Artificial Intelligence and the prohibition on the use of force: intention and causation

INTERNATIONAL LAW STUDIES

Published Since 1895

Artificial Intelligence and the Prohibition on the Use of Force: Intention and Causation

Russell Buchan

107 INT'L L. STUD. 9 (2026)

Volume 107 2026

Published jointly by the University of Reading and the Stockton Center for International Law

ISSN 2375-2831

Artificial Intelligence and the Prohibition on the Use of Force: Intention and Causation

Russell Buchan[*]

CONTENTS

* Russell Buchan is Professor of International Law at the University of Reading (United Kingdom), Senior Fellow at the NATO Cooperative Cyber Defence Centre of Excellence (Estonia), and former Senior Fellow at the Lieber Institute at the U.S. Military Academy at West Point. This article has benefited from discussions with Jonathan Kwik, Massimo Lando, Asaf Lubin, Iñaki Navarrete, Marco Roscini, and Nicholas Tsagourias. All errors remain my own. Contact: r.buchan@reading.ac.uk.

The thoughts and opinions expressed are those of the author and not necessarily those of the U.S. government, the U.S. Department of the Navy, or the U.S. Naval War College.

I. INTRODUCTION

Artificial intelligence (AI) describes programs, systems, and machines that are capable of performing tasks normally undertaken by humans. AI exists on a spectrum because different types of behavior require different levels of intelligence. AI is limited when it is controlled by an algorithm that requires the system to perform specific tasks in pre-determined scenarios and where the completion of these tasks is overseen by a human. At the other end of the spectrum, AI can utilize machine learning to accomplish broad objectives in complex and dynamic environments without the need for additional human input. While this type of AI is developed and deployed by humans and thus operates within a framework of planned behavior, it can make and execute decisions and is to a large extent self-governing.[1]

The benefits of AI are enormous and this technology is now widely used in various sectors including healthcare, transport, industry, and education. States have also recognized the advantages of AI and they frequently deploy AI-enabled systems in the kinetic and cyber domains to conduct operations against and within other States in pursuit of their national security objectives.[2]

A well-known drawback of AI is the potential for "unintended engagements," that is, where AI-enabled systems engage in activities that were not intended by those who developed and deployed them.[3] There are various reasons for why AI-enabled systems can carry out unintended engagements.

---

1. "[AI describes] systems that display intelligent behaviour by analysing their environment and taking actions—with some degree of autonomy—to achieve specific goals." EUROPEAN COMMISSION HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, A DEFINITION OF AI: MAIN CAPABILITIES AND SCIENTIFIC DISCIPLINES (Dec. 18, 2018), https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf.

2. OFFICE OF INTELLIGENCE AND ANALYSIS, DEP'T OF HOMELAND SECURITY, HOMELAND THREAT ASSESSMENT 2024, at 18 (2023), https://www.dhs.gov/sites/default/files/2023-09/23_0913_ia_23-333-ia_u_homeland-threat-assessment-2024_508C_V6_13Sep23.pdf; NAT'L CYBER SECURITY CENTRE, THE NEAR-TERM IMPACT OF AI ON THE CYBER THREAT 3 (Jan. 24, 2024), https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat.

3. U.S. DEP'T OF DEFENSE, DOD DIRECTIVE 3000.09, AUTONOMY IN WEAPONS SYSTEMS (Jan. 25, 2023), https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.

AI may be trained on biased, incomplete, discrepant, low-quality, flawed, or synthetic data.[4] This leads to AI being "brittle"[5] and prone to unexpected behavior when deployed into environments that it has not previously encountered (known as "distributional shift").[6] In fact, even comprehensive training has its limitations where AI is capable of machine learning because it can incrementally acquire, update, and exploit knowledge throughout its lifecycle. Thus it can engage in what is called "emergent behaviour,"[7] which makes these high-powered technologies "unpredictable by design."[8] Moreover, AI is often described as a "black box" technology.[9] This opacity prevents

---

4. "[Data can be] badly curated, making it challenging, time consuming and cost intensive to access sufficient levels of machine-ready data to train AI models. Data ownership and the ability to share data can also present significant challenges." UK Ministry of Defence, Written Evidence (AIW0035), ¶ 9.1 (June 2023), https://committees.parliament.uk/writtenevidence/121708/html/. On the risks of using biased data to train AI systems for military use, *see* ALEXANDER BLANCHARD & LAURA BRUNN, BIAS IN MILITARY ARTIFICIAL INTELLIGENCE (Stockholm Int'l Peace Research Inst. Background Paper, Dec. 2024), https://www.sipri.org/sites/default/files/2024-12/background_paper_bias_in_military_ai_0.pdf.

5. Brittleness occurs when an "algorithm cannot generalize or adapt to conditions outside a narrow set of assumptions." Mary L. Cummings, *Rethinking the Maturity of Artificial Intelligence in Safety-Critical Settings*, 42 AI MAGAZINE 6, 7 (Spring 2021).

6. Zachary Arnold & Helen Toner, *AI Accidents: An Emerging Threat. What Could Happen and What to Do*, CENTER FOR SECURITY AND EMERGING TECHNOLOGY, at 7 (July 2021), https://cset.georgetown.edu/publication/ai-accidents-an-emerging-threat/.

> The unpredictability of some AI systems, particularly when applied to new and challenging environments, increases the risks that unforeseen issues may arise with their use. The relative difficulties with interpreting how some forms of AI systems learn and make decisions present new challenges for the testing, evaluation and certification of such systems. In addition, the high potential impact of AI-enabled systems for Defence raises the stakes for potential side effects or unintended consequences, particularly when they could cause harms for those interacting with them.

UK MINISTRY OF DEFENCE, AMBITIOUS, SAFE, RESPONSIBLE: OUR APPROACH TO THE DELIVERY OF AI-ENABLED CAPABILITY IN DEFENCE (June 15, 2022), https://www.gov.uk/government/publications/ambitious-safe-responsible-our-approach-to-the-delivery-of-ai-enabled-capability-in-defence/.

7. Jakub Kraus, *Overview of Emergent and Novel Behaviour in AI Systems*, CENTER FOR AI POLICY (Mar. 26, 2024), https://www.centeraipolicy.org/work/emergence-overview.

8. VINCENT BOULANIN ET AL., LIMITS ON AUTONOMY IN WEAPONS SYSTEMS: IDENTIFYING PRACTICAL ELEMENTS OF HUMAN CONTROL 7 (June 2020), https://www.sipri.org/sites/default/files/2020-06/2006_limits_of_autonomy_0.pdf.

9. AI—at least in its modern, sophisticated form—is referred to as a "black box" in the sense that its internal decision-making processes cannot be easily explained and understood. In other words, while the inputs and outputs can be observed, how the technology moves

developers from being able to fully explain and understand why AI operates in the way that it does, and this makes it difficult for them to predict how the technology will behave once deployed.[10] Furthermore, malicious third-parties can use adversarial tactics against AI-enabled systems, which can lead to these systems carrying out unintended engagements.[11] For example, malicious actors can introduce AI-enabled systems to erroneous data during development and deployment ("data poisoning"[12]), conduct jamming operations that cause AI-enabled systems to malfunction, and launch "evasion attacks" that trick AI-enabled systems into misidentifying objects.[13] All of this means that AI-enabled systems are "deceptively capable"[14]—in short, developers falsely believe that they can accurately predict the activities of AI-enabled systems.[15]

This article examines the application of the prohibition on the use of force under Article 2(4) of the United Nations (UN) Charter 1945[16] to situations where States use AI-enabled systems to conduct operations against

---

from the input to the output is often concealed, complex, and difficult to explain. *See* Matthew Kosinski, *What is Black Box AI?*, IBM (Oct. 29, 2024), https://www.ibm.com/think/topics/black-box-ai; Arthur Holland Michel, *The Black Box, Unlocked: Predictability and Understandability in Military AI*, UNIDIR (2020), https://unidir.org/files/2020-09/BlackBoxUnlocked.pdf.

10. David Beer, *Why Humans Will Never Understand AI*, BBC (Apr. 7, 2023), https://www.bbc.com/future/article/20230405-why-ai-is-becoming-impossible-for-humans-to-understand.

11. *See generally* Marcus Comiter, *Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It*, BELFER CENTER (Aug. 2019), https://live-hksbelfer.pantheonsite.io/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf.

12. Tom Krantz & Alexandra Jonker, *What is Data Poisoning?*, IBM (Dec. 10, 2024), https://www.ibm.com/think/topics/data-poisoning; *see also* Micah Goldblum et al., *Dataset Security for Machine Learning: Data Poisoning, Backdoor Attacks, and Defenses*, ARXIV (last revised Mar. 31, 2021), https://arxiv.org/abs/2012.10544.

13. On evasion attacks, *see* Justin Gilmer et al., *Motivating the Rules of the Game for Adversarial Example Research*, ARXIV (last revised July 20, 2018), https://arxiv.org/abs/1807.06732.

14. Michael Horowitz & Paul Scharre, *AI and International Stability: Risks and Confidence-Building Measures*, CENTER FOR A NEW AMERICAN SECURITY (Jan. 12, 2021), https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures.

15. Roman V. Yampolskiy, *Unpredictability of AI: On the Impossibility of Accurately Predicting All Actions of a Smarter Agent*, 7 JOURNAL OF ARTIFICIAL INTELLIGENCE AND CONSCIOUSNESS 109 (2020).

16. The prohibition on the use of force is also established in customary international law. Military and Paramilitary Activities in and against Nicaragua (Nicar. v. U.S.), Judgment, 1986 I.C.J. 14, ¶ 34 (June 27).

and within other States. Where a State deploys an AI-enabled system with the intention to use force and the intended forcible effects in fact occur as a direct result of the State's conduct, there is little doubt that such conduct breaches Article 2(4). Rather, this article examines whether a breach of Article 2(4) emerges where a State deploys an AI-enabled system that then proceeds to undertake unintended engagements that give rise to forcible effects against or within another State.

To focus this article, it may be useful to sketch out several scenarios where the use of AI-enabled systems can lead to unintended engagements. Consider, for example, the situation where State A deploys an AI-enabled drone into the national airspace of State B with the intention of assassinating a member of a terrorist group. But due to faulty facial recognition technology, imagine that the drone mistakenly targets and kills a civilian in State B. If deployed along a border, it may even be the case that State A's drone mistakenly targets an individual located in the territory of State C. Indeed, this type of targeting error is particularly likely in cyberspace given the integrated nature of this domain and the fact that AI-enabled cyber operations can easily spread to the cyber infrastructure of other States. Moreover, the instantaneous nature of cyberspace increases the chances that third States may be affected because, even if the AI-enabled system is overseen by humans during deployment, there may be little opportunity for them to intervene and terminate the system's activities. Even if there is time to intervene, machine bias may mean that the operator defers to the decision of the system. Finally, consider the situation where, through reverse engineering, a malicious actor figures out the process by which an AI-enabled system identifies military objects. Imagine further that the malicious actor uses (so-called) "stickers" to change the appearance of a civilian object and in doing so lures the system into mischaracterizing it as a military installation and launching an attack against it.[17]

This article examines whether intention and causation are constitutive elements of the prohibition on the use of force. At the outset, it is important to emphasize that this article is concerned with whether causation must be established in order to found a breach of the prohibition on the use of force as a primary rule of international law. This article is not concerned with causation under the secondary rules of State responsibility, that is, to what extent

17. On "stickers," *see* Xingxing Wei et al., *Adversarial Sticker: A Stealthy Attack Method in the Physical World*, ARXIV (last revised Dec. 19, 2022), https://arxiv.org/abs/2104.06728.

a victim State can claim reparations for the injuries caused by a breach of the prohibition on the use of force.

To date, intention and causation have been given relatively little attention in the use of force scholarship. This is perhaps unsurprising given that, historically, States have used kinetic weapons to conduct forcible operations. When these types of weapons are used, intention and causation are usually present and a breach of Article 2(4) can be readily established. This means that there has been little need to consider whether intention and causation are preconditions for a use of force.[18] However, as already mentioned, whether intention and causation are *lex lata* requirements of the prohibition on the use of force is very important when it comes to the use of AI-enabled systems (and, indeed, to the use of new and emerging technologies more generally).

This article proceeds as follows. Part II sets the scene by briefly outlining what qualifies as a use of force under Article 2(4) of the UN Charter. Part III examines whether Article 2(4) is based on subjective or objective responsibility. Part IV analyzes the question of causation in the context of Article 2(4). Part V offers conclusions.

II. A PRIMER: WHAT QUALIFIES AS "FORCE"?

While this is not the place to engage in a lengthy discussion of when State conduct can amount to a use of force under Article 2(4) of the UN Charter,[19] it is first necessary to define the concept of "force" and develop a better understanding of how it applies to the use of AI-enabled systems.

18. "[I]n the context of the use of force through conventional weapons . . . the task of establishing a causal chain or link between the use of such a weapon and death, physical injury or destruction is straightforward." PRIYA URS ET AL., THE INTERNATIONAL LAW PROTECTIONS AGAINST CYBER OPERATIONS TARGETING THE HEALTHCARE SECTOR 52 (Feb. 2023), https://www.elac.ox.ac.uk/wp-content/uploads/2023/04/ELAC-Research-Report_International-Law-Protections-against-Cyber-Operations-Targeting-the-Healthcare-Sector.pdf. "Causation at the stage of breach is rarely discussed in international jurisprudence for the simple reason that it is usually obvious that the adverse consequences that gave rise to the breach were the result of the respondent state's conduct or omission." Henning Lahmann, *Infecting the Mind: Establishing Responsibility for Transboundary Disinformation*, 33 EUROPEAN JOURNAL OF INTERNATIONAL 411, 423 (2022).

19. For a discussion, *see* RUSSELL BUCHAN & NICHOLAS TSAGOURIAS, REGULATING THE USE OF FORCE IN INTERNATIONAL LAW: STABILITY AND CHANGE ch. 2 (2021).

Various provisions of the UN Charter indicate that the prohibition on the use of force applies to *armed* force,[20] and this interpretation is supported by the *travaux préparatoires* of Article 2(4).[21] Armed force requires the use of a weapon and, as the International Court of Justice (ICJ) has explained, Article 2(4) applies to the use of any weapon (e.g., a conventional, cyber, or AI-enabled weapon) provided the requisite effects are produced.[22] The critical question, then, is what *effects* qualify as force for the purpose of Article 2(4).

Incontrovertibly, State-backed operations causing physical harm to people or property can constitute a use of force. For example, a State can commit a breach of Article 2(4) where it uses an AI-enabled drone to launch missile strikes against targets within another State and in doing so causes harm to people or property. It is equally clear that cyber operations can amount to a use of force where they cause physical damage, such as the use of AI-enabled software to conduct a cyber attack against air traffic control services that causes airplanes to crash.[23]

A growing number of States maintain that cyber operations causing online virtual (non-physical) harm can qualify as a use of force where the effects produced are comparable to traditional kinetic attacks rising to the level of a use of force.[24] This approach makes sense given that States are nowadays hugely reliant on an effective and functioning cyberspace and thus online harm to computer networks, systems, and data can be as damaging as offline physical harm. Accordingly, cyber operations that impair the functionality of computer networks and systems, or that modify or delete data, can cross the use of force threshold.[25]

---

20. *See* U.N. Charter arts. 41, 43, 44, 46, 47 (referring to "armed force"). The preamble to the Charter also refers to "armed force."

21. *See, e.g.*, 6 DOCUMENTS OF THE UNITED NATIONS CONFERENCE ON INTERNATIONAL ORGANIZATION 559 (1945).

22. Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. 226, ¶ 39 (July 8). The UN General Assembly has explained that "international law, including the Charter of the United Nations . . . applies to matters governed by it that occur throughout all stages of the life cycle of artificial intelligence, including systems enabled by artificial intelligence, in the military domain." G.A. Res. 79/239, ¶ 1 (Dec. 31, 2024).

23. OFFICE OF THE GENERAL COUNSEL, U.S. DEP'T OF DEFENSE, LAW OF WAR MANUAL § 16.3.1 (updated ed. July 2023).

24. *See, e.g.*, EUROPEAN UNION, DECLARATION ON A COMMON UNDERSTANDING OF INTERNATIONAL LAW IN CYBERSPACE 6 (2024); NEW ZEALAND, THE APPLICATION OF INTERNATIONAL LAW TO STATE ACTIVITY IN CYBERSPACE ¶ 7 (2020); COSTA RICA, COSTA RICA'S POSITION ON THE APPLICATION OF INTERNATIONAL LAW IN CYBERSPACE 10–11 (2023).

25. Ireland explains that

Some States go even further and contend that cyber operations producing harmful economic effects—such as those causing stock markets to crash—can be a use of force.[26] However, only a limited number of States take this approach and, in my view, it represents an overly broad reading of Article 2(4). This is because the prohibition on the use of force is designed, fundamentally, to prohibit States from using "violence" in their international relations.[27] Operations causing harmful economic effects are better addressed by other rules of international law such as the principles of sovereignty and non-intervention.

Another issue is whether a de minimis threshold is built into Article 2(4): must the forcible effects (however defined) of an operation be sufficiently serious or grave to trigger a breach of this prohibition? State practice is admittedly divergent on this question. Some commentators[28] claim that the prohibition applies to all State-backed conduct amounting to a use of force because, in order to maintain international peace and security, Article 2(4) is

---

> although present day technology and our heavily digitised world may not have been contemplated at the time of the adoption of the UN Charter, it is appropriate to interpret Article 2(4) as applying to force emanating from cyber operations, notwithstanding the fact that the traditional physical or kinetic element may be lacking in terms of both means and impact.

IRELAND, DEP'T OF FOREIGN AFFAIRS, POSITION PAPER ON THE APPLICATION OF INTERNATIONAL LAW IN CYBERSPACE 18 (2023).

26. Norway explains that "the use of crypto viruses or other forms of digital sabotage against a State's financial and banking system, or other operations that cause widespread economic effects and destabilisation, may amount to the use of force in violation of Article 2(4)." Official Compendium of Voluntary National Contributions on the Subject of How International Law Applies to the Use of Information and Communications Technologies by States Submitted by Participating Governmental Experts in the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security Established Pursuant to General Assembly Resolution 73/266, U.N. Doc. A/76/136*, at 70 (July 13, 2021) [hereinafter 2021 Official Compendium]; *see also* Gov't of the Kingdom of the Netherlands, Letter from the Minister of Foreign Affairs to the President of the House of Representatives on the International Legal Order in Cyberspace 4 (2019); Denmark, *Denmark's Position Paper on the Application of International Law in Cyberspace*, 92 NORDIC JOURNAL OF INTERNATIONAL LAW 446, 451 (2023) [hereinafter Denmark Position]; FRANCE, DROIT INTERNATIONAL APPLIQUÉ AUX OPÉRATIONS DANS LE CYBERESPACE (2019).

27. YORAM DINSTEIN, WAR, AGGRESSION AND SELF-DEFENCE 90 (6th ed. 2017) ("the term 'force' in Article 2(4) must denote violence").

28. Tom Ruys, *The Meaning of "Force" and the Boundaries of the Jus Ad Bellum: Are "Minimal" Uses of Force Excluded from UN Charter Article 2(4)?*, 108 AMERICAN JOURNAL OF INTERNATIONAL LAW 159 (2014).

an "absolute all-inclusive prohibition"[29] that comprehensively protects States from the use of force.

The problem with this interpretation is that it trivializes the application of the prohibition on the use of force; and, after all, it is an *erga omnes* obligation the breach of which can have very serious consequences.[30] The better view is that Article 2(4) applies only to those uses of force that are sufficiently serious to justify the application of the prohibition regardless of whether the harmful effects manifest in the physical or virtual realm.[31] Ultimately, whether this threshold is reached depends on the context and requires a case-by-case assessment. Factors pointing to a sufficiently serious use of force include "the duration of the attack, the nature of the targets attacked, the locations of the targets attacked, and the types of weapons used, while the criterion of effects measures the extent of the damage caused by the attack."[32]

Finally, a tricky question is whether Article 2(4) is breached where a State sends its armed forces into the territory of another State even if they do not use their weapons to cause harm to people, property, or infrastructure. Some commentators argue that such conduct constitutes a use of force and in doing so cite Article 3(e) of the UN General Assembly's Resolution 3314 on the Definition of Aggression (1974).[33] This provision explains that "[t]he use of armed forces of one Sate which are within the territory of another State with the agreement of the receiving State, in contravention of the conditions provided for in the agreement or any extension of their presence in such territory beyond the termination of the agreement," amounts to an act of

---

29. 6 DOCUMENTS OF THE UNITED NATIONS CONFERENCE ON INTERNATIONAL ORGANIZATION 335 (1945).

30. Int'l Law Comm'n, *Draft Conclusions on Identification and Legal Consequences of Peremptory Norms of General International Law (jus cogens), with Commentaries*, concl. 17 cmt. ¶ 1, U.N. Doc A/77/10 (2022), https://legal.un.org/ilc/texts/instruments/english/commentaries/1_14_2022.pdf.

31. Mary Ellen O'Connell, *The True Meaning of Force*, 108 AJIL UNBOUND 141 (2014).

32. African Union Peace and Security Council, Common African Position on the Application of International Law to the Use of Information and Communication Technologies in Cyberspace, ¶ 41 (Jan. 29, 2024), https://papsrepository.africanunion.org/server/api/core/bitstreams/65bdfced-80d9-445b-b57f-31037616ccda/content [hereinafter Common African Position].

33. *See* Certain Activities Carried Out by Nicaragua in the Border Area (Costa Rica v. Nicar.) and Construction of a Road in Costa Rica along the San Juan River (Nicar. v. Costa Rica), Judgment, 2015 I.C.J. 665, 821, 822 (Dec. 16) (separate opinion of Judge Robinson) (¶¶ 47, 50). However, the ICJ's judgment did not examine whether Nicaragua's posting of troops on Costa Rica's territory amounted to a use of force.

aggression. It is correct that the Preamble to the resolution explains that aggression is "the most serious and dangerous form of the illegal use of force."

In my view, the mere dispatch of armed forces into another State's territory does not qualify as a use of force.[34] First, the Definition of Aggression defines the concept of aggression for the purposes of Article 39 of the UN Charter, not Article 2(4). Second, and as I have stressed, Article 2(4) is an effects-based prohibition that covers harm to people, property, or infrastructure. It is for this reason that States do not consider the flying of military aircraft into the national airspace of another State as a breach of Article 2(4).[35] Finally, State practice indicates that deploying software with attack capabilities into the cyber infrastructure of another State does not constitute a use of force irrespective of the scale or significance of the intrusion.[36] For these reasons, deploying an AI-enabled military system into the land territory, territorial waters, national airspace, or cyber infrastructure of another State does not constitute a use of force but may constitute an unlawful threat of force.[37]

III. IS ARTICLE 2(4) BASED ON SUBJECTIVE OR OBJECTIVE RESPONSIBILITY?

Primary rules of international law can be based on subjective or objective responsibility. Subjective responsibility holds States accountable where they are at fault, such as where they act intentionally or negligently. By contrast, objective responsibility holds States accountable without inquiring into

34. Russell Buchan & Nicholas Tsagourias, *The Crisis in Crimea and the Continuing Relevance of the Principle of Non-Intervention*, 19 INTERNATIONAL COMMUNITY LAW REVIEW 165 (2017).

35. Oliver J. Lissitzyn, *The Treatment of Aerial Intruders in Recent Practice and International Law*, 47 AMERICAN JOURNAL OF INTERNATIONAL LAW 559 (1953). In 1960, the Security Council rejected a proposed resolution by the Soviet Union that sought to condemn the United States' flying of a U2 spy plane (piloted by Gary Powers) in Soviet airspace as an act of aggression. U.N. SCOR, 15th Sess., 860th mtg., U.N. Doc. S/PV.860 (May 26, 1960).

36. For a discussion of whether the pre-positioning of cyber assets on another State's cyber infrastructure breaches the principle of non-use of force, *see* Juliet Skingsley, *Cyber-Rattling: Can 'Pre-Positioning' in Cyberspace Amount to a Threat of the Use of Force Under Article 2(4) of the United Nations Charter?*, 11 JOURNAL ON THE USE OF FORCE AND INTERNATIONAL LAW 50 (2024).

37. Article 2(4) of the UN Charter and customary law prohibit the threat of force. On cyber force, *see* Duncan B. Hollis & Tsvetelina van Benthem, *Threatening Force in Cyberspace*, *in* BIG DATA AND ARMED CONFLICT: LEGAL ISSUES ABOVE AND BELOW THE ARMED CONFLICT THRESHOLD (Laura A. Dickinson & Edward E. Berg eds., 2023).

whether they are at fault. The International Law Commission's Articles on State Responsibility make it clear that international law does not contain a hard and fast rule as to whether State responsibility is defined in subjective or objective terms. Instead, whether subjective or objective responsibility is required depends on the construction of the primary rule in question.[38]

The text of Article 2(4) does not provide any clues as to whether the use of force prohibition is based on subjective or objective responsibility. However, most use of force commentators aver that Article 2(4) is based on subjective responsibility and that, in particular, the use of force must be committed intentionally in order for the prohibition to apply.[39] It is important to recognize that intent can take on different meanings and perform different roles when it comes to the use of force prohibition.[40] Three approaches can be discerned.

- *Specific intent*: a State breaches the prohibition on the use of force where it intends to use force against a specific target in another State and in fact engages that target. However, Article 2(4) is not breached where State A intends to use force against State B but for whatever reason hits the wrong target in State B.

- *Hostile intent*: a State breaches the prohibition on the use of force where it intends to use force against a specific State. Provided State A intends to use force against State B, a breach arises even if State A mistakenly attacks the wrong target in State B. But State A does not commit a use of force where it intends to engage in a (non-forcible) cyber surveillance operation against State B but, for whatever reason, the cyber operation ends up causing

38. "Whether responsibility is 'objective' or 'subjective' in this sense depends on the circumstances, including the content of the primary obligation in question. The articles lay down no general rule in that regard." Int'l Law Comm'n, *Draft Articles on Responsibility of States for Internationally Wrongful Acts with Commentaries*, 56 U.N. GAOR Supp. No. 10, art. 2 cmt. ¶ 3, U.N. Doc. A/56/10 (2001), *reprinted in* [2001] 2 YEARBOOK OF THE INTERNATIONAL LAW COMMISSION 26, U.N. Doc. A/CN.4/SER.A/2001/Add.1 (Part 2), https://legal.un.org/ilc/documentation/english/reports/a_56_10.pdf [hereinafter *Articles on State Responsibility*].

39. *See, e.g.*, OLIVIER CORTEN, THE LAW AGAINST WAR 86–87 (2021); CHRISTIAN HENDERSON, THE USE OF FORCE AND INTERNATIONAL LAW 125–29 (2023); ERIN POBJIE, PROHIBITED FORCE: THE MEANING OF 'USE OF FORCE' IN INTERNATIONAL LAW 151 (2024); TERRY GILL & KINGA TIBOR-SZABO, THE USE OF FORCE AND THE INTERNATIONAL LEGAL SYSTEM 64 (2023).

40. "This criterion of intent may obviously lend itself to different interpretations." CORTEN, *supra* note 39, at 78.

forcible effects in State B (e.g., because it wipes critical data or affects the functionality of computer networks and systems), the reason being that State A does not harbor a hostile intent to use force against State B. For the same reason, Article 2(4) is not breached if State A intentionally launches a forcible operation against State B but mistakenly hits a target in State C.

- *General intent*: a State breaches the prohibition on the use of force where it launches an operation with the intent to use force. Where it does, a breach of Article 2(4) emerges regardless of where the forcible effects manifest. Consequently, State A is responsible for a breach of Article 2(4) where it intends to use force against State B but mistakenly hits a target in State C (or, for that matter, in States D, E, or F). However, State A does not breach Article 2(4) where, for example, a technical malfunction leads to a weapons system accidentally launching a missile that goes on to hit another State because, here, the missile was not launched with an intent to use force.

For the relatively few commentators who have turned their attention to the question of intention, they generally coalesce around the second approach: the prohibition on the use of force is breached only where a State harbors an *animus belligerandi* (hostile intent) to use force against the target State.[41] While these commentators cite several factors to support this conclusion, none of them are persuasive. Let us first consider these arguments before I demonstrate why Article 2(4) is—and should be—based on objective responsibility.

First, Article I of the Kellogg-Briand Pact explains that "[t]he High Contracting Parties solemnly declare in the names of their respective peoples that they condemn recourse to war for the solution of international controversies, and renounce it, as an instrument of national policy in their relations

41. "But while there is no express authority or primary rule on the element of *mens rea* in the determination that a prohibited use of force has occurred, it is arguable that an intention to use force is nonetheless required." HENDERSON, *supra* note 39, at 125. Elsewhere, Henderson explains that "an intention to use force against a state, or an *animus belligerandi*, is required in order to breach the prohibition on the threat or use of force." *Id.* "[F]or the article to apply to a particular situation presupposes that a State resorts to force against another, which supposes it intends to force the other State to do or not do something." CORTEN, *supra* note 39, at 85. A hostile intent to use force can be established where a State engages in an operation and it is virtually certain that the forcible effects would occur in the ordinary course of events (oblique intent). Here, intent is present even if the forcible effects are not directly intended.

with one another."[42] Given that the Pact is designed to prohibit "aggressive war," the requirement of *animus belligerandi* is inherent to the prohibition.[43] Commentators claim that, because Article I of the Pact was the forerunner to Article 2(4) of the UN Charter,[44] the former can be used to shed light on the meaning of the latter.[45] Consequently, as with the Kellogg-Briand Pact, a State must manifest an intention to use force against the target State, and go on to use force against that State, in order to breach Article 2(4).

In my view, using Article I of the Pact to interpret the meaning of Article 2(4) of the Charter is problematic because these provisions pursue very different aims and objectives. This is evident from the content of the two prohibitions. While Article I of the Pact prohibits "war . . . as an instrument of national policy," Article 2(4) of the Charter prohibits "force . . . against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purpose of the United Nations." Clearly, Article 2(4) of the Charter is wider in scope than Article I of the Pact, which cautions against invoking Article I of the Pact to help interpret the meaning of Article 2(4) of the Charter.

Second, the principle of non-intervention prohibits coercion within the *domaine réservé* of another State. Some commentators argue that, for a breach of this principle to occur, coercion must be applied intentionally because, as Christian Henderson explains, "an 'unintentional coercion' would seem to be something of a misnomer."[46] In the 1986 *Nicaragua* judgment, the ICJ explained that "[t]he element of coercion, which defines, and indeed forms the very essence of, prohibited intervention, is particularly obvious in the case of an intervention which uses force."[47] Citing this aspect of the judgment, commentators claim that the use of force represents an obvious and

---

42. General Treaty for Renunciation of War as an Instrument of National Policy art. I, Aug. 27, 1928, 46 Stat. 2343, 94 L.N.T.S. 57.

43. The Nuremburg Tribunal explained that "this Pact was violated by Germany in all the cases of aggressive war charged in the Indictment." International Military Tribunal (Nuremburg), Judgment, at 51 (Oct. 1, 1946), https://www.legal-tools.org/doc/45f18e/pdf/.

44. While the Kellogg-Briand Pact is still technically in force, it has been effectively replaced by the UN Charter system.

45. "This would also reflect the notion of 'use of force' as a broader concept but in many ways a continuation of the old concept of 'war' from the preceding treaty, the Kellogg–Briand Pact, which condemns 'recourse to war for the solution of international controversies' and embodies its renunciation 'as an instrument of national policy.'" POBJIE, *supra* note 39, at 154.

46. HENDERSON, *supra* note 39, at 126.

47. Military and Paramilitary Activities in and against *Nicaragua*, *supra* note 16, ¶ 205.

specific breach of the principle of non-intervention; in other words, the prohibition on the use of force is a subspecies of the principle of non-intervention.[48] The argument therefore runs that, if the non-intervention principle requires intention, the use of force prohibition must also require intention.

Historically, the prohibition on the use of force was subsumed within the principle of non-intervention. Over time, however, these provisions gradually de-coupled and assumed their own distinct ontological status under international law.[49] This means that the principles of non-intervention and non-use of force developed their own meaning and content—whereas the principle of non-intervention prohibits coercion, the principle of non-use of force prohibits violence.[50] Indeed, the ICJ's decision in *Nicaragua* does not contradict this position. The better reading of *Nicaragua* is that, while in many circumstances the use of force will involve the application of coercion, this need not always be the case.[51] Consequently, the fact that intention is a critical element of the principle of non-intervention does not automatically and necessarily mean that intention is a critical element of the prohibition on the use of force.

Third, it is clear that a threat of force requires a State to manifest an intention to use force against the target State.[52] This leads some commentators to conclude that, because the threat and use of force prohibitions are contained in the same provision (Article 2(4)) and given that these prohibitions are closely aligned insofar as threats of force are unlawful only if the

48. HENDERSON, *supra* note 39, at 125 ("the use of force is a specific form of intervention").

49. Buchan & Tsagourias, *supra* note 34.

50. "[Intervention has become] a word used to describe the sorts of behaviour not covered by Article 2(4) and hence non-intervention a rule not to be found there." ROBERT J. VINCENT, NON-INTERVENTION AND INTERNATIONAL ORDER 234 (1974).

51. "[I]t is not clear from the judgment whether a use of force must always be coercive. Just as an unlawful intervention can be forcible or non-forcible, it is arguable that a prohibited use of force can violate the principle of non-intervention or not." POBJIE, *supra* note 39, at 154.

52. Legality of the Threat or Use of Nuclear Weapons, *supra* note 22, ¶ 47; *see generally* Marco Roscini, *Threats of Armed Force and Contemporary International Law*, 54 NETHERLANDS INTERNATIONAL LAW REVIEW 229 (2007); Nicholas Tsagourias, *The Prohibition of Threats of Force*, *in* RESEARCH HANDBOOK ON INTERNATIONAL CONFLICT AND SECURITY LAW (Nigel D. White & Christian Henderson eds., 2013).

force would be unlawful if it were actually used,[53] a prohibited use of force must, like a prohibited threat of force, require intention.[54]

While the prohibitions on the threat and use of force are closely aligned and do indeed appear in the same provision of the UN Charter, they are nevertheless distinct and possess their own content and triggers.[55] Sure, these prohibitions share a common feature insofar as both pertain to "force." But what constitutes a "threat" and "use" of force can nevertheless differ. Thus, while the threat of force requires intention, this does ineluctably lead to the conclusion that the use of force also requires intention.

Finally, and perhaps most importantly, these commentators claim that State practice supports the argument that hostile intent forms a critical part of the prohibition on the use of force. The example that is typically given in this context is an incident that occurred during NATO's forcible intervention in the former Yugoslavia when it acted to prevent gross and systematic abuses of human rights.

In May 1999, and due to the pilots using outdated maps, a U.S. military aircraft mistakenly bombed the Chinese Embassy in Belgrade and killed several Chinese nationals and injured many others.[56] Commentators place significant emphasis on the fact that, while China criticized the bombing as "a gross violation of Chinese sovereignty and a random violation of the Vienna Conventions on Diplomatic Relations," it did not specifically claim that the United States had breached the prohibition on the use of force.[57] This has led these commentators to conclude that China did not invoke Article 2(4) because the United States intended to use force against Yugoslavia rather than China.[58] In short, the United States did not harbor hostile intent *towards* China.

---

53. Legality of the Threat or Use of Nuclear Weapons, *supra* note 22 , ¶ 47.

54. "If prohibited threats to use force require a coercive intent and the two prohibitions of threats and use of force are coupled, this would indicate that the latter also requires a coercive intent." POBJIE, *supra* note 39, at 146.

55. NIKOLAS STÜRCHLER, THE THREAT OF FORCE IN INTERNATIONAL LAW 262 (2009).

56. *NATO's Out-of-Date Map Caused Chinese Embassy Blunder*, THE GUARDIAN (May 11, 1999), https://www.theguardian.com/world/1999/may/11/balkans12.

57. U.N. Security Council Press Release SC/6674/Rev.1*, China, at Security Council Meeting, Registers Strongest Possible Protest Over Attack Against Its Embassy in Belgrade (May 8, 1999), https://press.un.org/en/1999/19990508.sc6674.r1.html.

58. HENDERSON, *supra* note 39, at 128.

However, a close inspection of Security Council debates[59] indicates that China (and other States) may have seen the bombing as a breach of Article 2(4). For instance, China explained that the United States had "attacked" its embassy and, "[w]hether deliberate or not, the incident was a blatant transgression of international law, and NATO must take responsibility for its actions."[60] In addition, Russia claimed that "the events were unconscionable and flagrantly violated the United Nations Charter."[61] Similarly, Iraq "condemned the barbaric act, which violated the United Nations Charter, international law, and the laws governing relations between countries."[62] Admittedly, these States did not specifically cite a breach of the prohibition on the use of force. But given that this incident involved one State bombing the embassy of another State ("attacked," as China explained), descriptions of this conduct as a breach of international law and in particular the UN Charter strongly indicate that they were referring to a violation of Article 2(4).

If we accept that these States characterized the embassy bombing as a violation of Article 2(4), there are three possible explanations for this conclusion. First, their view may have been that the United States launched an operation with a *general intent* to use force and it was therefore liable for a breach of Article 2(4) even though it did not harbor hostile intent towards China.

Second, they may have regarded the bombing as *negligent* because the U.S. pilots used an official CIA map that was reviewed and revised in 1997 and 1998 but that failed to identify the correct location of the Chinese Embassy even though it had moved to its new site in 1996.[63] Thus, any reasonable State in the United States' position would have been aware of the new location of the Chinese Embassy, updated its map correctly, and refrained from bombing those coordinates. As already noted, negligence is a type of subjective responsibility. According to this reading, negligence (as distinct from intention) is sufficient to establish a breach of Article 2(4). There is some support for this approach in State practice.

In 1994, Cameroon complained to the ICJ that Nigeria had breached Article 2(4) by sending its armed forces into the Bakassi Peninsula.[64] Given

59. Security Council Press Release SC/6674/Rev.1*, *supra* note 57.

60. *Id.*

61. *Id.*

62. *Id.*

63. *NATO's Out-of-Date Map Caused Chinese Embassy Blunder*, *supra* note 56.

64. Land and Maritime Boundary between Cameroon and Nigeria (Cameroon v. Nigeria), Judgment, 2002 I.C.J. 303, ¶ 310 (Oct. 10).

the unsettled nature of the boundary, Nigeria maintained that it held sovereign title over the peninsula but that, even if it did not, this was the product of an "honest belief" or "reasonable mistake."[65] Nigeria therefore appeared to argue that, even if Cameroon held sovereign title over the territory, it did not breach Article 2(4) because it did not intend to use force against Cameroon (because of its "honest belief" that it held sovereign title over the territory) or, even if it had made a mistake about who owned the territory, it did not use force negligently (because the mistake was "reasonable").

It would be interesting to consider Cameroon's and the ICJ's response to Nigeria's suggestion that Article 2(4) is not breached unless (at a minimum) force is used negligently. However, Cameroon did not engage with this issue because it maintained that Nigeria knew that it did not hold sovereign title over the peninsula. Put differently, it saw Nigeria's actions as intentional and thus the question of mistake (made negligently or otherwise) was not relevant. Moreover, the ICJ did not consider it necessary to determine whether Nigeria's actions breached Article 2(4) because it had already determined that Cameroon possessed sovereignty over the disputed territory and therefore "the injury suffered by Cameroon by reason of the occupation of its territory will in all events have been sufficiently addressed."[66]

Moreover, in 1988 the USS *Vincennes* shot down an Iranian civilian airliner after its air defense system mistakenly identified the aircraft as an incoming Iranian fighter jet. Iran rejected the U.S.'s claim that the incident was a mistake and maintained that the United States had intentionally shot down the civilian airliner. Yet, it proceeded to explain that, "[e]ven if there was a mistaken identification, this amounted to such gross negligence and recklessness on the part of the Vincennes that any characterization of the act as accidental or excusable is plainly wrong"[67] and "would not have made the act of shooting it down any less unlawful."[68] Iran's statement therefore suggests that a breach of Article 2(4) occurs where a use of force is committed negligently.

Third, it may be the case that China, Iraq, and Russia held the United States liable for a breach of Article 2(4) when it bombed the Chinese Embassy on the basis of objective responsibility. On this reading, the United

---

65. *Id.* ¶ 311.

66. *Id.* ¶ 319.

67. Memorial of Iran, Case Concerning the Aerial Incident of 3 July 1988 (Iran v. U.S.), ¶ 4.53, at 243–44 (July 24, 1990), https://icj-cij.org/sites/default/files/case-related/79/6629.pdf.

68. *Id.* ¶ 4.54, at 244.

States breached Article 2(4) because it engaged in conduct that resulted in the use of force against China regardless of whether it had acted intentionally or negligently.

For me, the third reading is most convincing because, when condemning the U.S.'s conduct as internationally wrongful, these States did not focus on whether the attack was committed intentionally or negligently. In fact, as seen, China criticized the U.S.'s conduct regardless of "[w]hether [it was] deliberate or not" and implored the United States to "take responsibility for its actions." Rather, these States appeared to condemn the United States solely on the basis that it had engaged in conduct that led to the use of force against China (i.e., objective responsibility).

Significantly, recent State practice indicates that Article 2(4) is based on objective rather than subjective responsibility. The potential for cyber operations to generate unintended reverberating effects involving the use of force has been a key driver of this practice. As we know, States and international organizations have published a flurry of statements on the application of international law to cyberspace in recent years.[69] These statements invariably contain a section on the use of force prohibition but, importantly, to date, *none* of them require that the forcible effects of a cyber operation must be committed intentionally or negligently in order for a breach of Article 2(4) to occur.[70]

In fact, Australia and New Zealand expressly state that the "intended or reasonably expected" effects of a cyber operation can be considered when determining whether a breach of Article 2(4) has been committed.[71] Similarly, for the African Union a cyber operation amounts to a use of force where the effects triggering the application of Article 2(4) were "expected"

---

69. For a repository of these statements, *see Use of Force*, INTERNATIONAL CYBER LAW IN PRACTICE: INTERACTIVE TOOLKIT, https://cyberlaw.ccdcoe.org/wiki/Use_of_force (last visited Oct. 28, 2025).

70. The United States has explained that, "[i]n assessing whether an event constituted a use of force in or through cyberspace, we must evaluate factors: including the context of the event, the actor perpetrating the action (recognizing challenging issues of attribution in cyberspace), the target and location, effects and intent, among other possible issues." Harold Hongjiu Koh, Legal Advisor, U.S. State Dep't, Remarks at the USCYBERCOM Inter-Agency Legal Conference: International Law in Cyberspace (Sept. 18, 2012), https://2009-2017.state.gov/s/l/releases/remarks/197924.htm. While this statement explains that intent can be a factor indicating that a cyber operation amounts to a use of force, it is clear that intent is not identified as a *precondition* for a breach of the use of force prohibition.

71. 2021 Official Compendium, *supra* note 26, at 5 (Australia); New Zealand, *supra* note 24, ¶ 7.

when the operation was launched.[72] As Australia and New Zealand make clear, whether the forcible effects were reasonably expected has nothing to do with intention ("intended *or* reasonably expected").[73] Moreover, the test of reasonable expectation has nothing to do with negligence because it does not require an assessment of whether the State failed to take reasonable measures to prevent the harm. Instead, reasonable expectation asks whether a reasonable State would have foreseen the harm in the circumstances and, as we shall see, expectation/foreseeability raises the question of causation rather than fault.

In sum, the trajectory of State practice suggests that Article 2(4) is premised on objective responsibility.[74] This assessment stands even though, when force is used as the result of a genuine mistake, the parties may prefer to resolve the dispute diplomatically rather than through formal legal processes.[75]

Furthermore, in my view there are very good reasons for why Article 2(4) should be based on objective responsibility. This is because integrating a requirement of intention or negligence into Article 2(4) can create gaps or loopholes in the prohibition.[76] This is particularly the case when it comes to AI-enabled systems given their potential to carry out unintended engagements involving the use of force. For example, a requirement of intention or negligence may mean that a State is not liable where its AI-enabled system mistakenly uses force against targets located in third States; where a warship's AI-enabled air defense system shoots down an aircraft belonging to a State after mistakenly believing that it belongs to another actor; where an AI-enabled system is designed to perform passive cyber defense within a State's own networks but it malfunctions and launches an attack against systems located on the cyber infrastructure of another State; where an AI-enabled

72. Common African Position, *supra* note 32, ¶ 39.

73. 2021 Official Compendium, *supra* note 26, at 5 (Australia); New Zealand, *supra* note 24, ¶ 7 (emphasis added).

74. An interpretation of Article 2(4) that requires subjective responsibility is also difficult to reconcile with the *travaux* of the UN Charter. As I have previously noted, the *travaux* indicate that the framers of the UN Charter intended to create an "absolute all-inclusive prohibition" on the use of force. 6 DOCUMENTS OF THE UNITED NATIONS CONFERENCE 335 (1945). "[N]o express authority supports the view that intent is needed to establish a violation of UN Charter Article 2(4)." Ruys, *supra* note 28, at 191.

75. BUCHAN & TSAGOURIAS, *supra* note 19, at 33.

76. IAN BROWNLIE, INTERNATIONAL LAW AND THE USE OF FORCE BY STATES 377 (1963).

combat drone experiences severe technical difficulties and, for safety reasons, the operator decides to jettison its weapons over a remote mountain range but they nevertheless hit a village on the ground; and where an AI-enabled reconnaissance drone is deployed into the national airspace of another State but, due to a technical malfunction, it crashes into the State's territory and causes extensive harm to people or property.

Provided causation is established (causation will be discussed in Part IV), there is no good reason why the deploying State should avoid responsibility for a breach of Article 2(4) simply because it did not intentionally or negligently use force against the victim State or, more to the point, there is no good reason why the victim State should not be entitled to the protection afforded by Article 2(4) simply because the force was not used intentionally or negligently. Indeed, the international community as a whole has an interest in categorizing this conduct as a breach of Article 2(4) given that the prohibition on the use of force is an *erga omnes* obligation aimed at the maintenance of international peace and security.

It should also be noted that a requirement of intention or negligence puts victim States at a disadvantage because of the difficulties in proving that entities such as States have acted intentionally or negligently. As Ian Brownlie observed, demonstrating "*animus aggressionis* in respect of a state is something of a chimera."[77] The same can be said with regard to negligence.

IV. CAUSATION

Some readers may feel that the objective approach sets the responsibility bar too low. In my view, these concerns are sufficiently assuaged by the requirement of causation.[78] Causation operates as an international legal tool that assigns responsibility to a State for those effects for which it is blameworthy. Causation thus rejects the assumption that a State is automatically responsible for all the harmful effects that follow from its conduct: "not all events which follow each other in invariable sequence are causally related."[79]

77. *Id.*

78. Note that causation is required regardless of whether Article 2(4) is based on subjective or objective responsibility. MALCOLM N. SHAW, INTERNATIONAL LAW 594 n.41 (2018) ("The question of intention is to be distinguished from the problem of causality, i.e., whether the act or omission actually caused the particular loss or damage").

79. HERBERT L. A. HART & TONY HONORÉ, CAUSATION IN THE LAW 15 (1985).

Causation must be distinguished from attribution.[80] Attribution is a secondary rule of international law and fundamentally a "normative operation":[81] Is the impugned conduct legally attributable to a State? As we know, where a de jure or de facto organ of a State engages in conduct (e.g., where the armed forces or an intelligence agency deploys an AI-enabled system),[82] that conduct is automatically attributable to the State even where the organ acts ultra vires.[83] As we shall see, causation (where required) forms part of the primary rule in question and determines the range of effects for which the State can be held responsible when its organs act. Following the structure of Article 2 of the Articles on State Responsibility,[84] determining attribution and causation is a two-step process:

> The first step is to establish whether and how the State is involved, and then establish what its involvement has caused. For, "the relationship between the person of the State and the action of an individual" is only a question as to the status of the person or entity acting in a particular way, and does not trouble the causation enquiry at all, because causation would enter the scene only after the involvement of State organs is identified. It is not just about *who* has acted but also about *what exactly* they have done, how their action altered status quo and caused the prohibited harm.[85]

---

80. Roberto Ago, Special Rapporteur, Second Report on State Responsibility, ¶ 38, U.N. Doc. A/CN.4/233 (Apr. 20, 1970) (attribution is "a legal connecting operation which has nothing in common with a link of natural causality").

81. *Articles on State Responsibility*, *supra* note 38, art. 2 cmt. ¶ 6.

82. *Id.* arts. 4, 5.

83. *Id.* art. 7.

84. Article 2 of the *Articles on State Responsibility*, *supra* note 38, explains: "There is an internationally wrongful act of a State when conduct consisting of an action or omission: (a) is attributable to the State under international law; and (b) constitutes a breach of an international obligation of the State."

85. ALEXANDER ORAKHELASHVILI, CAUSATION IN INTERNATIONAL LAW 59 (2022) (citing *Articles on State Responsibility*, *supra* note 38, art. 3 cmt. ¶ 6). "The relationship between causation and attribution is clear in the sense that causation serves as a limit on the scope of state responsibility for attributable acts." David Caron, *Attribution Amidst Revolution: The Experience of the Iran-United States Claims Tribunal*, 84 ASIL PROCEEDINGS 64, 68 (1990); *see also* Ago, *supra* note 80, ¶ 38; Vladyslav Lanovoy, *Causation in the Law of State Responsibility*, BRITISH YEARBOOK OF INTERNATIONAL LAW 1, 20 (2022), https://academic.oup.com/bybil/advance-article-abstract/doi/10.1093/bybil/brab008/6516063.

A. The Prohibition on the Use of Force and the Requirement of Causation

As indicated, international law leaves the question of causation to the primary rules of international law.[86] The immediate issue, therefore, is whether causation must be established in order for Article 2(4) to apply. While some primary rules of international law specifically require causation,[87] this is not the case with Article 2(4).[88] However, I have already explained that Article 2(4) is an effects-based prohibition; thus, it is the forcible effects generated by State activity that trigger a breach of the prohibition. As an effects-based prohibition, interpretive logic indicates that a State must cause the forcible effects in order for a breach of Article 2(4) to occur.[89]

More importantly, States seem to have worked a requirement of causation into Article 2(4). This practice has emerged in recent years given the potential for cyber operations to generate reverberating effects—the typical example being where a State commits a ransomware attack against a health care provider with the aim of extorting money but, by locking the provider out of its systems, patient care is affected leading to serious harm and even death. The question, then, is whether the cyber operation caused the ensuing harm to people and thus whether that harm can be considered when determining whether a breach of Article 2(4) has transpired.[90]

The *Tallinn Manual 2.0* explains that "directness" between a cyber operation and the resulting use of force is an important factor when determining whether a breach of Article 2(4) has emerged. The *Manual* goes on to explain that "directness examines the chain of causation. . . . Cyber operations in

86. Ilias Plakokefalos, *Causation in the Law of State Responsibility and the Problem of Overdetermination: In Search of Clarity*, 26 EUROPEAN JOURNAL OF INTERNATIONAL LAW 471, 481 (2015); Lanovoy, *supra* note 85, at 17, 21–22.

87. *See, e.g.*, United Nations Convention on the Law of the Sea art. 139(2), Dec. 10, 1982, 1833 U.N.T.S. 397.

88. Orakhelashvili notes that although it is not unusual for a primary rule to fail to specify that causation is required, State practice often makes it clear that causation is a condition precedent for establishing a breach of the rule. ORAKHELASHVILI, *supra* note 85, at 56 ("primary norms merely specify [the] standard of lawful conduct of a State, not the range of means and ways in which a primary norm is violated by a State").

89. PRIYA URS ET AL., *supra* note 18, at 47–48.

90. "When it comes to other forms of conduct that might amount to a use of force, however, such as the use of cyber operations, a suitable standard of causation is needed in the application of Article 2(4)." *Id.* at 52.

which cause and effect are clearly linked are more likely to be characterised as uses of force than those in which they are highly attenuated."[91]

As we know, the *Tallinn Manual* project has been very influential on States and international organizations when developing their positions on the application of international law to cyberspace. Indeed, certain States seem to adopt the *Tallinn Manual's* approach when they explicitly reference directness. For example, Norway explains that "directness" between a cyber operation and the resulting use of force is an important indicator of a breach of Article 2(4).[92] Other States go further and specifically require causation, even if they use different language. Austria, for instance, explains that "cyber activity that *leads to* injury, death or significant physical damage constitutes an unlawful use of force."[93] Denmark,[94] Estonia,[95] and the Czech Republic[96] explain that a cyber operation must "result in" the use of force in order for Article 2(4) to be engaged. Furthermore, the United States, which has produced a series of statements on the application of the international law to cyberspace, has explained that the cyber operation must "proximately result in" the use of force (in 2012[97] and again in 2021[98]) and that Article 2(4) applies where a cyber operation "causes" forcible effects (in 2020).[99] As we have already seen, Australia and New Zealand explain that a State is responsible for a breach of Article 2(4) where the forcible effects of a cyber operation are "reasonably expected," and the African Union uses similar language when it explains that Article 2(4) covers "expected" forcible effects. As we

---

91. TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS 334 (Michael N. Schmitt gen. ed., 2017).

92. Norway, *Norway's Position Paper on International Law and Cyberspace*, 92 NORDIC JOURNAL OF INTERNATIONAL LAW 470, 480 (2023)) [hereinafter Norway Position].

93. Austria, Position Paper of the Republic of Austria: Cyber Activities and International Law 6 (Apr. 2024), https://docs-library.unoda.org/Open-Ended_Working_Group_on_Information_and_Communication_Technologies_-_(2021)/Austrian_Position_Paper_-_Cyber_Activities_and_International_Law_(Final_23.04.2024).pdf (emphasis added).

94. Denmark Position, *supra* note 26, at 451.

95. 2021 Official Compendium, *supra* note 26, at 26 (Estonia).

96. Czech Republic, Position Paper on the Application of International Law in Cyberspace, ¶ 26 (Feb. 2024), https://mzv.gov.cz/file/5376858/_20240226___CZ_Position_paper_on_the_application_of_IL_cyberspace.pdf.

97. Koh, *supra* note 70.

98. 2021 Official Compendium, *supra* note 26, at 137 (U.S.).

99. Paul C. Ney Jr., General Counsel, U.S. Dep't of Defense, Remarks at US Cyber Command Legal Conference (Mar. 2, 2020), https://www.defense.gov/News/Speeches/Speech/Article/2099378/dod-general-counsel-remarks-at-us-cyber-command-legal-conference/.

shall see below, holding States responsible for expected harm is a clear reference to causation, to wit, legal causation. In light of this practice, it seems relatively well-settled that States have integrated a requirement of causation into the prohibition on the use of force.[100]

B. The Prohibition on the Use of Force and the Standard of Causation

The next task is to identify the standard of causation required by Article 2(4). Given that the text of Article 2(4) does not expressly require causation in order to find a breach of the prohibition on the use of force, it follows that Article 2(4) does not give any indication as to what standard of causation must be used when determining when a State can be said to have caused the forcible effects.[101]

Importantly, causation has been extensively considered by international courts and tribunals, as well as the International Law Commission in its Articles on State Responsibility, when determining the extent to which reparations are owed because wrongdoing States must provide reparations to injured States for the "injury caused" by their internationally wrongful acts.[102] This jurisprudence seeks to identify the damage for which reparations are owed under the secondary rules on State responsibility. However, its assessment of the concept of causation can be nevertheless used analogically to help inform our understanding of what causation may mean when it comes to determining whether a breach of a primary rule has occurred in the event that the rule requires causation, but it is silent as to what standard must be used.[103]

---

100. "[T]he notion of 'force' implies the need for some kind of effect in close relationship to a cause." Lahmann, *supra* note 18, at 425.

101. "Article 2(4) of the UN Charter does not specify any standard of causation with which to identify the legally relevant effects of an alleged use of force." PRIYA URS ET AL., *supra* note 18, at 52.

102. *Articles on State Responsibility*, *supra* note 38, art. 31, explains: "The responsible State is under an obligation to make full reparation for the injury caused by the internationally wrongful act" and that "[i]njury includes any damage . . . caused by the internationally wrongful act of a State."

103. Priya Urs, *The Causal Question in the Application of the Law on the Use of Force to Cyber Operations*, NATIONAL UNIVERSITY OF SINGAPORE: CIL DIALOGUES (Apr. 25, 2023), https://cil.nus.edu.sg/blogs/the-causal-question-in-the-application-of-the-law-on-the-use-of-force-to-cyber-operations/.

The International Law Commission explains that "the requirement of a causal link is not necessarily the same in relation to every breach of an international obligation."[104] Thus, when determining reparations international courts and tribunals have adopted different standards of causation depending on the primary rule in question and the nature and extent of the injury.[105] When determining whether a breach of a primary rule has emerged, this means that a standard of causation must be adopted that best meets the aims and objectives of that rule.[106] It should be therefore borne in mind that the aims and objectives of Article 2(4) are to protect the territorial integrity and political independence of States, and to maintain international peace and security more generally, by prohibiting the use of force in international relations.

Conceptually, causation contains two distinct cumulative elements: the State must be a *factual* and *legal* cause of the resulting harm.[107]

1. Factual Causation

Factual causation requires the production of empirical, scientific, or statistical evidence to demonstrate that there is a sufficient causal link between the

---

104. *Articles on State Responsibility*, *supra* note 38, art. 31 cmt. ¶ 10; *see also* Armed Activities on the Territory of the Congo (Dem. Rep. Congo v. Uganda), Judgment on Reparations, 2022 I.C.J. 13, ¶ 93 (Feb. 9) ("it should be noted that the causal nexus required may vary depending on the primary rule violated and the nature and extent of the injury").

105. Lanovoy, *supra* note 85, 43–60.

106. PRIYA URS ET AL., *supra* note 18, at 54; Urs, *supra* note 103.

107. Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosn. & Herz. v. Serb. & Montenegro), Judgment, 2007 I.C.J. 43, ¶ 462 (Feb. 26) (where the Court required a "sufficiently direct and certain causal nexus" to establish causation, and where the "certain causal nexus" aspect of this test speaks to factual causation and the "sufficiently direct" aspect speaks to legal causation).

> Causation has two aspects. The first, already alluded to, is factual causation. This requires a determination of whether the state's wrongful act or omission constituted a necessary link in the chain of circumstances leading to the claimant's injuries. The second element is legal or proximate causation, which involves analysis of whether the claimant's injury was a foreseeable consequence of the state's act or omission.

David J. Bederman, *Contributory Fault and State Responsibility*, 30 VIRGINIA JOURNAL OF INTERNATIONAL LAW 335, 349 (1990). The ILC's commentary on the Articles on State Responsibility notes that "causality in fact *is a necessary but not a sufficient* condition for reparation. There is a further element, associated with the exclusion of injury that is too 'remote' or 'consequential' to be the subject of reparation." *Articles on State Responsibility*, *supra* note 38, art. 31 cmt. ¶ 10 (emphasis added).

conduct and the effects giving rise to a breach of the international legal rule. The dominant approach to determining factual causation among international courts and tribunals is the so-called "but for" test.[108] The "but for" standard establishes factual causation where the harm would not have occurred *but for* the State's conduct. In this way, the State's conduct must be a necessary condition for the harm to occur. This means that factual causation is established where State A deploys an AI-enabled surveillance drone into the national airspace of State B but, due to adverse weather conditions, it crashes into State B's territory and causes harm to people or property. Here, factual causation is present because the resulting forcible effects would not have occurred but for State A's decision to deploy the drone. Similarly, factual causation is established where State A deploys an AI-enabled combat drone into State B in order to strike a target but, due to faulty facial recognition technology, it hits the wrong target. Again, the explanation lies in the fact that the forcible effects would not have occurred but for State A's decision to deploy the drone. Equally, factual causation is established where State A deploys an AI-enabled drone into the territory of State B but it launches an attack against the wrong target due to an intervention by a malicious third-party, such as where an adversary commits an evasion attack. As with before, factual causation is present because the forcible effects would not have occurred but for State A's decision to deploy the drone. That said, in all these scenarios legal causation looms large, and will be returned to below.

Incidentally, where a *State*[109] uses adversarial tactics against an AI-enabled system—such as an evasion attack that causes the system to misidentify a target or a jamming operation against an AI-enabled drone that causes it to crash—it can be also regarded as the factual cause of the harm (i.e., the

108. *See, e.g.*, LG&E Energy Corp. v. Argentine Republic, ICSID Case No. ARB/02/1, Award, ¶ 48 (July 25, 2007); Ioan Micula v. Romania [I], ICSID Case No. ARB/05/20, Award, ¶ 1117 (Dec. 11, 2013); Chevron Corporation (USA) v. Ecuador, Case No. 34877, Partial Award on the Merits, ¶ 374 (Perm. Ct. Arb. 2010); Bilcon of Delaware v. Canada, Case No. 2009-04, Award on Damages, ¶ 94 (Perm. Ct. Arb. 2019); Suez, Sociedad General de Aguas de Barcelona, S.A. v. Argentina, ICSID Case No. ARB/03/19, Award, ¶ 53 (Apr. 9, 2015) (although the tribunal implicitly adopted the "but for" test).

109. Whether *non-State actors* are bound by the customary prohibition on the use of force is contested and falls beyond the scope of this article. On this debate, *see* BUCHAN & TSAGOURIAS, *supra* note 19, at 18–19. That parties to armed conflicts can be held responsible for breaches of international humanitarian law where they use adversarial tactics against AI-enabled systems operated by other actors, *see* Jonathan Kwik, *Is Wearing These Sunglasses an Attack? Obligations Under IHL Related to Anti-AI Countermeasures*, 926 INTERNATIONAL REVIEW OF THE RED CROSS 732 (2024).

force) because, but for its intervention, the harm would not have occurred. Conceptually, under the "but for" test multiple actors can be designated as the factual cause of the resulting harm provided their actions together are necessary for the harm to occur; that is, but for each of their actions, the harm would not have occurred.

The situation is different, however, where each action *alone* would be sufficient to bring about the harm. For example, factual causation cannot be established where State A launches an AI-enabled cyber operation against State B in order to disable its military communications systems but it is revealed that a number of other States also launched similar operations against those systems. In this scenario, State A would not be responsible for a use of force because the harm (i.e., the forcible effects) would have occurred but for its conduct.

This scenario reveals the limits of the "but for" test; where there are *multiple independent sufficient causes* of the resulting harm, the "but for" test is not met.[110] It is for this reason that certain commentators draw on developments in national legal systems and propose alternative tests for factual causation. The most popular in this regard is the so-called "necessary element of a sufficient set" test: factual causation is established where the impugned conduct "was a necessary element of a set of antecedent actual conditions that was sufficient for the occurrence of the result."[111] This test may be alien to international lawyers, however, because there is little indication that international courts and tribunals (let alone States) have adopted it when determining factual causation.

2. Legal Causation

If factual causation is established, the next step is to consider legal causation. Legal causation asks whether there are any normative, policy, or pragmatic factors that justify severing the factual chain of causation at any intermediate point between the conduct and the effects because the resulting damage is "too indirect, remote, and uncertain to be appraised."[112] In other words, legal

110. Plakokefalos, *supra* note 86, at 477.

111. *Id.* at 478 (citing HART & HONORÉ, *supra* note 79, at 110–29; Richard W. Wright, *Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts*, 73 IOWA LAW REVIEW 1001, 1019 (1987–1988)).

112. Trail Smelter (U.S. v. Can.), 3 R.I.A.A. 1905, 1931 (Trail Smelter Arb. Trib. 1938 & 1941).

causation seeks to delimit the scope of State responsibility to that which is fair and just in the circumstances.

Despite international courts and tribunals routinely recognizing that causation contains two discrete elements (factual and legal causation), with its emphasis on fairness and justice the reality is that the requirement of legal causation effectively consumes the requirement of factual causation and it is for this reason that legal causation is often the main focus of debates.[113] Indeed, and as we shall see, this has certainly been the approach of States when determining whether a cyber operation can be said to have caused the resulting use of force: factual causation is basically assumed and the focus quickly shifts to legal causation.[114]

International courts and tribunals have articulated different standards for determining legal causation.[115] In the *Bosnian Genocide* case, the ICJ concluded that Serbia had failed to comply with its obligations under the Genocide Convention to prevent and punish acts of genocide and then had to identify the damage or harm caused to Bosnia and Herzegovina for which reparations were owed. In addressing the question of legal causation, the ICJ explained that reparations are owed where there is a "sufficiently direct and certain causal nexus" between the wrongful conduct and the injury suffered.[116] As we have seen in the cyber context, States such as Norway have used the concept of "directness" to determine whether the causal chain can

---

113. As the tribunal in *Burlington* explained, "if the injury was objectively foreseeable (i.e., because the act was objectively capable of causing the injury), *then the test for both factual and legal causation will normally be met*." Burlington Resources Inc. v. Ecuador, ICSID Case No. ARB/08/5, Decision on Reconsideration and Award, ¶ 333 (Feb. 7, 2017) (emphasis added). "The standard of legal causation may even do away entirely with any requirement of factual causation." Urs, *supra* note 103.

114. See, for example, the cyber statements by Australia, *supra* note 32; New Zealand, *supra* note 24; Common African Position, *supra* note 32.

115. "International courts and tribunals have adopted a variety of views on the standard of legal causation." Lanovoy, *supra* note 85, at 14.

116. Application of the Convention on the Prevention and Punishment of the Crime of Genocide, *supra* note 107, ¶ 462. For further support for the test of directness in the ICJ's jurisprudence, *see* Ahmadou Sadio Diallo (Guinea v. Dem. Rep. Congo), Compensation Judgment, 2012 I.C.J. 324, ¶ 14 (June 19); Certain Activities Carried Out by Nicaragua in the Border Area (Costa Rica v. Nicar.), Compensation Judgment, 2018 I.C.J. 15, ¶ 32 (Feb. 2). The International Tribunal for the Law of the Sea has also adopted the test of directness. *See, e.g.*, M/V Saiga (No. 2) (St. Vincent v. Guinea), Case No. 2, Judgment of July 1, 1999, ITLOS Rep. 1999, at 10, ¶ 172; M/V Virginia G (Pan./Guinea-Bissau), Case No. 19, Judgment of Apr. 14, 2014, ITLOS Rep. 2014, at 4, ¶ 436; M/V Norstar (Pan. v. Italy), Case No. 25, Judgment of Apr. 10, 2019, ITLOS Rep. 2018–2019, at 10, ¶ 334.

be established.[117] The advantage of this test lies in its simplicity insofar as it covers effects that are brought about "in one causal step or moment."[118] The drawback, however, is that it excludes indirect effects and is thus underinclusive.[119] Evidently, this problem is particularly acute when it comes to new and emerging technologies given their potential for reverberating (i.e., indirect) effects.[120]

Perhaps for this reason, other courts have favored the "proximate"[121] or "sufficiently proximate"[122] standard, which is a "somewhat less restrictive alternative" to the test of directness.[123] This test asks whether the harmful effects are sufficiently proximate to the impugned conduct to warrant the imposition of legal responsibility. When it comes to determining whether a breach of Article 2(4) has occurred, we find support for this approach in the cyber context where the United States has consistently explained that a State is responsible for a breach of Article 2(4) where its cyber operations "proximately result in" the use of force.[124]

That said, the proximity test is not problem-free. The main difficulty is that it is vague and imprecise and does not provide a sufficiently clear test to determine which effects can be considered when assessing whether a State has committed a breach of the prohibition on the use of force: "The standard of proximity admits of varied application, permitting the drawing of what

117. Norway Position, *supra* note 92, at 480. Kreß explains that the use of force must be "sufficiently direct" in order for a breach of Article 2(4) to arise. Claus Kreß, *The State Conduct Element*, *in* THE CRIME OF AGGRESSION: A COMMENTARY 412, 425 (Claus Kreß & Stefan Barriga eds., 2017).

118. Lanovoy, *supra* note 85, at 53. Lanovoy goes on to explain that "[t]he adoption of such standard may substantially simplify the adjudicator's task." *Id.*

119. *See* War-Risk Insurance Premium Claims Arbitration (U.S. v. Ger.), 7 R.I.A.A. 44, 62–63 (1923) (where the umpire explained that the distinction between direct and indirect causes of damage is "illusory and fanciful" and "should have no place in international law").

120. PRIYA URS ET AL., *supra* note 18, at 56–57.

121. LG&E Energy Corp. v. Argentine Republic, *supra* note 108, ¶ 50; S.D. Myers, Inc. v. Canada, Second Partial Award, ¶ 140 (NAFTA Arb. Trib., Oct. 21, 2002). Certain commentators have also favored this approach with regard to Article 2(4). *See* POBJIE, *supra* note 39, at 134 ("the use of force should be the proximate cause of harm").

122. Victor Pey Casado and President Allende Foundation v. Chile (I), ICSID Case No. ARB/98/2, Award II, ¶ 218 (Sept. 13, 2016).

123. PRIYA URS ET AL., *supra* note 18, at 57.

124. 2021 Official Compendium, *supra* note 26, at 137 (U.S.); Koh, *supra* note 70.

are ultimately arbitrary distinctions between proximate and remote causes."[125]

Another standard used by international courts and tribunals to assess legal causation is reasonable foreseeability.[126] Importantly, and as we have seen, this standard has been endorsed by Australia, New Zealand, and the African Union when determining whether a State can be held responsible for the forcible effects of a cyber operation.[127] This standard asks whether the force was objectively foreseeable in the circumstances, that is, would a reasonable State in the impugned State's position have foreseen that the forcible effects would have occurred in the ordinary course of events?[128] The objective nature of this test means that a State must be judged according to

---

125. Urs, *supra* note 103. Honoré refers to proximity as a "rough and ready" standard. Anthony M. Honoré, *Theories of Causation and Remoteness of Damage*, *in* 11 INTERNATIONAL ENCYCLOPEDIA OF COMPARATIVE LAW ch. 7, ¶ 76 (André Tunc ed., 1971).

126. Naulilaa Arbitration (Port. v. Ger.), 2 R.I.A.A. 1011, 1013 (1928) ("The uprising . . . thus constitutes an injury which the author of the initial act . . . should have foreseen as a necessary consequence of its military operation"); Lighthouses Arbitration (Greece v. Fr.), 12 R.I.A.A. 155, 218 (1956) ("The damage was neither a foreseeable nor a normal consequence of the evacuation"); CME Czech Republic B.V. v. Czech Republic, Partial Award, 9 ICSID Rep. 113, ¶ 527 (Sept. 13, 2001) (referring to the "foreseeable consequences" of the conduct); Iran v. U.S., Partial Award, Award No. 604-A15 (II:A)/A26 (IV)/B43-FT, ¶ 2088 (Mar. 10, 2020) ("Such delay [in shipment], and possible damages, were or should have been foreseeable by the United States"); Amco Asia Corp. v. Indonesia, ICSID Case No. ARB/81/1, Award in Resubmitted Proceeding, ¶ 172 (May 31, 1990) (requiring the injury to be "foreseeable"); Burlington Resources Inc. v. Ecuador, *supra* note 113, ¶ 333 (requiring the injury to be "objectively foreseeable"). Note that the Eritrea-Ethiopia Claims Commission saw foreseeability as an essential element of the proximity test rather than a standalone test of legal causation. Eritrea-Ethiopia Claims Commission, Case No. 2001-02, Decision No. 7: Guidance Regarding Jus ad Bellum Liability, ¶ 13 (Perm. Ct. Arb., July 27, 2007).

127. In the context of self-defense and when determining whether the armed attack threshold has been met, the *Tallinn Manual 2.0* explains that States are responsible for all "reasonably foreseeable" effects of their cyber operations. TALLINN MANUAL 2.0, *supra* note 91, at 343.

128.

> If an injury was not objectively foreseeable because it was caused by an unusual chain of events that could not foreseeably derive from the act, legal causation may be absent and recovery may be excluded. However, if the injury was objectively foreseeable (i.e., because the act was objectively capable of causing the injury), then the test for both factual and legal causation will normally be met.

Burlington Resources Inc. v. Ecuador, *supra* note 113, ¶ 333; *see also* BING CHENG, GENERAL PRINCIPLES OF LAW AS APPLIED BY INTERNATIONAL COURTS AND TRIBUNALS 250–51 (1953); Lanovoy, *supra* note 85, at 63.

what a reasonable State would have known—and thus what a reasonable State would have foreseen—in the circumstances.[129] Unlike the negligence test discussed above, reasonable foreseeability does not exclude responsibility where the State has done everything reasonable to prevent the occurrence of the harm; rather, a State is responsible if a reasonable State in its position would have realized that those effects would have occurred. However, the test of reasonable expectation means that a State cannot be held responsible for a breach of Article 2(4) for unforeseeable accidents involving the use of force.[130]

I am in favor of the "reasonably expected" test because, first, it allows States to assess the legality of their operations *before* they are undertaken and, second, it enables actors (such as States, international courts, etc.) to assess the legality of operations *after* they have been carried out with greater consistency, predictability, and fairness.[131] In other words, the reasonable expectation standard strikes an appropriate balance between a strict liability test where the State is held responsible for all the harmful effects that its operation causes in a factual sense (and is thus *over*inclusive) and narrow and arbitrary standards such as directness and proximity that may exclude liability for harmful effects even though those effects were reasonably foreseeable in the circumstances (and are therefore *under*inclusive).[132] In this way, the reasonable expectation standard meets the aims and objectives of Article 2(4) because it holds a State responsible for a breach of the prohibition on the use of force (and provides a victim State with redress) where it pushes forward

---

129. PRIYA URS ET AL., *supra* note 18, at 61.

> The foreseeability analysis would not require inquiring into the actual knowledge or intention of the organ or agent of the state at the time of the breach. Rather, it would be sufficient to inquire whether the organ or agent of the state could or should have envisaged that certain consequences would flow from their actions or omissions in the normal course of events.

Lanovoy, *supra* note 85, at 64.

130. PRIYA URS ET AL., *supra* note 18, at 59 ("This excludes unforeseeable accidents having relevant effects").

131. *Id.* at 60; Urs, *supra* note 103; Lanovoy, *supra* note 85, at 62–63.

132. Some commentators claim that the reasonable expectation standard is overinclusive because potentially everything can be foreseen and, consequently, this test can lead to "infinite liability." Arthur W. Rovine & Grant Hanessian, *Towards a Foreseeability Approach to Causation Questions at the United Nations Compensation Commission*, *in* THE UNITED NATIONS COMPENSATION COMMISSION 244 (Richard B. Lillich ed., 1995). For me, this characterization of the reasonable expectation test fails to appreciate its objective nature, which adequately delimits its scope of application.

with an operation even though it should have been reasonably aware that forcible effects would occur.

What are the implications of the reasonable expectation standard for the use of AI-enabled systems? The first issue is that if States deploy these systems without adequately training and testing them, and therefore do so without being able to accurately predict their behavior, almost all resulting effects can be considered reasonably foreseeable. For example, if a State deploys AI-enabled software into another State's cyber infrastructure without being reasonably confident as to what types of targets it can engage, or where a State is not reasonably confident as to the software's scope of operations and whether it can spread to different networks and systems, the State can be held liable for a breach of Article 2(4) where the software conducts unintended attacks against networks, systems, or data in the target State or spreads across the Internet and carries out unintended attacks against networks, systems, or data in third States.

Consider also the situation where a State deploys a surveillance drone without being reasonably confident that it will remain airborne in adverse weather conditions such as heavy rain and strong winds. If State A deploys the drone without considering the forecasted weather conditions and, as forecast, the drone encounters heavy rain and strong winds, crashes, and causes harm to people or property in State B, State A will be responsible for the forcible effects because any reasonable State in its position would have been aware of the likelihood of bad weather and thus the potential for the drone to crash. The conclusion would be different, however, if the weather forecast was good but the drone encountered freakishly bad weather and, as a result, crashed into the territory of State B. Here, State A will not be responsible for a breach of Article 2(4) because the potential for the drone to crash as a result of bad weather was not reasonably foreseeable.

What about the situation where a malicious actor hacks into the systems that a State uses to train AI-enabled systems and manipulates the training data, leading to AI-enabled systems conducting unintended uses of force once deployed? What about the case where an AI-enabled system is hacked during deployment and reprogrammed to attack civilians? In these scenarios, does the intervention by the third-party break the chain of causation, or can the deploying State still be regarded as the legal cause of the forcible effects? Again, for me, the critical issue is whether the third-party intervention was reasonably foreseeable in the circumstances. If a State was reasonably aware that the system could be hacked and manipulated during development and deployment, the potential for intervention was reasonably foreseeable and

the State is liable for any resulting use of force. By contrast, if a State was reasonably confident that the training processes were rigorous and effective, and that the AI-enabled system was safe and secure once deployed, the third-party intervention was not reasonably foreseeable and thus the State is not liable for a breach of Article 2(4). In short, the third-party intervention amounts to a new and independent cause of the resulting harm—a so-called *novus actus interveniens*—and the deploying State cannot be held responsible for a breach of Article 2(4).

The same rationale applies where an AI-enabled reconnaissance drone is deployed into the airspace of another State but, due to a jamming operation by a third-party, it crashes and causes extensive harm to people and property. If the deploying State should have reasonably known that the drone was susceptible to a jamming operation because of inadequate training and testing and that such an interference could cause it to crash, or because a similar drone had crashed in a previous deployment after being subject to a jamming operation, the potential for a jamming operation that could cause the drone to crash was reasonably foreseeable and thus the deploying State is responsible for the resulting harm. If, however, the jamming operation was not reasonably foreseeable because this type of intervention was highly sophisticated or unprecedented, legal causation cannot be established.

What about the situation where a malicious actor conducts an evasion attack against an AI-enabled system? For example, where a malicious actor manipulates the environment in which the system operates and, in doing so, tricks it into misidentifying an individual or object. Again, the reasonable expectation standard asks whether, once deployed, it was reasonably foreseeable that the system could be tricked in this way. If a State deploys an AI-enabled system without adequate training and testing and is not therefore reasonably confident that the system can overcome evasion attacks, it is responsible for the resulting use of force. Equally, if the feedback loop indicates to the operator that the AI-enabled system was previously tricked into misidentifying targets, the deploying State will be responsible for the resulting use of force if it re-deploys the system and the system is (again) subject to an evasion attack. If, however, the system went through comprehensive testing and training in line with prevailing industry standards and during previous deployments it proved resistant to evasion attacks, the intervention by the third-party is unforeseeable and thus constitutes a *novus actus interveniens*. To be clear, the deploying State cannot be designated the legal cause of the use of force.

Also consider a Stuxnet-type scenario where a State manages to surreptitiously deploy an AI cyber capability into a nuclear facility's air-gapped network. If a technician in the facility inadvertently releases the capability into the wider Internet and it goes on to cause harm in other States, the State can be regarded as the legal cause of that harm even if it took reasonable steps to ensure that the capability would only operate on the facility's network. This is because, despite the State's best efforts to contain the operation, it was nevertheless reasonably foreseeable that a clandestine capability (a virus) could escape from the facility and cause harm. In this situation, the act of the technician does not constitute a *novus actus interveniens*.

Let us again consider the responsibility of a State that maliciously interferes with the performance of an AI-enabled system and that leads to forcible effects within another State. Imagine, for instance, the situation where an adversarial State conducts an evasion attack against an AI-enabled drone in order to trick it into misidentifying targets. I have already said that, according to the "but for" test, the adversarial State can be regarded as a factual cause of the force. In my view, legal causation is also established because any reasonable State in the adversarial State's position would have been aware that the interference would result in the use of force.

Consider further the situation where an adversarial State commits a jamming operation against an AI-enabled drone, leading it to crash and cause harm to people or property. While factual causation can be established using the "but for" test, legal causation hinges on what a reasonable State would have foreseen in the adversarial State's position. If a reasonable State would have foreseen that the jamming operation would merely prevent the drone from communicating with its operator, or prevent it from launching attacks, the adversarial State would not be responsible if the jamming operation unexpectedly prevented the drone from remaining airborne and led to it crashing. But if it was reasonable to assume that the jamming operation would prevent the drone from remaining airborne, the adversarial State is a legal cause of the forcible effects generated by the downing of the drone.[133]

133. *Articles on State Responsibility*, *supra* note 38, art. 47(1), explains that, "[w]here several States are responsible for the same internationally wrongful act, the responsibility of each State may be invoked in relation to that act." Determining the extent to which these States must provide reparations to the injured State is a different question and falls beyond the scope of this article. That said, Article 47(2)(a) states that Article 47(1) "does not permit any injured State to recover, by way of compensation, more than the damage it has suffered." On reparations where there are multiple causes, *see* Lanovoy, *supra* note 85, at 65–78.

Finally, what about the situation where an AI-enabled system engages its decision-making capacity to use force where this capability was not granted to it by the developer? Consider the scenario where AI-enabled software is programmed to covertly exfiltrate data from another State's cyber infrastructure but, through machine learning that allows it to adapt to inputs received from the environment in which it operates, it decides to delete data or shut down networks in order to meet perceived objectives. As far I understand it, current technology is not at this advanced stage. However, the reality is that AI technology is developing rapidly and, as we move towards the next generation of AI, it may be possible that AI-enabled systems take on a life of their own and develop capabilities that are not programmed by developers, such as the ability to conduct attacks.

It is therefore worth considering how the prohibition on the use of force applies in this type of scenario. Does the development of these types of attack capabilities constitute an unforeseen event that breaks the chain of legal causation and thus precludes the imposition of responsibility for a breach of Article 2(4)? Or can it be said that, if developers and operators lack the confidence that the technologies they deploy will *not* go on to develop use of force capabilities, the potential for these technologies to do so means that the State is responsible for the forcible attacks they commit because such attacks were reasonably foreseeable? To avoid the emergence of an accountability gap—which, after all, is one of the most pressing concerns when it comes to the use of AI[134]—my view is that States should not deploy AI-enabled systems unless they are reasonably confident as to the activities or behavior that they will engage in. If they decide to launch these systems without this confidence, they should be held responsible for any resulting forcible effects because a reasonable State in their position would have foreseen the potential for force to be used.

V. CONCLUSION

Despite the copious amounts of academic literature written on the nature, content, and scope of the prohibition on the use of force, very little attention has focused on the elements of intention and causation. Developments in

134. *See, e.g.*, Rebecca Crootof, *AI and the Actual IHL Accountability Gap*, CENTRE FOR INTERNATIONAL GOVERNANCE INNOVATION (Nov. 28, 2022), https://www.cigionline.org/articles/ai-and-the-actual-ihl-accountability-gap/.

new technologies such as cyberspace and AI have placed intention and causation under an intense international legal spotlight and, in recent years, important State practice has emerged.

It seems that States have, for very good reasons, rejected the contention that intention or negligence forms part of the prohibition on the use of force, and have instead placed more emphasis on whether the impugned conduct caused the forcible effects. Causation comprises two elements. This article has explained that factual causation, which asks whether the forcible effects would have occurred but for the State's conduct, will be readily established where a State deploys an AI-enabled system and that the critical issue is legal causation. Legal causation asks whether the forcible effects were reasonably expected when the AI-enabled system was launched. This article has explored in particular how this test applies when AI-enabled systems commit unintended uses of force.

By examining the issues of intention and causation, it is hoped that this article will help States formulate their positions on how the prohibition on the use of force applies to AI-enabled systems and, indeed, to new and emerging technologies more generally.