1).

⁷⁶ Human Rights Committee, General Comment No. 36: Article 6: Right to Life, UN Doc. CCPR/C/GC/36, 30 October 2018, p. 2, para. 7.

⁷⁷ *Ibid.*, p. 6, para. 26.

⁷⁸ AP I, Art. 51(1).

subsequent jurisprudence.⁷⁹ It clearly **exists**, but the nature of its relevance is the matter for debate. It is the opinion of the authors that, if a State merely regards the Martens Clause as a principle (rather than a legal obligation), then it is not effective. It actually risks weakening the structure of IHL, and furthering the divide between those that do, and those that think.

Far from a symbolic relic of humanitarian idealism, the Clause exists precisely to govern situations not yet contemplated by codified law. Whilst applying the Clause does not create bright-line rules or resolve attributional or enforcement issues, applying it in a manner consistent with the preamble to the 1899 Hague Convention, Additional Protocol I and decisions of the ICJ, allow it to provide a principled basis for restraint on action both prior to and during the existence of armed conflict.⁸⁰ By invoking the principles of humanity and dictates of public conscience, the Clause provides a framework capable of evaluating emerging technologies whose effects transcend traditional categories of peace, war, civilian infrastructure, and military capability. In this respect, data poisoning resembles previous moments of technological rupture, where law struggled to keep pace with innovation, but humanitarian principles remained operative.

Ultimately, the challenge posed by pre-conflict data poisoning is not simply one of cyber regulation, but of legal imagination. International law cannot remain credible if hostile actions capable of shaping future conflict are permitted to evade scrutiny merely because they occur before the formal outbreak of hostilities. The Martens Clause offers the most coherent existing mechanism for preventing that outcome. By applying the principles of humanity and the dictates of public conscience to the manipulation of datasets that shape automated decision-making, the Clause restores a measure of normative oversight to an otherwise unregulated domain of strategic competition. In doing so, it reaffirms a central proposition of humanitarian law: that the evolution of technology must never be allowed to outpace the protection of humanity.

⁷⁹ *Legality of the Threat or Use of Nuclear Weapons*, above note 13; *Kupreškić*, above note 70; A. Cassese, above note 68.

⁸⁰ Hague Convention II, above note 18, Preamble; AP I, Art. 1(2); *Legality of the Threat or Use of Nuclear Weapons*, above note 13.