GENEVA ACADEMY
Académie de droit international humanitaire et de droits humains
Academy of International Humanitarian Law and Human Rights

RESEARCH BRIEF

DIGITAL DISINFORMATION OPERATIONS: PART II – CHARTING THE INTERNATIONAL LEGAL FRAMEWORKS

ABSTRACT

Digital disinformation operations (disinfo-ops) exploit automated influence tactics, online vulnerabilities and legal ambiguities to distort public discourse, manipulate governance processes, and evade accountability. As these campaigns operate below the traditional thresholds of armed conflict, existing international legal frameworks are not fully adapted to effectively address the threat. This report examines how international law applies to disinfo-ops, highlighting gaps in legal accountability, challenges in attribution, and the evolving role of empirical research in strengthening enforcement mechanisms. By integrating empirical data into legal interpretations, States can develop more objective and durable legal obligations to counter what is a rapidly expanding crisis.

JANUARY 2025 | STEVEN J. BARELA

This publication has been externally peer-reviewed

THE GENEVA ACADEMY A JOINT CENTER OF
GENEVA GRADUATE INSTITUTE
INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES
UNIVERSITÉ DE GENÈVE
FACULTY OF LAW

INTRODUCTION

The first report in this series examined how mis- and disinformation thrive in a digital ecosystem characterized by inauthentic activity, data-driven microtargeting and algorithmic amplification. It explored the sharp tension between the fundamental rights to freedom of expression and opinion and the urgent need to mitigate the harms of disinformation. Through an analysis of the mechanisms of proliferation, it detailed how bots, trolls, deepfakes, AI-generated content and precise targeting contribute to a distortion of individualized newsfeeds. The fact that these devices represent synthetic activity is important; a key recommendation is to ensure that real humans are interacting in digital spaces.[1] While this phenomenon raises critical human rights concerns, it also presents profound regulatory challenges across legal frameworks. In response, this second report examines the ways in which international law applies to digital operations to seed manipulated information – referred to here as disinfo-ops. These campaigns raise critical legal questions, particularly regarding sovereignty, non-intervention, and self-determination, while also exposing gaps in attribution and accountability due to the absence of a clear, binding legal framework.

The fragmented nature of the legal response complicates efforts to mitigate these threats. Disinformation often violates multiple legal paradigms simultaneously, creating compounded harms that transcend the scope of any single framework. Viewing it through a narrow legal lens risks overlooking its interconnected impacts and allows malicious actors to continue undermining confidence in the digital information space. This report navigates these complexities by identifying where existing legal principles provide guidance, where ambiguities persist, and where progressive legal development may be necessary, advocating for a holistic approach to ensure international law evolves effectively to address the multifaceted dangers posed.

Given the borderless nature of digital information, this report also emphasizes the importance of empirical data and interdisciplinary collaboration in understanding the "scale, scope, and precision" of these campaigns spanning multiple jurisdictions.[2] Without transparent access to platform data to reveal the quantitative extent and impact, the legal and policy responses to disinfo-ops remain hindered by limited visibility into the mechanisms driving its spread. Indeed, accurate interpretations of the law and its evolution must be based on factual data to ensure that it is both objective and durable.

MULTIPLE INTERNATIONAL LEGAL FRAMEWORKS

Disinformation campaigns operate below the threshold of an "armed attack" or conventional war as defined under international law. Unlike cyberwarfare, which is generally understood to involve acts that result in physical destruction, injury, or loss of life, the effects here are subtle but pernicious.[3] This ambiguity creates a significant challenge in finding the proper legal paradigm to address these activities.

The application of international law to cyberspace has been a focal point of legal debate following the 2007 cyber-attacks on Estonia.[4] Two key questions have dominated the discourse: whether international law applies in the new space created by information and communication technologies (ICTs), and if so, how. The United Nations (UN) Group of Governmental Experts (GGE) has played a pivotal role in addressing these questions. Its 2013 report affirmed that international law, including the UN Charter, applies to cyberspace, emphasizing that state sovereignty and its associated norms govern ICT-related activities. It declared that efforts to enhance cybersecurity must align with human rights and fundamental freedoms, and that states bear responsibility for internationally wrongful acts committed using ICTs, including acts by proxies or non-state actors operating from their territory – a stance endorsed by key States such as China, Russia, and the United States, as well as the European Union.[5] Beyond that, progress on the second question – how international law applies – has been more contentious. The 2015 UN GGE report explicitly confirmed that states have jurisdiction over ICT infrastructure within their territory, clarifying that sovereignty constrains state behavior in cyberspace, and reinforcing the principle of non-intervention by emphasizing that states must not use ICTs to interfere in the protected affairs of other states.[6] Disagreements over self-defense, international humanitarian law, and countermeasures led to a breakdown in 2017 negotiations,[7] though the 2021 GGE report renewed optimism for advancing consensus.[8]

In response, the Open-Ended Working Group (OEWG), inclusive of all interested states, was established. In 2021, the group consolidated existing norms of responsible state behavior, emphasizing confidence-building measures, but made limited progress on accountability mechanisms due to persistent geopolitical divisions.[9] The 2023 OEWG report expanded state participation and political engagement,

with increased focus on cyber threats in active conflicts (particularly in relation to Russia's cyber operations in Ukraine) while advancing discussions on a potential programme of action despite ongoing disagreements over the application of international humanitarian law to cyberspace.[10]

THE PRINCIPLE OF NON-INTERVENTION

The principle of non-intervention lies at the heart of the Westphalian system, embodying the fundamental notion of sovereign equality and the right of each State to determine its internal and external affairs free from coercion. It is a cornerstone of international law, serving as a corollary to state sovereignty, territorial integrity, and political independence. Simply put, it prohibits forceful influence on other States. While firmly rooted in customary international law and reinforced by key decisions of the International Court of Justice (ICJ), the principle's contours remain elusive.[11] In view of its amorphous status, especially in the context of emerging information operations, a proposal for the progressive development of international law on this point is discussed below.

The unlawfulness of interference and intervention has been included in over thirty-five resolutions passed by the UN General Assembly (UNGA).[12] However, many have often been passed on divided votes and cannot be said to be authoritative. Among these instruments, the most significant would be those passed with a substantial majority: the 1965 Declaration on the Inadmissibility of Intervention,[13] the 1970 Declaration on the Principles of Friendly Relations (agreed upon without formal voting, signifying unanimity),[14] and the 1981 Declaration on the Inadmissibility of Intervention and Interference.[15]

Moreover, this principle exists alongside treaty-based norms, particularly the UN Charter's prohibition of force in Article 2(4). While the Charter explicitly addresses uses of force, customary international law provides a broader safeguard against non-forcible coercion. The ICJ's recognition of the principle as a standalone norm in the *Nicaragua* case underscores its independent legal status.[16] This case is seen as an important moment in the development of the principle and brought a significant degree of clarity since meddling across borders has frequently occurred.[17] The *Nicaragua* decision emphasized coercion as the defining element of unlawful intervention, stating that acts interfering with a State's choice of political, economic, or social systems are wrongful if they force action against the State's will.[18] In a nutshell, the Court found that for an act to qualify as unlawful intervention, two elements must be present:

1. **Domaine Réservé**: An action must be taken within the sphere of another State where it has the authority to choose policy freely.
2. **Coercion**: The act must compel the target State to adopt policies or actions it would not otherwise choose.

While these criteria have provided clarity, they have also exposed limitations. Coercion in traditional settings often involved economic measures, support for subversive groups, or direct military action. However, this framework is less equipped to address the subtler and indirect methods emerging in the digital age.

Operations in Cyberspace

In the digital realm, States increasingly resort to low-intensity operations, which fall below the threshold of the use of force but can significantly impact other States' affairs. These operations challenge existing legal frameworks by exploiting gaps in the principle of non-intervention.[19] Digital operations in cyberspace may not involve direct compulsion but still manipulate political, economic, or social outcomes. Influence campaigns or disinformation operations to undermine public trust in governance have been discussed in Part I of this report. Moreover, one can also include cyber espionage to extract sensitive information to gain strategic advantages and economic disruptions using cyber-attacks to target financial or infrastructure systems. These actions blur the lines between permissible and impermissible conduct, as they do not always manifest as overt coercion yet disrupt the internal affairs of targeted States.

The *Tallinn Manual Project* is a critical resource for understanding how international law applies to cyber operations.[20] As interpreted by a panel of international legal experts, the *Tallinn Manual 2.0* clarifies that the principle of state sovereignty extends to cyberspace – as asserted by the UN GGE. This includes a clear prohibition on cyber operations that infringe on another State's sovereignty, affirming that such actions violate fundamental principles of international law.[21] Regarding the principle of non-intervention, Rule 66 specifies that States "may not intervene, including by cyber means, in the internal or external affairs of another State."[22] This interpretation draws directly on the ICJ's ruling in the *Nicaragua* case, underscoring that prohibited intervention requires coercion and must affect the *domaine réservé* of the target State.[23]

The commentary to Rule 66 elaborates on what

constitutes a State's internal affairs, reaffirming that decisions related to the political system and its organization lie at the core of sovereignty.[24] For instance, the experts highlight cyber operations designed to "alter electronic ballots" as a clear example of unlawful intervention.[25] While the foundational rules outlined in *Tallinn Manual 2.0* remain unchanged from its earlier iteration, it is noteworthy that *Tallinn Manual 1.0* included explicit identification of the spread of "false news" on the eve of elections as a possible form of unlawful interference.[26] There is no indication that the absence of this language in the more recent version is a change of view on the law.

In discussing the elements of unlawful intervention, the concept of coercion has been a particular point of contention.[27] The *Tallinn Manual 2.0* defines coercion as an act intentionally "designed to deprive another State of its freedom of choice," effectively forcing it to act against its will or to abstain from an action it would otherwise undertake.[28] The experts explicitly differentiate coercion from less intrusive forms of influence, such as "persuasion, criticism, public diplomacy, propaganda," or malicious acts, which, while potentially disruptive, do not compel action from the target State.[29] Unlike coercion, these activities either aim to influence voluntary actions or do not seek to elicit a specific response at all. For an act to constitute coercion, it must have the capacity to compel the target State to take or refrain from taking a specific action against its intended course of behavior.[30]

Several States have now adopted a view aligning with *Tallinn 2.0* that cyber operations are unlawful when they are coercive and interfere with a State's internal or external affairs.[31] A few States have also assumed broader approaches to coercion than what was espoused in the 2017 manual. For instance, Australia approved that coercion involves actions that "effectively deprive the State of the ability to control, decide upon or govern matters of an inherently sovereign nature."[32] Germany adopted the position that coercion in cyberspace involves situations where a State's internal processes are "significantly influenced or thwarted" and its "will is manifestly bent" by the conduct.[33] The UK has advocated for a broader interpretation of coercion than that outlined in the *Tallinn Manual 2.0*, suggesting that in some instances, "disruptive cyber behaviors" could qualify as coercive even without clear evidence of specific actions a State was compelled to take or refrain from undertaking.[34] It can thus be seen that there is some appetite in the international community for considering a broader view in the context of cyber operations.

The *Domaine Réservé* and Legitimacy

At the core of the non-intervention principle is the concept of the *domaine réservé*: the realm of affairs in which States retain exclusive control. Historically, the Permanent Court of International Justice (PCIJ) defined the *domaine réservé* as matters over which each State is the singular arbiter and "not, in principle, regulated by international law".[35] This concept is inherently dynamic, shifting as international law is created to cover evolving shared areas of concern such as trade, human rights, and environmental regulation.[36]

Legitimacy is an essential sphere for every State, and a keystone of governance. For any group to act in concert, obedience must flow freely to the authority in power.[37] When legitimacy is sufficiently eroded, a society can become inert.[38] Elections, as mechanisms for conferring legitimacy, are particularly vulnerable to digital interference. Manipulating the decision-making of individuals during election processes, or undermining public confidence in their integrity, can destabilize governance and disrupt the social contract. Such actions, even when conducted below the threshold of force, strike at the heart of a State's sovereign prerogatives and – it is argued here – operate with the *domaine réservé* of a State.

A Quantifiable Standard of Coercion

The requirement of coercion distinguishes prohibited intervention from other forms of influence. Yet it is useful to underscore the fact that "international law provides no conclusive definition of the term."[39] Conventionally understood it can be taken as a form of compelling another actor to do one's will.[40] This definition is the standard approach and generally understood to be the *lex lata*.[41] However, as there is no decisive meaning of coercion it is argued here that in the context of new digital circumstances there is a pressing need to interpret the term progressively.[42] International legal scholar Tom Farer offers a compelling explanation of the issue posed by this specific idiom:

The nub of the matter is that the word 'coercion' has no normative significance; there is nothing illegal about coercion. Coercion is normal in all human relationships, including those between lovers. It's part of life. So is cooperation. Indeed, every human relationship is some mixture of coercion and cooperation.[43]

In cyberspace, coercion takes on new forms, challenging traditional interpretations. There has been cogent argument for a spectrum-based understanding of coercion,

emphasizing factors such as the significance of State interests affected, the scale of disruption, and the extent of involuntary harm inflicted.[44] Furthermore, a minority view in *Tallinn 2.0* put forward that "it is impossible to prejudge whether an act constitutes intervention without knowing its specific *context and consequences*."[45] It is therefore argued that cyber operations that significantly alter electoral processes, compromise confidence in election results, or widely disseminate precisely targeted disinformation to manipulate public opinion, can meet these criteria, even if their methods appear less invasive.

ICJ Judge Rosalyn Higgins has written on the idea of expanding the notion of coercion to include violations of jurisdiction.[46] This approach could align with the dimensions of "consequentiality" put forward in 1958 by McDougal and Feliciano which consider the values affected, the extent of their disruption, and the actors involved.[47] By applying such frameworks to cyber operations, low-intensity actions such as disinfo-ops could qualify as coercive interventions, warranting tighter scrutiny under international law.

To address this problem, some international jurists (including this author) have proposed infusing the concept of "coercion" with empirical content to address the challenges posed by cyber operations.[48] This approach would involve assessing coercion based on measurable factors, such as the scale, reach and accuracy of an operation, to determine its impact on a target State's autonomy. By focusing on empirical criteria, this framework bridges the gap between traditional understandings of coercion emphasizing direct compulsion, and subtler indirect methods of influence enabled by cyberspace. Data access for independent study would be critical to operationalizing such a framework, enabling the collection of quantitative evidence to assess the size and impact of cyber operations.[49] Access to detailed datasets can illuminate the precise mechanisms and impacts of disinfo-ops and other cyber operations, providing the empirical foundation needed for more precise legal interpretations. Such an approach would provide a more objective basis for analyzing whether a cyber operation undermines a State's decision-making freedom, creating a measurable and verifiable standard to reduce abuse in the self-administered international legal system.

SELF-DETERMINATION

The right to self-determination is another cornerstone of international law, affirming that peoples have the collective right to freely determine their political status and pursue their social, cultural, and economic development without external interference.[50] Self-determination is not only about a nation's independence, but there is a space for it to be also understood as ensuring the integrity of its internal political processes. Information campaigns that distort democratic processes or manipulate public opinion strike at the heart of this right. It has been argued by Jens David Ohlin that by impersonating citizens, amplifying divisive rhetoric, or disseminating false narratives, these campaigns interfere with the electorate's ability to make informed decisions, undermining the authenticity of democratic deliberation.[51]

Historically, the right to self-determination has been associated with colonized or stateless peoples striving for autonomy and independence.[52] However, its relevance extends beyond statehood and decolonization as affirmed by leading judicial authorities.[53] Ohlin has argued that it should encompass the protection of democratic processes within established States. By analyzing election interference through the lens of self-determination, the collective harm inflicted on a population's ability to freely shape its political destiny becomes clearer.[54] It is put forward here that this approach complements the principles of sovereignty and non-intervention, offering a more nuanced understanding of the distinct challenges posed by digital interference. Their reciprocal nature will be discussed below.

While the concept of self-determination provides a valuable analytical framework, several challenges hinder its practical application to election interference. First, international legal practice has traditionally limited self-determination to contexts such as decolonization and the rights of stateless peoples, leaving its relevance to established States underexplored. Second, the lack of widespread State practice or *opinio juris* recognizing election interference as a breach of self-determination complicates its enforcement under customary international law. Third, the extraterritorial nature of election interference raises questions about whether self-determination applies when the offending State acts entirely outside the territorial boundaries of the affected State.[55]

Progressive Development for the Digital Realm

Digital disinformation can disrupt the mechanisms through which self-determination is exercised, turning democratic participation into a manipulated spectacle rather than a genuine expression of popular will. The digital age has heightened these challenges, as foreign actors exploit online platforms to infiltrate information spaces shared by citizens and policymakers alike. Modern election interference, as illustrated by the Russian meddling in the 2016 U.S. elections, represents a confluence of cyber-attacks, social media manipulation, and strategic propaganda efforts; it involved activities such as hacking email accounts, deploying troll farms on social media platforms, and infiltrating domestic advocacy groups to subtly alter public opinion and political discourse.[56] The paradigm of self-determination can provide an important understanding of the legal obligations of States. Through this lens, self-determination emerges as an additional legal framework for addressing the unique harm posed by influence operations to democracy in the modern era.[57]

Scholars and practitioners must explore a broader interpretation of self-determination that reflects contemporary threats, particularly those emerging from digital technologies. This entails recognizing that self-determination is not merely a right tied to national liberation, but also a principle that holds a great deal of potential to protect the democratic processes through which a population expresses its collective will. The expansion of this framework aligns with the evolving realities of modern governance, where information manipulation transcends borders and undermines the integrity of political systems and pluralistic societies. Nevertheless, it is important to note that at this point, "neither the practice of States nor their expressions of *opinio juris* are sufficiently uniform and consistent" to support the conclusion that such a view has been embraced.[58]

Integrating Sovereignty and Self-Determination

The principles of sovereignty and self-determination, while distinct, are deeply interconnected. Sovereignty establishes a state's supreme authority over its affairs, protecting it from external manipulation, while self-determination safeguards the collective will of a population to freely shape its political, social, and economic future. Digital disinformation challenges both principles simultaneously by blurring the lines between foreign interference and internal manipulation, undermining the legitimacy of governance and democratic processes.

Election interference serves as a prime example of this dual violation. As Nicholas Tsagourias has argued, while the principle of non-intervention seeks to protect "the integrity and autonomy of a state's authority and will", self-determination reflects "the process of authority and will formation" within a population.[59] Viewed through this lens, their interconnectedness becomes evident – sovereignty provides the legal framework that enables self-determination, while self-determination gives legitimacy to sovereignty through the free and unmanipulated expression of the people's will. Disinformation campaigns disrupt both principles by infringing on a state's internal affairs and distorting the processes through which the collective will is formed. By integrating these principles in this context, international law can better address the multifaceted threats posed by election interference, recognizing both its structural and human impacts.

However, applying these frameworks to digital interference presents legal challenges. As Jens Ohlin has noted, there is an unresolved question of whether digital operations should be considered "coercive" or merely "corrosive" to democratic functions.[60] Traditional interpretations of sovereignty and non-intervention are territorial in nature, making them a more difficult fit to non-physical cyber incursions. In contrast, self-determination better captures the political and psychological harm inflicted by digital interference, offering a legal framework that aligns more closely with the unique nature of election manipulation.

The UN General Assembly Declaration on Principles of International Law concerning Friendly Relations and Cooperation among States recognizes that self-determination and sovereignty reinforce one another. It affirms that "the principle of equal rights and self-determination of peoples constitutes a significant contribution to contemporary international law and... its effective application is of paramount importance for the promotion of friendly relations among States, based on respect for the principle of sovereign equality."[61] This underscores the reasoning behind a holistic approach to the two legal concepts in this context. Moreover, addressing these challenges requires greater clarity on how non-intervention applies to digital operations and how self-determination can be safeguarded against manipulation at scale.

Ultimately, advancing legal responses to digital disinformation will depend on empirical data that helps define thresholds for intervention and the manipulation of authority and will formation. Identifying measurable

indicators – such as the reach and individualized accuracy of disinfo-ops – will be crucial in determining when an operation constitutes mere influence or an outright violation of both sovereignty and self-determination. Embedding quantifiable limits into legal frameworks will allow international law to evolve in response to the growing challenges.

LEGAL ACCOUNTABILITY: ATTRIBUTION AND DUE DILIGENCE

Attribution is a critical foundation for accountability. It underpins the application of numerous international legal frameworks, including those governing State responsibility, human rights, armed conflict and cyber activities, providing a baramoter for assessing legal consequences in various contexts. These paradigms define State obligations and individual rights, but their effectiveness depends on the ability to identify and assign responsibility for harmful actions.[62]

While attribution clarifies accountability and enables enforcement, it is often hindered by technical and evidentiary challenges. Many cyber operations cannot be directly or legally linked to a State or non-state actor. In such cases, the principle of due diligence provides a complementary mechanism, obliging States to take reasonable steps to prevent unlawful acts originating from their territory that harm other States. This section examines the legal foundations of attribution, the challenges unique to cyberspace, and emerging proposals to strengthen accountability.

Challenges of Attribution in Cyberspace

Attribution in international law is primarily grounded in the principle of state responsibility. The International Law Commission's Draft Articles on Responsibility of States for Internationally Wrongful Acts requires that a State either directly conducts or exercises effective control over non-state actors responsible for an action.[63] However, these standards, derived from analog legal contexts, are often ill-suited to the anonymity and technical complexity of cyber activities. To hold a State accountable for a cyber operation, there must be a sufficiently close link between the harmful act and the State in question. This requires clear and convincing evidence to meet evidentiary thresholds.[64]

Non-binding instruments like the *Tallinn Manual 2.0* offer guidance on applying international law to cyberspace. Rule 15, for example, outlines criteria for attributing cyber operations to States, emphasizing the necessity of sufficient evidence and State involvement.[65] Similarly, efforts by the United Nations GGE and OEWG have promoted norms for responsible State behavior in cyberspace, focusing on transparency, confidence-building measures, and capacity building to enhance trust and collaboration in addressing cyber incidents, including attribution.[66] While these provide valuable frameworks for guiding international conduct, both leave States with significant flexibility in interpretation and implementation. This gap contributes to a fragmented legal landscape, where unclear applications of international law hinder accountability.

Cyber operations frequently employ anonymizing technologies, proxy networks, and false flags to obscure their origin, making it difficult to attribute actions to specific actors.[67] Many such operations also exploit legal gray zones, operating in spaces where existing legal frameworks are unclear or enforcement mechanisms are weak, allowing perpetrators to avoid direct violations of international law.[68] Advanced persistent threat (APT) groups further complicate attribution by mimicking other entities, using deceptive tactics to mislead investigators and obscure their true identity. Technical attribution relies on forensic methods, such as analyzing indicators of compromise (IoCs), identifying unique malware signatures, and comparing tactics, techniques, and procedures (TTPs) to known threat actors.[69] However, these methods rarely produce definitive proof, and real-time observations of malicious activities are often unavailable or incomplete, making attribution an inherently complex and uncertain process.

Beyond technical challenges, attribution decisions must balance transparency with the protection of intelligence sources. Publicly disclosing evidence can enhance credibility and deter further attacks but also risks exposing sensitive methods, compromising ongoing investigations, or revealing state capabilities. This tension often leads to underreporting of incidents or cautious, low-confidence attributions, which in turn erodes trust in the attribution process and limits the ability to hold perpetrators accountable.[70] The lack of definitive attribution can also create opportunities for deniability, geopolitical maneuvering, and further exploitation of digital vulnerabilities, reinforcing the strategic advantages cyber operations offer to malicious actors.

Advancing Cyber Attribution

Useful proposals have been advanced to address the gaps and challenges in attribution while promoting accountability in cyberspace. One approach is the development of a

specialized legal framework (*lex specialis*) tailored to cyber attribution. Such a framework would establish consistent evidentiary standards, clarify thresholds for responsibility, and differentiate between lawful and unlawful cyber operations. Codifying these standards could reduce the risk of misattribution and foster greater trust in attribution processes.[71]

Enhancing multilateral collaboration can also be essential for attribution credibility. Building on existing soft law frameworks, States could establish formal mechanisms for joint attributions, including coordinated information-sharing and joint investigations. These confidence-building measures would foster trust among allies, reduce geopolitical tensions, and enhance collective responses to cyber threats. Coordinated attributions by like-minded States would increase credibility, limit adversarial deniability, and impose greater costs on malicious actors, reinforcing norms of responsible state behavior in cyberspace.[72]

Moreover, private entities, including cybersecurity firms, academics, and think tanks, play a crucial role in attribution. Their technical expertise and access to global networks enable faster and more detailed investigations than many governments can achieve. Integrating public-private partnerships into attribution efforts would enhance cross-verification, transparency, and reliability.[73]

Digital Due Diligence

Due diligence also offers a compelling mechanism for addressing cross-border cyber harm. Unlike the principle of state responsibility, which hinges on attributing specific actions to States, due diligence emphasizes a State's obligation to prevent harmful activities originating from its jurisdiction. This principle holds particular promise in cyberspace, where the anonymity of actors and the transboundary nature of cyber operations often complicate traditional frameworks of accountability.[74]

Even if its precise contours are not clear, the concept of due diligence is firmly rooted in international law dating back to the 19th century,[75] and affirmed by the ICJ in 1949.[76] Today it is present in areas like environmental law where States are required to ensure that activities within their territory do not cause harm beyond their borders.[77] In the cyber domain, due diligence obliges States to take reasonable measures to prevent, mitigate, or address malicious cyber activities that could harm another State. While not codified as a binding legal rule, due diligence has gained traction in soft law instruments like the *Tallinn Manual 2.0* where it is articulated in Rule 6 as an obligation for States to prevent their territory from being used to conduct harmful cyber operations against other States, emphasizing reasonable measures and knowledge thresholds as key components of this principle.[78]

The promise of due diligence in cyberspace lies in its practical adaptability. Unlike State responsibility, due diligence does not require definitive attribution of harmful acts to a State. Instead, it imposes an obligation on States to act once they are aware of malicious activities emanating from their territory or infrastructure. The principle of due diligence creates a pathway for addressing harmful cyber operations without relying on precise attribution, focusing instead on a State's capacity to exercise control over its cyber infrastructure and mitigate risks.[79] States are expected to take feasible steps to enhance cybersecurity, including developing capabilities, engaging in international cooperation, and, where appropriate, implementing monitoring measures to mitigate cyber threats. While due diligence does not impose a strict duty to monitor all cyber activities, States should take reasonable and proportional measures to prevent serious harm originating from within their jurisdiction, regardless of whether the actors involved are State-affiliated or private.[80]

Despite its potential, the principle of due diligence faces challenges, including the lack of universally agreed thresholds for harm and the difficulty of defining "reasonable" State action in the highly technical cyber domain. While the duty does not impose an obligation to actively monitor all cyber activity, States should act when they possess actual or constructive knowledge of harmful operations. This standard, while useful, leaves room for interpretation, creating uncertainty around enforcement and compliance.[81]

Antonio Coco and Talita de Souza Dias propose a progressive approach to due diligence, advocating for a tiered framework that reflects varying degrees of State involvement and control. They argue that States should bear graduated obligations based on their capacity to act and the extent to which malicious cyber activities can be traced to their territory or infrastructure. For instance, when a State has full control over infrastructure enabling cyber operations, it must take robust preventive measures. Where its control is more limited, such as activities by non-state actors, States may instead focus on cooperative measures, including intelligence sharing and international coordination. This nuanced model aims to balance accountability with practicality, ensuring that the principle of due diligence remains both adaptable and enforceable in an increasingly complex cyber landscape.[82]

CONCLUSION

Disinfo-ops thrive by anonymously flooding tailored information streams with synthetic and polluted content, which real people engage with as though it were authentic. Identifying polluters and curbing their contamination fosters trust and enables accountability. In turn, these efforts preserve the integrity necessary for informed discourse needed for democratic governance and scientific advancement. Like any other ecosystem, the digital information space requires methodical study, coordinated intervention, and vigilant stewardship. However, without a clear legal and empirical foundation, efforts to address disinformation remain reactive and fragmented, underscoring the urgent need for structured research and internationally coordinated legal responses.

Central to these efforts is enabling data access for independent scientific study. Advancing cross-disciplinary and cross-community collaboration is essential to ensuring that research reflects diverse perspectives and that underrepresented communities play a role in shaping solutions. The Geneva Academy's Info-Brief on EU Data Access to Study Digital Disinfo-Ops highlights a key initiative in this area, demonstrating how structured researcher access to privacy-preserving platform data contributes directly to addressing disinformation operations. Supporting initiatives that foster such collaboration – including mechanisms that provide resources for long-term, inclusive research – will be critical in fully understanding how disinformation spreads and impacts societies, strengthening every dimension of the response.

The international legal framework addressing these operations would greatly benefit from enhanced empirical understanding. Foundational principles such as non-intervention and self-determination are increasingly tested by digital disinformation campaigns, which challenge traditional definitions of coercion, distort the democratic processes of will formation, and blur accountability lines. Independent research can help quantify coercive acts, build measurable benchmarks for electoral interference that violates self-determination, and help create identification schemes for nefarious actors.

By integrating rigorous empirical research into legal interpretations, these frameworks can become more objective, enforceable, and durable, equipping States with practical tools to uphold their international obligations while effectively countering digital disinformation. Attribution and due diligence are particularly reliant on robust empirical evidence. The anonymity and technical complexity of disinformation campaigns often obscure responsibility, frustrating efforts to assign accountability or prevent harm. Independent study can support the development of more precise attribution mechanisms and enhance due diligence standards, enabling States to take proactive measures against harmful activities. As digital disinfo-ops increasingly exploit legal gray zones to evade traditional frameworks, empirical research that quantifies thresholds of harm can bring sharper definition to what is lawful and unlawful, turning ambiguity into enforceable standards.

END NOTES

1 For a proposal to curb automated and synthetic activity, preserve privacy and promote human interaction, see S Hallensleben, Trust in the European Digital Space in the Age of Automated Bots and Fakes (European Observatory of ICT Standardisation, January 2022) <https://www.standict.eu/news/trusted-information-digital-space> accessed Dec 2024.

2 S Bradshaw and P Howard, The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation (Project on Computational Propaganda 2019) 11.

3 JD Ohlin, C Finkelstein, and K Govern (eds), Cyber War: Law and Ethics for Virtual Conflicts (OUP 2015).

4 D Hollis, 'Why States Need an International Law for Information Operations' (2007) 11(4) Lewis & Clark Law Review 1023–1061; E Tikk, K Kaska, and L Vihul, International Cyber Incidents: Legal Considerations (Cooperative Cyber Defence Centre of Excellence, 2010); C Droege, 'Get Off My Cloud: Cyber Warfare, International Humanitarian Law, and the Protection of Civilians' (2012) 94(866) International Review of the Red Cross 533–578.

5 UN Group of Governmental Experts (UN GGE), Report on Developments in the Field of Information and Telecommunications in the Context of International Security (24 June 2013) UN Doc A/68/98. Notably, there was not agreement to refer to international humanitarian law as applicable in cyberspace.

6 UN GGE, Report on Developments in the Field of Information and Telecommunications in the Context of International Security (22 July 2015) UN Doc A/70/174, paras 24–28. The 2013 report did not address IHL explicitly, but the 2015 report introduces the principles of humanity, necessity, proportionality, and distinction.

7 For commentary, see M Schmitt and L Vihul, 'International Cyber Law Politicized: The UN GGE's Failure to Advance Cyber Norms' (Just Security, 30 June 2017) <https://www.justsecurity.org/42768/international-cyber-law-politicized-gges-failure-advance-cyber-norms/> accessed Dev 2024.

8 UN GGE, Report on Developments in the Field of Information and Telecommunications in the Context of International Security (14 July 2021) UN Doc A/76/135. For commentary, see M Schmitt, 'The Sixth United Nations GGE and International Law in Cyberspace' (Just Security 10 June 2021) <https://www.justsecurity.org/76864/the-sixth-united-nations-gge-and-international-law-in-cyberspace/> accessed Dec 2024.

9 UN Open-Ended Working Group (OEWG), Final Substantive Report on Developments in the Field of ICTs in the Context of International Security (10 March 2021) UN Doc A/AC.290/2021/CRP.2; for commentary see P Meyer and D Stauffacher, ICT4Peace and the United Nations Open-Ended Working Group on International Cybersecurity (UN OEWG) 2019-2021 (ICT4Peace Foundation, 2021).

10 UN OEWG, Annual Progress Report of the OEWG (2021–2025) on Developments in the Field of ICTs in the Context of International Security (11 December 2023) UN Doc A/78/180; for commentary see Paul Meyer, The Plot Thickens: The UN Open-Ended Working Group on ICTs – Fourth Session (ICT4Peace Foundation, 2023).

11 M Roscini, International Law and the Principle of Non-Intervention: History, Theory, and Interactions with Other Principles (OUP, 2024) ch 1-2; M Jamnejad and M Wood, 'The Principle of Non-Intervention' (2009) 22 Leiden Journal of International Law 345.

12 ibid Jamnejad and Wood, 349–351; C Keitner, 'Foreign Election Interference and International Law' in Hollis and Ohlin, Defending Democracies: Combating Foreign Election Interference in a Digital Age (OUP 2021) 181-187.

13 United Nations General Assembly (UNGA), 'Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of their Independence and Sovereignty', UN Doc. A/2131 (XX) (December 21, 1965), passed with 109 to 0, with one abstention.

14 UNGA, 'Declaration on Principles of International Law concerning Friendly Relations and Cooperation among States in accordance with the Charter of the United Nations', UN Doc. A/RES/2625 (XXV) (October 24, 1970).

15 UNGA, 'Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States', UN Doc. A/Res/36/103 (December 9, 1981); passed 120 for, 22 against, and 6 abstentions.

16 International Court of Justice, Case Concerning Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v. United States of America) (Merits) [1986] ICJ Rep. 14, para 174.

17 One scholar has tallied that the United States and the USSR/Russia have engaged in partisan electoral meddling 117 times between 1946 and 2000, see D Levin, 'Partisan Electoral Interventions by the Great Powers: Introducing the PEIG Dataset' (2019) 36 Conflict Management and Peace Science 88.

18 Nicaragua (n 16) para. 205.

19 See eg, S Watts, 'Low-Intensity Cyber Operations and the Principle of Non-Intervention' in Cyber War (n 3) 249–270.

20 M Schmitt (ed.), Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (CUP 2017) (hereinafter Tallinn 2.0).

21 ibid, Rule 4, 17-27.

22 ibid, 312.

23 ibid, 314.

24 ibid, 315.

25 ibid, 313.

26 M Schmitt (ed), *Tallinn Manual on International Law Applicable to Cyber Warfare* (CUP 2013) at 45.

27 See I Kilovaty, 'The Elephant in the Room: Coercion' (2019) 113 AJIL Unbound 87.

28 Tallinn 2.0 (n 20) 317, citing Robert Y. Jennings and Arthur Watts (eds), Oppenheim's International Law, Volume 1, Peace (9th edn, Longman 1992), 430–431.

29 ibid, 318.

30 Ibid, 319.

31 See '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' UN Doc A/76/136* (July 13, 2021). For example, statements demonstrating this trend come from Estonia, Norway and Switzerland, along with New Zealand (New Zealand Foreign Affairs and Trade, 'The Application of International Law to State Activity in Cyberspace' (1 December 2020) 2 <https://dpmc.govt.nz/publications/application-international-law-state-activity-cyberspace>).

32 Australian Government, 'Australia's Position on how International Law applies to State Conduct in Cyberspace' (2017) <https://www.dfat.gov.au/sites/default/files/application-of-international-law-to-cyberspace.pdf> accessed Nov 2024.

33 Federal Government of Germany, 'On the Application of International Law in Cyberspace' (March 2021), <https://www.auswaertiges-amt.de/blob/2446304/32e7b2498e10b74fb17204c54665bdf0/on-the-application-of-international-law-in-cyberspace-data.pdf> accessed Nov 2024.

34 S Braverman, 'International Law in Future Frontiers' (Attorney General's Office, May 19, 2022) <https://www.gov.uk/government/speeches/international-law-in-future-frontiers> accessed Nov 2024.

35 Nationality Decrees Issued in Tunis and Morocco, Advisory Opinion, (7 February 1923), 1923 PCIJ Series B No. 4, at 24.

36 Roscini, Principle of Non-Intervention (n 11) ch 3.

37 The classic source on legitimacy is sociologist, philosopher, jurist, and political economist, Max Weber. His best-known articulation of this is 'Politics as a Vocation' in HH Gerth and C Wright Mills (eds and trs), From Max Weber: Essays on Sociology (OUP 1946) para 6. This speech was delivered at Munich University in January 1919, and published October of that year. See also, J Habermas, Legitimation Crisis (T McCarthy tr, Heinemann Educational Books 1976); D Beetham, The Legitimation of Power (Palgrave McMillian,1991).

38 G Ferrero, The Principles of Power (GP Putnam's Sons 1942); D Cook, 'Legitimacy and Political Violence: A Habermasian Perspective,' (2003) 30 Social Justice 108; M Crenshaw, Terrorism, Legitimacy, and Power: The Consequences of Political Violence (Wesleyan University Press 1983).

39 M Schmitt, '"Virtual" Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law' (2018) 19 Chicago Journal of International Law 30, 49-50.

40 UNGA, Friendly Relations Declaration (n 14): 'No State may use or encourage the use of economic political or any other type of measures to coerce another State in order to obtain from it the subordination of the exercise of its sovereign rights and to secure from it advantages of any kind.'; and Tallinn 2.0 (n 19) 317: 'an affirmative act designed to deprive another State of its freedom of choice, that is, to force that State to act in an involuntary manner or involuntarily refrain from acting in a particular way'.

41 Schmitt, '"Virtual" Disenfranchisement…' (n 39) 52-3.

42 For one expression of the need to eyes open in novel circumstances, see Rosalyn Higgins, 'Intervention and International Law' in Rosalyn Higgins (ed), Themes and Theories (OUP 2009) 272: "Rules are really only the accumulated body of past decisions, which, while an essential starting point, tell us little about variables and still less about changing circumstances".

43 T Farer, 'Political and Economic Coercion in Contemporary International Law' (1985) 79 American Journal of International Law 405, 406.

44 For compelling argument on the cyberspace context see S Watts, 'Low-Intensity Cyber Operations and the Principle of Non-Intervention' in Ohlin, Govern, and Finkelstein, Cyber War (n 3) 256-261.

45 Tallinn 2.0 (n 20) 319 (emphasis added).

46 Rosalyn Higgins, 'Intervention and International Law,' in Hedley Bull (ed), Intervention in World

Politics (Clarendon Press, 1984).

47 M McDougal and F Feliciano, 'International Coercion and World Public Order: The General Principles of the Law of War' (1958) 67 Yale Law Journal 771, 782–783. For discussions on disinformation in the cyber context see SJ Barela, 'Cyber Ops to Erode Legitimacy: An Act of Coercion' (Just Security, 12 January 2017) <https://www.justsecurity.org/36212/cross-border-cyber-ops-erode-legitimacy-act-coercion/> accessed Jan 2025; and I Kilovaty, 'Doxfare: Politically Motivated Leaks and the Future of the Norm on Non-Intervention in the Era of Weaponized Information' (2018) 9 Harvard National Security Journal 146. Kilovaty suggests that 'doxfare' should be interpreted as coercive when it causes 'disruption,' yet this criterion lacks verifiable constraints to avoid abuse.

48 SJ Barela and S Haataja 'Rethinking the International Law of Interference in the Digital Age,' in M Regan and A Sari (eds), Hybrid Threats and Grey Zone Conflict (OUP 2024) 393-424.

49 See S. Barela, Info-Brief on EU Data Access for Study of Digital Disinfo-Ops, Geneva Academy of International Humanitarian Law and Human Rights (2025).

50 International Covenant on Civil and Political Rights (ICCPR) (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171, art 1; International Covenant on Economic, Social and Cultural Rights (adopted 16 December 1966, entered into force 3 January 1976) 993 UNTS 3, art 1; Charter of the United Nations (adopted 26 June 1945, entered into force 24 October 1945) 1 UNTS XVI, arts 1(2) and 55; UNGA, Friendly Relations Declaration (above n 14).

51 JD Ohlin, 'Did Russian Cyber-Interference in the 2016 Election Violate International Law?' (2017) 95 Texas Law Review 1579; JD Ohlin, Election Interference International Law and the Future of Democracy (CUP 2020). See also H Lahmann, 'Information Operations and the Question of Illegitimate Interference under International Law' (2020) 53(2) Israel Law Review 189–224.

52 A Cassese, Self-Determination of Peoples: A Legal Appraisal (CUP 1995).

53 Secession of Quebec [1998] 2 SCR 217 (Supreme Court of Canada); International Court of Justice, Accordance with International Law of the Unilateral Declaration of Independence in Respect of Kosovo (Advisory Opinion) [2010] ICJ Rep 403.

54 Ohlin, Election Interference (n 51) 90-117.

55 ibid.

56 R Mueller, Report on the Investigation Into Russian Interference in the 2016 Presidential Election (Volume 1) (US Department of Justice 2019) 14-35; US v Internet Research Agency et al No 1:18-cr-32-DLF (D DC, 16 February 2018).

57 This analysis has become increasingly complex since some domestic actors now spread rumors and mistruths online and operations from abroad help amplify these internal efforts.

58 Lahmann, 'Information Operations…' (n 51) 216.

59 N Tsagourias, 'Electoral Cyber Interference, Self-Determination, and the Principle of Non-Intervention in Cyberspace' in D Broeders and B van den Berg (eds), Governing Cyberspace: Behaviour, Power, and Diplomacy (Rowman & Littlefield 2020) 51.

60 Ohlin, 'Russian Cyber-Interference' (n 51) 1593.

61 UNGA, Friendly Relations Declaration (n 14).

62 SJ Shackelford, 'From Nuclear War to Net War: Analogizing Cyber Attacks in International Law' (2009) 27 Berkeley J Intl L 191. It is usefully pointed out here: "The laws of war require states attacking another state to identify themselves" 232-3.

63 International Law Commission, Draft Articles on Responsibility of States for Internationally Wrongful Acts, with Commentaries (2001), UN Doc A/56/10, 47-48; see also Nicaragua (n 16), para 115.

64 D Tran, 'The Law of Attribution: Rules for Attributing the Source of a Cyber-Attack' (2018) 20 Yale Journal Law & Tech 376; P Margulies, 'Sovereignty and Cyber Attacks: Technology's Challenge to the Law of State Responsibility' (2013) 14 Melbourne J Intl L 496.

65 Tallinn Manual 2.0 (n 20) Rule 15, 87-92.

66 UN GGE, Report on Developments… (n 5-8); UN OEWG, Final Substantive Reports (n 9-10).

67 Tran, 'The Law of Attribution' (n 63) 383-391; JS Davis et al., Stateless Attribution: Toward International Accountability in Cyberspace (RAND Corporation, 2017) 9-24

<https://www.rand.org/pubs/research_reports/RR2081.html> accessed Dec 2024.

68 See C Marsh 'The Grey Zone & Hybrid Conflict: A Conceptual Introduction' in Sari and Regan (n 48) 31-44.

69 A Kuehn, D Decker, and K Rauhut, Whodunit in Cyberspace: The Rocky Road from Attribution to Accountability (2023) ORF America Background Paper No 18.

70 K Eichensehr, 'The Law and Politics of Cyberattack Attribution' (2020) 67 UCLA L Rev 520.

71 ibid 586-7.

72 For a proposal to create a Multilateral Cyber Adjudication and Attribution Council, see J Healey et al., Confidence-Building Measures in Cyberspace (Atlantic Council, 2014) 10-12 <https://www.files.ethz.ch/isn/185487/Confidence-Building_Measures_in_Cyberspace.pdf> accessed 16 Dec 2024; Davis et al., Stateless Attribution (n 67) 29.

73 K Eichensehr, 'Decentralized Cyberattack Attribution' (2019) 113 AJIL Unbound 213, 216; Kuehn, Decker, and Rauhut, Whodunit in Cyberspace (n 69) 5-6.

74 N Tsagourias, 'Cyber Attacks, Self-Defence and the Problem of Attribution' (2012) 17(2) J Conflict & Security L 229, 242.

75 US claims against Great Britain for breaching its neutrality during the American Civil War by allowing Confederate warships, like the Alabama, to be built and operated from its territory, resulting in an award of $15.5 million in damages to the United States, Alabama Claims Arbitration (United States v Great Britain) (1872) 29 RIAA 125 <https://legal.un.org/riaa/cases/vol_XXIX/125-134.pdf> accessed Dec 2024.

76 International Court of Justice, Corfu Channel (United Kingdom v Albania) (Merits) [1949] ICJ Rep 4.

77 See eg, Convention on Biological Diversity (adopted 5 June 1992, entered into force 29 December 1993) 1760 UNTS 79, art 3; United Nations Framework Convention on Climate Change (adopted 9 May 1992, entered into force 21 March 1994) 1771 UNTS 107, art 2; **Convention on Long-Range Transboundary Air Pollution** (adopted 13 November 1979, entered into force 16 March 1983) 1302 UNTS 217, art 2.

78 Tallinn Manual 2.0 (n 20) Rule 6, 30-50.

79 ET Jensen and S Watts, 'A Cyber Duty of Due Diligence: Gentle Civilizer or Crude Destabilizer?' (2017) 95 Tex L Rev 1555, 1565-8.

80 ibid.

81 H Lahmann, Unilateral Remedies to Cyber Operations: Self-Defence, Countermeasures, Necessity, and the Question of Attribution (CUP 2020) ch 5.

82 A Coco and T de Souza Dias, 'Cyber Due Diligence: A Patchwork of Protective Obligations in International Law' (2021) 32 EJIL 771.

THE GENEVA ACADEMY

The Geneva Academy provides post-graduate education, conducts academic legal research and policy studies, and organizes training courses and expert meetings. We concentrate on branches of international law that relate to situations of armed conflict, protracted violence, and protection of human rights.

DISCLAIMER

The Geneva Academy of International Humanitarian Law and Human Rights is an independent academic centre. Our publications seek to provide insights, analysis and recommendations, based on open and primary sources, to policymakers, researchers, media, the private sector and the interested public. The designations and presentation of materials used, including their respective citations, do not imply the expression of any opinion on the part of the Geneva Academy concerning the legal status of any country, territory, or area or of its authorities, or concerning the delimitation of its boundaries. The views expressed in this publication represent those of the authors and not necessarily those of the Geneva Academy, its donors, parent institutions, the board or those who have provided input or participated in peer review. The Geneva Academy welcomes the consideration of a wide range of perspectives in pursuing a well-informed debate on critical policies, issues and developments in international human rights and humanitarian law.

The Geneva Academy of International Humanitarian Law and Human Rights

Villa Moynier
Rue de Lausanne 120B
CP 1063 - 1211 Geneva 1 - Switzerland
Phone: +41 (22) 908 44 83
Email: info@geneva-academy.ch
www.geneva-academy.ch