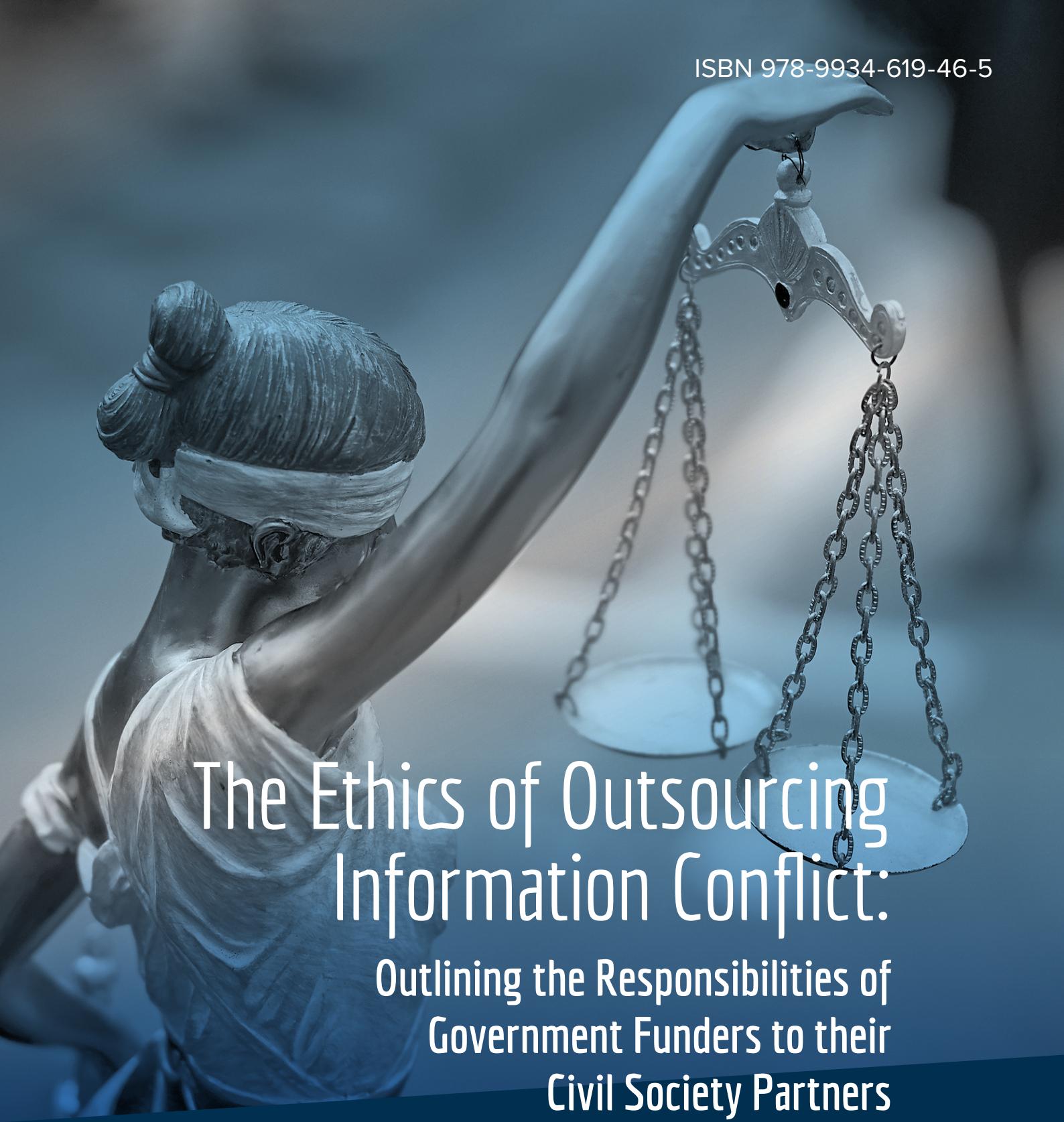
The Ethics of Outsourcing Information Conflict: Outlining the Responsibilities of Government Funders to their Civil Society Partners

Pamment, James; Ahonen, Anneli

2023

Document Version:
Publisher's PDF, also known as Version of record

ISBN 978-9934-619-46-5



The Ethics of Outsourcing Information Conflict:

Outlining the Responsibilities of Government Funders to their Civil Society Partners

PREPARED AND PUBLISHED BY THE

NATO STRATEGIC COMMUNICATIONS CENTRE OF EXCELLENCE

ISBN: 978-9934-619-46-5

Authors: James Pamment, Anneli Ahonen
Project Manager: Tomass Pildegovičs
Content Editor: Monika Hanley
Design: Inga Ropša

Riga, August 2023

NATO STRATCOM COE
11b Kalnciema iela,
Riga, LV1048, Latvia
stratcomcoe.org
@stratcomcoe

The Ethics of Outsourcing Information Conflict:

Outlining the Responsibilities of Government Funders to their Civil Society Partners

Contents

Introduction

Non-governmental organisations (NGOs), researchers, journalists, and the private sector are often the main actors actively countering disinformation and influence operations. While governments maintain some counter-disinformation capabilities, they tend to outsource much of the day-to-day work through, for example, programmatic funding. It is more cost effective and credible to fund independent, non-partisan NGOs to debunk disinformation than for a government to get caught up in trying to correct the sensitive issues that disinformation often entails. Indeed, in a recent *Washington Post* op-ed, former Radio Free Europe/Radio Liberty President Thomas Kent lauds the role of NGOs in countering disinformation and urges governments to keep finding funding for them: "Volunteer activists fight it out with trolls online, penetrate and disrupt conspiracy chat rooms, campaign for companies to stop advertising on disinformation sites, and post memes ridiculing Russian propaganda."1 They do all the things, in other words, that governments can't or won't do at scale.

Similarly, journalists play a central role in exposing and countering disinformation, though they are often not directly backed by governments.[2] Nor are private sector actors such as intelligence firms and social media companies, although the relationship can be close. For the purposes of this report, outsourced (operators/agencies) is then best defined as NGOs, researchers, and other actors who receive direct tasking from a government, as opposed to those who participate in countering disinformation for other (personal, commercial, or ideological) motivations.

This report investigates the roles and responsibilities governments assume when they collaborate in areas of information conflict. In particular, it assesses the risks to civil society and the private sector when they engage in countering hostile foreign influence operations with funding from governments. What are governments' options and limitations when supporting civilian populations to counter information attacks? To what extent can and should governments outsource these activities? And what are governments' responsibilities to civil society and the private sector if and when they come under attack by hostile actors?

Conceptually, this report contributes to the field by developing the term *information conflict*. While disinformation and influence operations are typically the preferred vocabulary for this policy area, both lack a sense of adversarial interaction that characterises the operational realities of countering influence operations. Information conflict is used to reposition actors engaged in countering disinformation and influence operations as

participants in *adversarial contestation* over questions such as asserting matters of fact and truth, determining the legitimacy of public influence methods, and in the ability to take effective countermeasures.

In practical terms, this report contributes to the field by mapping out a range of adversarial measures that can be taken against non-governmental actors who directly or indirectly support the objectives of governments in information conflict. This is not a question of connecting influence operations to methods of countering them. It is about the attacks organisations face for participating in information conflict, often designed to remove them from the disinformation-countermeasure dynamic. It reflects the practical realities of adversarial contestation as it is faced by civil society and the private sector when they take responsibility for engaging adversaries head on.

In addressing several areas of information conflict targeting civilian activities, this report maps out some of the most serious risks and makes recommendations for improving the ways in which civil society is protected. This includes better understanding of vulnerabilities such as legal and regulatory measures, application of terms of service on tech platforms, hack and leak attacks, political and reputational attacks, and harassment of individuals. There are at present no international norms or standards governing the responsibilities of governments over the organisations they fund or support in information conflict; this report may be seen as the start of a conversation about what best practice could and should look like.

The concept of information conflict

In recent years, a number of concepts have come to define aspects of the contemporary struggle over information. *Disinformation* and *influence operations* – to name but two commonly used terms – characterise the sense of a so-called information disorder that has been widely debated since Russia's 2016 intervention in the US Presidential Election.[3] However, the term *information conflict* has by-and-large vanished from the agenda. This is unfortunate since it possesses a unique conceptualisation that other terms do not sufficiently cover. Rather than emphasising hostile campaign content, tactics, and intentions, information conflict encapsulates the ongoing struggle between two or more actors over information and information systems. This missing perspective can help to refocus debates on different problems to other terms.

NATO defines *disinformation* as the 'deliberate creation and dissemination of false and/or manipulated information with the intent to deceive and/or mislead'.[4] As other studies have noted, disinformation is often used as a synonym for two closely related terms: misinformation and malinformation.[5] Misinformation refers to verifiably false information that is shared without an intent to mislead, whereas malinformation refers to true or partially true information that is twisted or taken out of context to support false, and potentially harmful, interpretations.[6] This group of terms can be characterised by an emphasis on information content. Analysis of information content is focused on why the messages were produced (intent) and whether they are truthful.

Influence operations refer to efforts to influence democratic processes using illegitimate, but not necessarily illegal, methods to the benefit of an often hidden or disguised set of interests. Whereas disinformation emphasises factualness and intent in the content of messages, influence operations cover the often covert and manipulative communication techniques that make up a coordinated effort to influence a society.[7] It is a concept favoured by social media platforms. Meta, for example, defines influence operations as 'coordinated efforts to manipulate or corrupt public debate for a strategic goal."[8] Analyses once again focus on how the campaigns are constructed, with more emphasis on the techniques and behaviours used by an adversary than on content per se.

Information conflict has its origins in international law dating back to the 1990s. In this context, it refers broadly to "a state's rights and responsibilities when it conducts operations that affect another state's information or information systems."[9] Traditional applications situate it at what has been termed the grey zone between peace and war; indeed, use of information *conflict*, as opposed to information *warfare*, is intended to emphasise that such operations are not necessarily governed by the rules of war, and that they are regularly deployed outside of the battlefield. Inspired particularly by the development of cyberspace as a domain of contestation in both civilian and military sectors, debates about information conflict foreshadow present discourses of "grey zone" hybrid, cyber, and influence operations.

Though it could hardly be considered a widely used term even during the 1990s and early 2000s, information conflict appears to have been used in North American legal contracts to cover issues such as conflicting data sources held by two or more databases, or differences between two or more parties concerning what they consider to be correct information.[10] The term has also been applied to wikis (online information sources such as Wikipedia which are updated by communities of users). One academic paper characterises inconsistent and opposing updates to wikis as a form of information conflict.[11] A sense of active contestation between actors over the

validity of their preferred information sources and content permeates these different uses of the term.

In academic discourse on information warfare, information conflict has been used to refer to the application of information warfare tactics in both military and civilian contexts. Cyberwarfare, command and control warfare (targeting coordination capabilities), intelligence-based warfare, and psychological operations are considered possible techniques of information conflict. For example, an article from 2013 predicts that social media is likely to become a central battlefield for information conflict to be carried out through platform manipulation.[12] This adds a dimension that helps to align the term with influence operations, in the sense that the latter often involve covert and manipulative methods developed initially for the battlefield but frequently deployed in civilian contexts.

Bringing the concept fully up-to-date would suggest that elements of disinformation and influence operations could be better understood through the lens of contestation over information and information systems. The niche occupied by the concept of information conflict within an already crowded terminology could be considered as the sense of *adversarial contestation* between two or more parties. While much of the term's meaning is covered by contemporary understandings of hybrid, cyber, and influence operations, it is worth underscoring that the sense of adversarial contestation in these terms is comparatively poorly explicated. A concept such as information conflict can account for the threats and harassment targeting civil society actors who participate in countering disinformation in a way that the terms *disinformation* and *influence operations* gloss over.

This is because of a tendency to divide the contemporary field into hostile operations contra countermeasures. Antagonists are often referred to as hostile actors, threat actors, or adversaries, whose activities involve manipulation or misuse of information. Those responsible for countermeasures sometimes refer to themselves as "the defender community",[13] on the basis that their activities are intended to protect the integrity of information and information systems. The *conflict* over information *between actors* is glossed over by publications that either talk in forensic detail about how disinformation or influence operations are designed,[14] or through bland statements of policy that explain why some content might be blocked or removed.[15] Rarely is the *information conflict* – the game of cat and mouse between two or more actors over the prominence of information and integrity of information systems – discussed in detail. Takedowns are presented as *fait accompli*; the actual means by which one side won out over the other tends to be glossed over.[16] If we are to better understand the responsibilities of governments to its outsourced operators in the defender community, we need to think about disinformation as something bigger than content, intent, or manipulation. It is also about contestation, including efforts to remove actors in the defender community from the conflict by, for example, threatening, harassing, or discrediting them.

There are many reasons why this problem has been underplayed conceptually. Influence operations are often identified and disrupted using countermeasures that cease to be effective if adversarial actors know about them beforehand. For example, if a social media platform has an automated system that triggers an alert if more than 50 accounts are created within a day from the same IP address, a motivated adversary would stick to creating 49 accounts every 24 hours. Sensitivities mean that both defenders and adversaries wish to keep their methods a secret, and only expose what they know about the methods of the other side if it is in their interests to do so. In other words, research into information conflict is under-represented because its focus falls on a real-time struggle using tradecraft that needs to be protected for operational reasons.[17]

Furthermore, exposés, spurious legal challenges, and harassment are difficult to talk about. They are often seen as an anomaly; an embarrassing one-off event rather than a

systemic issue connected to both information conflict and vulnerabilities in the government-outsourcer relationship. For example, the victim of a hack-and-leak rarely wants to discuss in detail the contents of that leak. Nor is it likely that a government funder would want to be particularly visible while one of its civil society partners is in the midst of a crisis.

The premise of this report is that information conflict is an ongoing phenomenon in which all actors with some involvement in the disinformation and influence operations fields engage. Yet, its actual practices and modalities are under-represented within research. There is a lack of studies of the ongoing conflict that targets the actors themselves rather than solely over the narratives, content, or methods of a campaign. If those actors can be attacked or discredited outside of the boundaries of the campaign, it can be easier for one side to gain an information advantage. What then are some of the main areas in which information conflict takes place? Who is involved, and what happens when they come under attack from an adversary? And what are the ethical consequences of governments outsourcing aspects of information conflict to civilian organisations such as the private sector, nongovernmental sector, and research?

Types of information conflict

For the purposes of this report, it is valuable to first outline the major areas where information conflict occurs in a civilian context, emphasizing that this work is predominantly conducted through collaboration between governments and nongovernmental partners. Three areas of work are particularly worthy of further discussion. The first is the area of information conflict characterised by *contested truths*; including, for example, organised efforts to debunk false or misleading statements that can cause public harm. A second area of information conflict is characterised by *contested influence methods*, such as the systematic analysis and attribution of adversarial behaviour in conjunction with content removal, including tech platform takedowns. A third area of information conflict refers to *contested counter-operations*; for example, activities conducted for or by democratic institutions that are designed to disrupt adversaries' influence capabilities.

Contested truths

The term *contested truths* refers to the battle of narratives, truth, and media content. Several techniques are used to conduct information conflict in this space. Perhaps the clearest and most widely understood tenet of information conflict is the efforts of fact-checkers and debunkers to correct false and misleading information. Fact-checking refers to the long-standing process of verifying that all facts in a piece of writing, news article, speech, etc. are correct. It derives from a need to hold those in power accountable for their claims, and is traditionally conducted by journalists, newsrooms and political analysts. Debunking refers to the process of exposing falseness or showing that something is less important or accurate than it has been portrayed. Debunking is often conducted in order to expose the lies of a specific actor or focuses on a specific area of expertise in order to reduce the harm that some falsehoods can present to the public. It is often funded by governments or intergovernmental organisations, but conducted by researchers, think tanks, and advocacy groups.[18]

Fact-checking and debunking as a tenet of information conflict is relatively straightforward to exemplify. If two or more actors go onto a social media platform and assert information either for or against a vaccine, there is a clear public interest—and, therefore, a governmental responsibility—to intervene in that discussion with the correct information. In other words, governments and interest groups have a mandate to engage in information conflict in order to protect the public from harm. Ideally, the actor intervening should be credible; for example, a doctor or medical researcher can speak more credibly about the pros and cons of a vaccine than a government representative. It is, therefore, common for governments to fund debunking initiatives in areas of strategic importance, particularly where there is a risk of mis- and disinformation causing significant public harm.

At present, there are over 100 specific initiatives around the world that do fact-checking or debunking targeted on antagonistic or harmful mis- and disinformation.[19] Each of them involve some form of purposeful financing for the strategic goal of engaging in information conflict over factualness. For example, in Lithuania, 5,000 volunteers work in a loose coalition to combat and expose false claims made by pro-Kremlin trolls. Known as the Elves, this anti-troll network works to actively contest false content and is funded by the government via an NGO.[20] Many of the organisations involved in asserting the truth in contested spaces will testify to concerted efforts to undermine their reputations in order to discredit them and ultimately remove them from the public sphere.

Media literacy refers to education or training in how to critically interpret media, including social media. For example, the European Digital Media Observatory is a network of universities and think-tanks funded by the EU to build capacity for media literacy initiatives in Europe.[21] Such initiatives perform a public education role, encouraging media users to become more resilient to disinformation.

Pre-bunking is another area of information conflict in which targeted segments of the public are prepared with information about the spread of potentially harmful disinformation so that they can reject the falsehoods before they are exposed to them. For example, during the COVID-19 pandemic, researchers prepared a guide for dealing with vaccine hesitancy, designed to pre-empt likely fears and misgivings about the vaccine with early interventions based on factual arguments.[22] Activities such as these are often conducted by NGOs with government funding on the understanding that they contest mis- and disinformation with facts by strengthening the public's ability to interpret information and information sources.

Content moderators are often challenged in terms of their political allegiances or agendas. Studies have noted that information conflict takes place in news and information sources outside of social media. Wikis, which are updated by communities of users, are increasingly seen as targets for manipulation in a phenomenon known as conflict-of-interest editing.[23] The comment sections of online news sites have been systematically targeted by threat actors who seek to give a false picture of public sentiment.[24] Deep fakes, which refer to use of digital technology to fabricate facial movements and voice, and deep text, which uses machine learning to create automated texts that take on the characteristics of genuine texts, are other areas where falsehoods must be debunked quickly and efficiently in the public interest. In each case, it is credible actors in civil society and the private sector who elect to participate in information conflict on behalf of wider public interest, to correct falsehoods and support the integrity of the information environment.

Contested influence methods

Another area of information conflict is characterised by *contested influence methods*, such as the systematic analysis and attribution of adversarial behaviour. Discovering and exposing manipulative influence techniques is a key area of information conflict. It is often assumed that casting light on covert activities such as influence operations is sufficient to embarrass the source and undermine their ability to manipulate target audiences. Much effort in information conflict is, therefore, dedicated to two aspects of attribution: the technical ability to connect activities to actors, and the strategic ability to expose–with or without supporting evidence–those actors as the sources of the influence operation. A major consideration in information conflict is in protecting the tradecraft that provides the technical basis for an attribution, contra the political value of being able to announce what an analysis has revealed.

When dealing with influence operations on digital media, attribution is often the purview of non-governmental actors such as the digital platform owners. In most cases, the results of attributions are presented to the public without details of how the information conflict was discovered, exposed, and countered. Announcements and reports simply offer examples of why the adversarial behaviour was in violation of platform terms of service so as to justify content removal. For example, the Disinfodex[25] database houses over 360 disclosures made by tech companies about threat actors –state or non-state–who have conducted influence operations on their platforms. There is no information about how the adversarial actors were identified; only, in some cases, a name. This fits a broad pattern in which attributions tend to be presented by tech companies as simple statements of fact, in order to protect tradecraft.[26]

Although tech platforms and government intelligence agencies may share some information, particularly when it comes to criminal activity or espionage, it is ultimately the tech companies that finance and oversee takedowns and attributions based on their profit-oriented goals as corporations. Their mandates, therefore, rest somewhere between self-interest and the public interest, with different companies applying those factors to different degrees. In such cases, responsibility is not so much outsourced by governments to the private sector as fully owned by the private sector.

Exposure of influence operations on tech platforms is, in some cases, supported by investigations and reports by third parties who were either tipped off by the social media companies, or who discovered the initial abuse and are granted the opportunity to further investigate. In these reports, researchers and analysts, usually based at think-tanks, universities, news media, or in private sector intelligence or communication companies, are granted limited access to some examples of the influence operations shortly before the content is removed.

The resulting analysis offers some insights into the techniques used by the adversary on a given tech platform, but often disproportionately focuses on clear-cut examples of problematic content rather than revealing the technical details that the platforms used to make their assessments.[27] Furthermore, many of these independent reports are carried out by organisations that have signed legal agreements such as an NDA with the tech platforms.[28] In exchange for limited access to data, independent organisations must stick to the scope of investigation agreed with the platform, which in essence means agreeing with the platform's assessment. This has the effect of juxtaposing the public interest mandate of independent researchers with a legal framework tied to the commercial and self-preservation objectives of the tech company. In such instances, the information conflict is not just between threat actors and tech platforms, but also between the tech platforms and public interest research.

Some research takes place independently of both tech platforms and governments. So-called open-source intelligence (OSINT) involves combining investigative journalism with intelligence tradecraft to gather information. Following the Salisbury poisonings, Bellingcat made use of publicly available, leaked, and declassified intelligence sources to identify the alleged perpetrators.[29] The resulting reports are some of the most detailed in terms of information conflict, since they often go into forensic detail about adversary tactics, which countermeasures or investigative methods helped to expose them, and what evidence was used.[30] Bellingcat does not accept direct funding from governments but receives support from a number of foundations and intergovernmental organisations, as well as individual donations and in-kind support.[31]

The Mueller Report on Russian interference in the 2016 US Presidential Election[32] is a rare exception to analyses of influence operations that limit their analysis of information conflict. Conducted by the Special Council's legal team in conjunction with independent researchers, forensic analysis of how Russia and its proxies interfered in the election is laid out in full. Rather than simply presenting the influence operation as though it were conducted in a vacuum, the report had the mandate to explain how Russian activities circumvented and exploited existing measures, thereby revealing the methods used by all parties in the information conflict. Many operational details were, however, redacted to protect tradecraft or ongoing investigations.

While the Mueller Report remains one of the most important examples of information conflict available to researchers and analysts, the circumstances of its production mean that it is unlikely to become a regular process for gaining insight into information conflict. For example, the French government elected not to publish a public report into its experiences of election interference in 2017.[33]

Contested counter-operations

Another area of information conflict refers to *contested counter-operations*, including activities designed to disrupt adversaries' influence capabilities. Some countermeasures targeting influence operations take the form of proactive campaigns that are designed to assert counter-messaging, counter-narratives, or seek to diminish the brands of adversaries.[34] Often, there is a tendency for analyses to focus on the proactive campaign and its objectives rather than how it is contested in practice. For example, the European External Action Service's EUvsDisinfo[35] exposes and debunks disinformation as part of a public campaign to raise awareness and push back against false Russian narratives spread within the EU. Following a scandal in which three Dutch media companies sued the EEAS for claiming that they spread disinformation, some articles were removed from the database.[36] It could be argued that the political mandate of such a campaign is decisive in determining the level of political support it receives when it becomes heavily embroiled in information conflict; in this case, external political pressure to shut the campaign down.

Sometimes, countermeasures draw on legal and regulatory support. In March 2022, the EU imposed sanctions[37] on Russian state media, including RT and Sputnik, in response to disinformation spread about Ukraine prior to and during the invasion. A report by the NATO Strategic Communications Centre of Excellence demonstrates how the Kremlin continually changed tactics to stay ahead of the tech platforms' efforts to restrict access to Russian state media.[38] A recent academic article produced by a coalition of civil society actors that was monitoring implementation of the sanctions on behalf of the EU Commission published interim results that demonstrate how the cat-and-mouse game was played by the Kremlin to circumvent various enforcement measures.[39] As a form of counter-operation that also engages with attribution, sanctions provide a means of assessing information conflict, in the sense that their application becomes a clear, legally-mandated site of contestation.

Other counter-operations draw upon secret intelligence to influence the calculus of a hostile state actor. For example, prior to the Russian invasion of Ukraine in February 2022, the US and UK released intelligence suggesting that Russia intended to use so-called false flag operations as a pretext for war.[40] Exposing these plans pre-emptively reduced Russia's opportunities for using such pretexts, which, in essence, captures a moment of contestation between two parties over information dominance. Although principally the work of governments, Bellingcat was, for example, involved in collecting and analysing open-source data that corroborated the intelligence assessments.[41] Other civil society actors, such as the Centre for Information Resilience[42] and the Atlantic Council[43] have produced reports on Russian activities in Ukraine that leveraged open-source intelligence in order to counter Russia's false narratives about war crimes.

At the harder end of counter-operations are countermeasures which draw on clandestine or covert methods, known as information operations, psychological operations, or more recently, cognitive warfare.[44] Many armed forces possess the capability to run operations targeting adversaries in conflict zones and civilian targets if they are mandated to do so. For example, during the 2018 midterm elections, the US allegedly disrupted the internet access of the notorious St. Petersburg troll farm behind the 2016 election interference, the Internet Research Agency.[45] In some cases, these operations are run by experts within the military. In others, there may be some level of outsourcing to private sector communication companies who may have staff with security clearances and typically military or intelligence backgrounds. Their work in information conflict is closer to the design and process of military operations but is notably conducted at a level removed from government.

Challenges for the defender community

These areas are, of course, not exhaustive, but give a sense of where, how, and on what basis information conflict takes place in the disinformation policy area. Three challenges for the defender community stand out as significant. The first is that the mandate for participation in information conflict derives from the moral responsibility of liberal western democratic governments to protect the public and the integrity of public debate. In some cases, it also involves protection from harm. This moral responsibility is, in democracies, often shared with civil society actors who work in the public interest. In some cases, the mandate has to negotiate with commercial objectives, which adds a layer of complication to the equation; while tech platforms may ideally have an ethical responsibility to their users, this is often weighed up against commercial interests.

Second, the field has grown comfortable avoiding talk of information conflict, usually to protect tradecraft or the details of the way actors organise themselves. This creates tensions over the quality and validity of investigations, as well as between transparency and secrecy. It is often a hard choice to decide whether to promote an organisation's great work in countering influence operations versus protecting them from being targeted as fallout from the information conflict. In simple terms, we do not talk enough about information conflict in order to establish shared norms because it often seems easier for all parties involved not to talk about it at all.

The third point is that the field has evolved in such a way that most analysis glosses over information conflict. The position of influence operations in relation to cyber and hybrid threats is often simplified or glossed over. Rather, information conflict is perceived as the unfortunate by-product of angry people who go too far on social media. Nothing could be further from the truth. Information conflict is part of the game, and it involves harsh, often distasteful measures against actors perceived as vulnerable in order to remove them from the playing field. The key questions should therefore be what risks do these structures pose to those who work within them, and what responsibilities do governments and other funders have when a hostile actor retaliates?

Retaliation in information conflict

When conducted by democratic countries, countering disinformation is rarely limited to one actor alone. Invariably, there is some level of collaboration between governments, the private sector, civil society, and research to strengthen the efforts. Often, but not always, a mixture of governments, the private sector, and public interest foundations fund the activities. This gives the impression that information conflict is sometimes outsourced, for example by governments to civil society, or by tech platforms to universities. A key question for this report is, therefore, what level of protection does outsourced information conflict deserve from its funders if and when they are specifically targeted by threat actors with, for example, cyber or intelligence capabilities.

Funding an NGO to participate, or which is participating, in information conflict risks making it appear as a legitimate target to an adversary. If that adversary has state-backed influence, cyber, or intelligence capabilities, the risks are considerable. The purpose of this section is to offer some examples of when participants in information conflict have become targets of adversaries either because of what they were doing, or because of who they were funded by. It raises questions about what levels of protection outsourced information conflict participants should expect from funders, and, more specifically, what the role of government is in protecting democratic societies when motivated and capable adversaries seek to remove civil society actors from the information environment.

Legal, diplomatic & regulatory measures

Reciprocity refers to the idea that punitive actions, such as sanctions, can receive an in-kind response, even if those caught up in the response are not involved in any decisions or are not directly connected to the issue.

- A month before Russia's invasion of Ukraine in early 2022, Germany's independent media regulator ruled that RT DE was operating in Europe without a valid broadcasting permit. The Kremlin immediately took retaliatory measures against Germany's international public service broadcaster Deutsche Welle, closing its Russian offices and revoking its operating licences, which were valid until 2025 and 2027.[46] The distinction is significant since Russia revoked DW's permit as a political response to an independent regulatory decision, thereby intentionally positioning DW as a casualty of an information conflict between Germany and Russia over the right to broadcast in each other's jurisdictions.

- When Russian state media was sanctioned in Europe following the invasion of Ukraine, Western social media platforms were pressured by the Kremlin to censor material that referred to their "special operation" as a war or invasion. Some Western tech platforms, including Instagram, LinkedIn, and Twitter were either heavily throttled within the Russian Federation or blocked entirely because of their compliance with the EU's sanctions regime.[47]

In these examples, laws and regulations are used as a pretext to punish actors who are ultimately not responsible for the broader political conflict, and who are independent from government. They are punished because of their perceived role in information conflict. For example, in most democracies, public service broadcasters are considered

separate from government direction, unlike autocratic state-media. Likewise, tech companies act independently of the government of the country in which their headquarters are based, at least in liberal western democracies.

In addition to legal and regulatory measures designed to single-out and damage independent actors for their potential role in information conflict by nation states, some regulations are designed to protect researchers as well as the public but can be used against the spirit in which they are intended.

- EU Disinfolab, an independent Brussels-based NGO which analyses disinformation, came under scrutiny in 2018 when accounts identified as spreading pro-Kremlin disinformation were identifiable in one of their publications despite efforts to anonymise them. Some of the account owners reported the non-partisan civil society organisation to the French Data Protection Authority in an effort to discredit the findings of the research.[48] These regulatory processes provide valuable protections to all parties but can be used to distract from the findings of research by seeking to politicise, invalidate, or censor it.

- Use of freedom of information requests is a well-established method of turning an important regulatory mechanism against actors in the public sector, including government and research. The Election Integrity Partnership, a collaboration between academia, non-profits, government, and social media platforms detailed so-called super spreaders of false claims about election processes during the 2020 US Presidential Election.[49] One prominent researcher in the Partnership claimed that some FOIA requests made by political interest groups would take years to fulfil. "They have now discovered they can weaponize our transparency laws to harass my colleagues and me (at a public university)."[50] Similar experiences have been reported by government departments and agencies working with disinformation, who sometimes receive freedom of information requests at a rate that would permanently tie up their entire staff.

The weaponisation or exploitation of legal and regulatory measures, which can be used against public office activities, can also present a significant risk to nongovernmental actors who participate in information conflict. For some organisations, such as public sector broadcasters or think tanks with offices in authoritarian countries, it would be valuable to prepare risk assessments of the grounds upon which a hostile actor could invoke false reciprocity. While some organizations already conduct risk assessments, it is possible that they may not specifically prioritize threat vectors arising from information conflict linked to disinformation and its associated methods.

Likewise, NGOs should be interested in which laws and regulations could be twisted as a form of "lawfare" against them. Sudden labelling as a foreign agent or representatives of a foreign actor, for example, might present significant risks to local staff in those countries. Mass reporting via freedom of information requests, GDPR, ethics boards, and other technical regulatory issues around data handling are all capable of impacting an NGO's work. Legal provisions for good faith research exist in most democracies, yet they seem to not always protect organisations when they need it most.

Overall, the main point is that any civil society or private sector actor is at risk of not only undeclared, covert, and discreet attacks as part of the information conflict, but also legal and/or regulatory interventions from hostile actors. The challenge then is to better understand which information conflicts an actor is deliberately (from your side) or perceived (from their side) to be participating in. Which of these measures within your own country, in the countries where you are active, and in hostile states, can be taken against you?

Terms of Service

Tech platforms have the ability to enforce their terms of service based on the norms and values they wish to assert on their platforms. This sometimes puts them in the firing line.

- In December 2020, Meta announced the removal of over 100 Facebook and Instagram assets that had been attributed to "individuals associated with the French military."[51] The operation targeted Francophone Africa and appears to have been designed to push back on Russian influence operations in the region. This was the first time Facebook publicly announced a takedown of coordinated inauthentic behaviour connected to a Western government, and one of the first to acknowledge one or more actors engaging in information conflict over the same target audiences. A similar number of Russian-backed assets were also removed, some of which were connected to Russian oligarch Yevgeni Prigozhin's Wagner Group. In response, Prigozhin posted on social media that Facebook is "nothing more than a tool of US intelligence services," adding that he had never used it personally.[52] The response, in other words, legitimised further attacks on Meta as a participant in information conflict by conflating it with US intelligence.

Other Western countries were subject to similar censure by tech platforms in August 2022.

- Twitter and Meta announced the removal of over 150 assets that had been active for several years, and that had their origins in the US and UK. The accounts mimicked the influence operations methods used by the Kremlin.[53] Follow-up reporting claimed that US Central Command was "facing scrutiny" following the announcement and that an internal audit of its covert information warfare activities had begun. The Washington Post claimed that other covert campaigns conducted by the US military had in fact been removed previously by Facebook, and that Facebook had on more than one occasion warned US officials that their efforts were too easy to detect.[54]

In this case, a private sector company based in a given country is put in a challenging position vis-à-vis how it applies its terms of service against a government with which it is simultaneously negotiating over a host of regulatory, financial, and policy matters.

- In late 2021, RT DE was given a one-week restriction from uploading new videos on YouTube for violating its COVID-19 misinformation policy. RT then tried to circumvent the punishment by using a second, connected channel to disseminate more misinformation. When they were caught, both channels were banned, which led to an immediate response by the Russian Foreign Ministry. They accused YouTube of "unprecedented information aggression" against Russia and threatened "retaliatory measures" against both YouTube and German media. As a private company owned by Google, it is unlikely that they took instruction from the German government or acted in collusion with German media, though the Kremlin's response indicated that YouTube's invocation of its terms of service would be interpreted as an act of aggression with state backing.[55]

While tech platforms are sometimes the victims of information conflict when they act in the public interest, at other times their commercial mandates position them as adversaries against independent researchers. Terms of service have been used by tech platforms to restrict independent research into influence operations on the platforms.

- Following the Cambridge Analytica scandal,[56] Facebook sought to improve data security and reduce breaches, including restricting apps that facilitated opt-in data donated by users for legitimate academic research purposes.[57] The Netvizz application,[58] a research tool created at Amsterdam University and used by hundreds of academics around the world, was banned in 2018 (there are currently nearly 2,000 academic citations of data produced by the tool).[59] A joint statement signed by hundreds of scholars argued that the changes were more about strengthening Facebook's control over research about the platform than protecting users. It argued, "the platform providers ... cannot be allowed to position themselves as the gatekeepers for the research that investigates how their platforms are used."[60] However, Facebook's terms of service provided legal protection to do just that; to control who was able to analyse information conflict on their platforms.

- In August 2021, Meta removed the open-source browser extension NYU Ad Observatory in a further example of a tech company censoring academic research that attempts to be in some way independent of the reports and tools Facebook provides.[61] Launched in 2017 as ProPublica's *Facebook Political Ad Collector*, the extension allowed Facebook users to voluntarily submit anonymised data showing which advertisements targeted their user profiles.[62] Following legal threats,[63] backend changes designed to break the tool,[64] and some public backtracking,[65] Facebook finally banned the extension and suspended accounts shortly after the lead researchers notified that they would study the January 6 riots.[66] Facebook blamed the Federal Trade Commission,[67] who in return clarified that good-faith research in the public interest has protections.[68]

Such incidents demonstrate that terms of service are powerful tools that can be used to achieve a variety of objectives within information conflict, including removing content, banning accounts, cataloguing issues subject to information conflict, exposing influence operations and their methods, and attributing actors. Tech platforms can be placed under massive pressure by governments— friendly and unfriendly—to turn a blind eye to their information operations and to expose only those conducted by others. In turn, researchers can be placed under tremendous pressure by tech platforms, which may, for example, restrict access to API's or analysis tools such as CrowdTangle that can be essential to an organisation's livelihood.

Tech platforms should not be able to decide who gets to conduct research using their data, just as democratic governments should not be able to stop tech platforms from exposing and removing their information operations. In both cases, governments have a responsibility to guarantee protections to the actors; from censure for removing problematic content even if it is yours, and protections of researchers to be able to fully investigate tech platforms and their data. There are currently important efforts to ensure that platform regulation takes this into account. For example, the EU's Digital Services Act promises access to data from very large online platforms and very large online search engines to vetted researchers as one of its key components[69]. The act has not yet been implemented in practice and public feedback is still being collected on how independent audits should be conducted[70].

Cyberattacks

As part of the information conflict, adversaries may sense an opportunity to remove an actor from the conflict by, for example, attempting to discredit them through a hack and leak.

- Scottish charity, The Institute for Statecraft, created an international civil society network in 2015 called Integrity Initiative, which was designed to counter Russian disinformation in multiple countries. In late 2018, the Institute was hacked by the hacker group Anonymous, though it was later revealed that hackers connected to the Russian state were likely posing as Anonymous.[71] Russian state media, as well as other Kremlin-friendly actors, used the seven tranches of released documents to suggest a nefarious conspiracy organised by current and former intelligence officers.[72] The UK had backed the network with over £2 million[73] and the leaked documents suggested that other countries, NATO, and Facebook, among others, had made financial or in-kind contributions to the network.[74] While the outrage from Russian state media sought to aggressively discredit individuals named in the planning documents, the budding political scandal within the UK centred on a handful of the Institute's politically-motivated tweets, including a retweet of a story that claimed then-Labour leader Jeremy Corbyn was a "useful idiot" for the Kremlin.[75]

Clear from these efforts to discredit the networks is the sense that the NGOs were considered active players in information conflict, and therefore legitimate targets for Kremlin-backed hackers. Once emails and internal documents had been exfiltrated, they could be used to support spurious narratives. Discretion around the NGO networks and their activities simply added justification to the idea that their work was secret, and hence linked to intelligence.

- EU Disinfolab was subjected to a cyberattack which took its website offline in July 2020. The attack was allegedly connected to the Solarwinds hack and involved a successful phishing attempt.[76] According to a statement by the NGO, "When a Russian cyberattack targeted us, only US actors helped us [...] FireEye helped us for free for days when we really needed help (tremendous help). Meanwhile, we did not get any support from any EU-based organisations."[77]

This points to a significant problem for nongovernmental actors involved in information conflict. While it is reasonable for them to deal with many risks associated with their work, they cannot be expected to deal alone with a hostile foreign state's intelligence and cyber capabilities. Here there is a reasonable expectation for governmental support and protection. In the case of Integrity Initiative, the connection to the UK government's Russia work is what made the network of interest to Kremlin hackers. Some degree of protection from the Kremlin's advanced cyber capabilities would seem like a reasonable starting point for all contracts made in such a risky area.

What is also clear from these examples is that organisations' state backing can be easily and effectively exploited as a vulnerability and discrediting factor by hostile states. In general, only a handful of government funders provide workshops and training on security issues for their contractors. Overall, there appears to be a lack of monitoring of threats targeting funded organisations and contractors, and hence a lack of protection. Based on our interviews with NGOs working in this area, more could be done to raise the baseline competence of grantees to work in such sensitive areas; for example, with training support for operational security, cyber security, crisis communication, and physical security.

Political & reputational attacks

As in the case of the hack and leak attack on Integrity Initiative, an objective of information conflict can be to politicise aspects of an actor so that political partisanship comes into play.

- In the case of Integrity Initiative, this related to some political tweets made by the hosting organisation entirely unrelated to the project. The politically motivated tweets eventually motivated a Parliamentary question from a representative of the Labour Party.[78] Later on during the height of the Pandemic, the UK announced that its PsyOps brigade, the 77th, had supported governmental efforts to counter COVID-19 disinformation, albeit with a focus on foreign actors.[79] Allegedly, a Scottish politician accused the 77th of working alongside Integrity Initiative to target domestic Scottish audiences. By extension, this implied a conspiratorial connection between British military PsyOps, Integrity Initiative, and suppression of the Scottish independence movement. The tweet was later deleted.[80]

- Ukrainian NGO, StopFake, has been falsely characterised as a neo-Nazi organisation in pro-Kremlin disinformation, and at one point, these efforts almost led to a parliamentary inquiry into their fact checking activities.[81] Rumours of connections to foreign intelligence agencies were spread to cast doubt on their work. DDOS attacks were directed at the website to disrupt its work, and some staff were targeted on social media after public appearances. Ultimately, StopFake's protection from the political criticism is believed to have come from the content of their fact checking, which is politically neutral and meets international standards. As an already well-established fact-checker with six years of experience, StopFake joined the International Fact-Checking Network (IFCN) in 2020 and follows its Code of Principles. According to Editor-in-Chief, Yevhen Fedchenko, "We are now also financially sustainable and don't depend on donors, so casting a shadow on us to disrupt our international partnerships does not work."[82] Currently, StopFake funds its work through commercial clients, training, and consultancy to cover its fact-checking operations. Apart from the Ukrainian website, it has websites in 13 different languages, which are covered through a grant from the National Democratic Institute.

Other political criticisms are borne of coincidence, mistakes, or political opportunism.

- In September 2020, Canadian military personnel conducting psychological operations training produced a fake letter from the Nova Scotia government about a pack of grey wolves running loose in the area. In conjunction with the letter, they practised generating wolf sounds through a loudspeaker. A copy of the training material, which was not marked as such, was found by a resident and quickly shared within the local community. Local government debunked the letter and the military admitted it was their error.[83] While some media coverage emphasised the humorous nature of the mistake, other coverage observed that testing of PsyOps on civilian targets should be considered unethical.[84] This linked to other recent debates in Canada about military communications in general, and particularly which capabilities should or

shouldn't be used within the country, for example, for public affairs purposes, social media monitoring, or for analysing COVID-19-related misinformation.[85] While the mistake was by all accounts innocent, it encroached on a delicate area of domestic political controversy.

- EUvsDisinfo is a website and campaign that has the objective to "increase public awareness and understanding of the Kremlin's disinformation operations."[86] Since 2015, it has been run by the East Stratcom Task Force within the European External Action Service, a team whose mandate is to address Russia's ongoing disinformation campaigns. The team was originally created as part of the EU's measures to respond to Russia's aggression in Ukraine in 2014.[87] Naturally, the Kremlin has sought to discredit it. In 2015, Russia's representative to the EU claimed that the project was conducted by "ideological special forces" whose job was to spread Russophobia and anti-Russian myths.[88] Russian foreign minister Sergey Lavrov referred to this and the European Centre of Excellence for Countering Hybrid Threats as "reminiscent of a hunt for dissidents and is unlikely to help restore confidence."[89]

- In 2018, EUvsDisinfo was drawn into a political controversy when three Dutch media outlets took it to court for labelling their articles as pro-Kremlin disinformation. As a result, EUvsDisinfo removed the articles from its database[90] and the outlets dropped the court case. Demands to shut down EUvsDisinfo followed.[91] The resulting political debates eventually reached the settlement that EUvsDisinfo would not include European outlets in its database of pro-Kremlin disinformation, which, it could be argued, significantly weakened its ability to fulfil the original mandate.[92]

- During the COVID-19 pandemic, the European External Action Service widened its disinformation reporting to cover other relevant foreign actors such as China. Following alleged pressure from the Chinese government to soften a report into its disinformation, information about the internal debates was leaked. Unusual for this type of coverage, the New York Times printed the name of an analyst whose internal emails about the Chinese pressure were leaked.[93] The analyst soon left the EEAS' Stratcom unit as a result. Her treatment was criticised by several members of the European Parliament. For example, MEP Sergey Lagodinsky spoke of "frankly a disgrace for our EU diplomatic service. Mobbing out the one who has uncovered irregularities despite the parliamentarians' concerns. This is not OK."[94]

These examples demonstrate how readily participation in information conflict can be utilised for political ends. In some cases, hostile influence operations might seek to inflame these fears, by identifying genuine political interests or concerns and exacerbating tensions. In others, simple errors can open the door for political opportunism. Reputational or other vulnerabilities in the actors within the defender community can be relatively easily turned into content for an influence operation by a motivated threat actor.

In addition, these examples showcase how carefully and thoroughly any democratic government or international organisation should prepare and craft their defence measures. Most of the Western governments started developing their work in countering influence operations only after Russia provoked a war in Ukraine and illegally annexed Crimea in 2014. New governmental bodies

have been created, for example, in the US, EU, France, and Sweden. The latest attempt to create a "Disinformation Governance Board" in the US failed after it was dragged into political controversy[95]. There has been a lengthy trial-and-error period for developing countermeasures and the partnerships capable of pursuing them.

More work needs to be done in understanding which parts of the work should be classified to increase safety and protection, and where transparency is the best course of action. We argue that an important lesson that can be drawn from this ongoing process of developing a defender community centres on defining mandates. The purpose must be clear enough to ensure that the tasking can be fulfilled, even when placed under opportunistic—or for that matter, systematic—political pressure. For example, if partners in the defender community are mandated to conduct independent work to shape the information environment, how robust is that mandate likely to be if the objectives of the agreement were to be leaked? [96] If independent actors are given licence to conduct covert information operations to disrupt/degrade malign foreign actors, how much support will remain if their efforts are intercepted by a tech platform or adversary?[97]

There needs to be wide political consensus and will to back a mandate and protect the individuals working to implement it. There have to be enough resources reserved to ensure the quality of countermeasures. The public should have access to information about the mandate and governments should be able to communicate clearly why these efforts are needed. What exactly is being monitored by the government and why? How is the information stored and further used? If any of these parts fail, it is much more likely that these vulnerabilities will be used by a hostile actor, ultimately putting nongovernmental partners at risk.

For government-funded civil society organisations, adherence to international standards can offer a measure of protection against politically partisan attacks. Other simple lessons are to avoid any kind of public statements that could be interpreted as political if an organisation receives government funding. Due diligence in this area is probably something to be considered prior to signing contracts.

Harassment

Apart from targeting institutions and organisations engaging in countering disinformation efforts, systematic attacks are often directed at individuals working on the topic or related issues. Targeting may include, for example, doxing (publishing and spreading personal information about the individuals and enabling further attacks), public smear campaigns and hate speech, death or other violent threats, misogynistic, racist or sexual online abuse, and taking spurious legal action against the individuals. Gender-based disinformation has emerged as a particularly important attack vector for hostile actors based on harassment methods.[98]

- Well-known incidents include the lengthy harassment campaign of Finnish journalist Jessikka Aro who investigated Russia's troll factory and influence techniques for the Finnish public broadcaster,[99] and Nobel Peace Laureate, Maria Ressa, who has faced a continuous online abuse campaign over the course of several years.[100]

- Nina Jankowicz faced large-scale online harassment after she was named the executive director of the newly created Disinformation Governance Board of the US Department of Homeland Security. Jankowicz resigned from the position and the board was disbanded.[101] In this latter case, it was clear that DHS was unprepared for the political attacks that followed the announcement of the board and failed to adequately defend Jankowicz from harassment.

Such attacks are often aimed at disrupting the work the individual is doing, but also to systematically control and prevent free public debate over certain issues. Here, litigation and harassment overlap as an information conflict strategy.

- In 2017, Russia scholar Martin Kragh published an agenda-setting study on Russian influence methods in Sweden in the journal *Security Studies*.[102] In response, he was reported to the university ethics board for academic dishonesty, accused of being an MI6 agent, and had his reputation attacked in several publications.[103]

- Three Russian oligarchs initiated libel proceedings against journalist Catherine Belton and HarperCollins, the publisher of her book "Putin's People". The defendants settled or withdrew their claims in the end. The publisher agreed to several amendments related to the Russian oligarch Roman Abramovich.[104]

- Following her revelations about Cambridge Analytica and related scandals, journalist Carole Cadwalladr was sued in what several international journalist associations termed as "vexatious in nature and intended to silence Cadwalladr's courageous investigative journalism."[105] So-called Strategic Litigation Against Public Participation (SLAPP) lawsuits have also been levelled against NGOs engaged in exposing disinformation.

The problem with these kinds of attacks on individuals is that their employers handle the threats inconsistently. In some of the cases mentioned above, the employers offered tremendous support. In others, and in a number of further cases not mentioned here, employers have not supported their staff at all. It is difficult to determine general rules of best practices for what governments can do in these cases, since organisations have their own provisions. However, it should be clear that harassment efforts backed by foreign intelligence agencies demand heightened government protection and response to the extent that such support is possible.

The ethics of outsourcing information conflict

Information conflict refers to a wide array of activities that seek to contest information and information systems in order to influence decision-makers at various levels, across all sectors of society. Democracies depend on civil society and the private sector to complement government in providing for society, and hence it is unavoidable that these actors are part of the counter-disinformation community. Civil society, researchers, media and the private sector should and must be involved in the protection of democratic debate.

Governments are perhaps the only actors in these collaborative partnerships with the means and capabilities to monitor and respond to some of the harshest retaliation efforts, particularly if espionage and cyber-attacks are involved. Indeed, they have a responsibility to support all non-governmental actors that they fund, but not necessarily equally. In our opinion, the following distinctions may be observed:

- **Projects where information conflict is likely to lead to retaliation.** Some delegated tasks are more likely to place the organisation at risk of retaliation. Organisations performing such roles should receive more support for conducting risk assessments, if necessary, with the support of intelligence analysts with a detailed understanding of the threat actor.

- **Projects where information conflict has domestic political consequences.** Some tasks of the defender community can occasionally overlap with polarised domestic political controversies, such as with COVID-19. All parties should get better at assessing associated risks, particularly given that much disinformation seeks to exploit precisely these kinds of polarised topics.

- **Directly funded contractors and subcontractors.** Governments would expect to assume a greater responsibility to the people and organisations they fund directly than to those receiving indirect funding, for example, as subcontractors of their grantees. However, it is not always possible to delegate security responsibilities to non-governmental organisations. At times, a government security "umbrella" for an entire project or work package may be appropriate.

- **Development of transparent practices.** In a number of the cases mentioned above, claims about "secret dealings" became one of the main vectors through which the threat actor sought to discredit members of the defender community. The general trend to tackle this vulnerability has been to develop as transparent of practices as possible, and we argue that more thought needs to be given to the public disclosure of outsourced counter-disinformation activities.

Beyond the ethical responsibility to protect members of the defender community, what are the governments' options and limitations to intervene? In which situations should the governments exercise their power to intervene, and what are the limitations?

When democratic governments and international organisations set up new bodies (such as units within governmental departments) to address malign influence, they should spend considerable time and resources in preparation of these efforts. Unless aspects

of the work are secret, the public should have access to information about the mandate and the resources, and governments should be able to communicate clearly about the powers and limitations of these new bodies, while protecting national security interests.

If any of these parts fail, it seems more likely that vulnerabilities can be exploited by a hostile actor to, for example, sow distrust of these efforts. As the resources backing the defender community increase, particularly in highly contested areas such as foreign interference, governments must be able to convincingly communicate about the efficiency and results of the work, not only to their civil society partners but to the public. This need for transparency and accountability is also important when designing countermeasures, which may be secret initially but are likely to end up as a matter of public record.

The limitation or issue governments face is the inconsistent sharing of pertinent information, which can lead to the often disjointed and uncoordinated nature of defensive activities in the information environment. This creates an environment where a hostile actor or actors can conduct systematic strategic and expansive influence operations, while the defensive measures are fragmented and uneven, spread out in different countries, and not spoken about in public.[106] While it is clear that part of this problem is also the strength of democracies – there is no single body that can take full responsibility for the response - democratic governments should take more responsibility in demonstrating how they achieve the aim of raising the cost and deterring hostile actors' operations.

Yet, as with all sensitive activities, it often seems that governments want civil society programmes to be "deniable" if and when anything goes wrong. This seems disingenuous. Governments and civil society are partners and collaborators in protecting democratic societies. Longstanding tendencies within governments to over-classify information, in conjunction with the security and hostile state actor aspects of influence operations, have perhaps pushed some of this work too far out of the public eye.[107]

Recommendations:

- Governments should make **comprehensive risk assessments** when engaging in partnerships with the private sector and civil society, where counter-disinformation activities present a clear risk of information conflict. Currently, there is no best practice on how to do this. Only a few governments are perceived to have sufficient understanding of the risks and carry out systematic work with their partners to mitigate them.

- The government and the funded partner should be aware of and **reach a mutual understanding on at least the following areas of risks:** legal, diplomatic, and regulatory measures; the impact of social media platforms' policies; cyber security; political and reputational attacks; online and offline harassment.

- A risk assessment should **clarify the responsibilities of each party** in the case of a legal, reputational, or cyber-attack in order to increase preparedness in a situation when a certain risk, or several at once, materialise. Some questions to consider include: Who will investigate or cover the cost for investigating a cyber-attack? Under which circumstances is the funder ready to provide support for a court case or reputational attack? Who is responsible for monitoring the threats to an organisation or individual? How is reporting these threats organised? Who is best positioned to coordinate with the social media platforms in case of an online harassment campaign?

- The relationship and legal responsibilities between the funder and

the organisation, company, or an individual receiving the funding, varies from case to case. In some cases, the partnership operates in full transparency, in others a level of confidentiality is applied, and in some cases parts of the information is classified. **Does increased security in a project entitle the grantee to enhanced support in the case of security breaches?** If secrecy awakens the interest of a hostile intelligence agency, this seems essential.

- If a fact-checker or an NGO engaging in countering disinformation receives government funding, the accountability, independence, and integrity of the receiver are essential. Funding could also be included for the receiver to **comply with the international standards and code of ethics** in the field; or for building capacity towards achieving that. In the case of fact-checking, the standards are already well-developed, while for others the best practice still needs to be developed. In most cases, the best way to preserve accountability and independence is by reference to national or international standards.

- The **level of anonymity** guaranteed by the funder should be discussed and agreed upon. Some of the legal and regulatory measures, like freedom of information requests, may involve trying to trick governments into passing on information jeopardising the third party.

- In all cases, **collaboration agreements should include consideration of the best interests of the civil society partner** in the case of information conflict. While governments can outsource counter-disinformation activities, they should not be able to outsource responsibility for these activities. The field must mature in this respect, and quickly.

Endnotes

1 Thomas Kent, 'In the global meme wars, it's time to side with the elves against the trolls'. *Washington Post*, November 16 2022

2 Stephanie Kirchgaessner, Manisha Ganguly, David Pegg, Carole Cadwalladr and Jason Burke, 'Revealed: the hacking and disinformation team meddling in elections', *The Guardian*, 15 Feb 2023. DFRLab, 'The Russians Who Exposed Russia's Trolls', *Medium*, 8 March 2018

3 Clare Wardle, Hossein Derakhshan, Information Disorder, Council of Europe report DGI(2017)09

4 NATO's approach to countering disinformation: a focus on COVID-19

5 Pamment, James (2020) The EU's Role in Fighting Disinformation: Taking Back the Initiative (Part 1). Washington DC: Carnegie Endowment for International Peace.

6 Media Defence, 'Misinformation, Disinformation and Mal-Information'

7 Pamment, J., Nothhaft, H., Agardh-Twetman, H., & Fjällhed, A. (2018) Countering Information Influence Activities: The State of the Art. Swedish Civil Contingencies Agency (MSB). Stockholm: MSB.

8 Facebook (2021) Threat Report The State of Influence Operations 2017-2020. No place of publication.

9 Thomas C. Wingfield (2000) *The Law of Information Conflict: National Security Law in Cyberspace*. Aegis Research Corp

10 Law Insider, 'Information conflict definition'

11 Alguliyev, Rasim M., Ramiz M. Aliguliyev, and Irada Y. Alakbarova. "Extraction of hidden social networks from wiki-environment involved in information conflict." *International Journal of Intelligent Systems and Applications* 8.2 (2016): 20

12 Brett van Niekerk & Manoj Mahara, Social Media and Information Conflict. *International Journal of Communication* 7 (2013).

13 Meta, 'Threat Report: Combating Influence Operations'

14 Graphika, 'Secondary Infektion'

15 Meta, 'Taking down violating content'

16 Meta, 'Removing Coordinated Inauthentic Behavior From China and Russia'

17 Pamment, J. & Smith, V. (2022) *Attributing Influence Operations: toward a community framework*. Riga: NATO Strategic Communications Centre of Excellence & EU-NATO Hybrid Centre of Excellence

18 Pamment, J. & Lindwall Kimber, A. (2021) *Fact-checking and debunking: a best practice guide to dealing with disinformation*. Riga: NATO Strategic Communications Centre of Excellence.

19 Pamment, J. & Lindwall Kimber, A. (2021) *Fact-checking and debunking: a best practice guide to dealing with disinformation*. Riga: NATO Strategic Communications Centre of Excellence.

20 Debunk, 'About Elves'

21 EDMO, 'Media literacy'

22 Lewandowsky, Cook, Schmid, Holford, Finn et al, 'The COVID-19 Vaccine Communication Handbook'

23 Nadiya Ivanenko, 'The Russian trace in "suspicious" edits in Wikipedia has been exposed', *Mezha* 19 October 2022

24 Martin Innes, 'High-profile Western media outlets repeatedly infiltrated by pro-Kremlin trolls'

25 Disinfodex.org

26 Pamment, J. & Smith, V. (2022) *Attributing Influence Operations: toward a community framework*. Riga: NATO Strategic Communications Centre of Excellence & EU-NATO Hybrid Centre of Excellence

27 Pamment, J. & Smith, V. (2022) *Attributing Influence Operations: toward a community framework*. Riga: NATO Strategic Communications Centre of Excellence & EU-NATO Hybrid Centre of Excellence

28 Non-disclosure agreements, which assert the confidentiality of shared data and limit how it may be used.

29 Bellingcat

30 Bellingcat, 'Socialite, Widow, Jeweller, Spy: How a GRU Agent Charmed Her Way Into NATO Circles in Italy'

31 Bellingcat

32 US Department of Justice, 'Report On The Investigation Into Russian Interference In The 2016 Presidential Election'

33 Ninon Bulckaert, 'How France successfully countered Russian interference during the presidential election', Euractiv 17 July 2018

34 Pamment, James (2021) *RESIST 2*. London: UK Government Communication Service

35 EUvsDisinfo

36 EUvsDisinfo, 'Removal of three cases further to complaints by Dutch media'

37 Council of the EU, 'EU imposes sanctions on state-owned outlets RT/Russia Today and Sputnik's broadcasting in the EU'

38 Fredheim, R., Stolze, M. Virtual Manipulation Brief, Issue 1/2022: Russia's Struggle to Circumvent Sanctions and Communicate Its War Against Ukraine. Riga: NATO Strategic Communications Centre of Excellence

39 Pamment, James (2022) How the Kremlin circumvented the EU sanctions on Russian state media in the first weeks of the illegal invasion of Ukraine. *Journal of Place Branding & Public Diplomacy*

40 Kayleen Devlin, Jake Horton and Olga Robinson, 'Ukraine crisis: Is Russia staging 'false flag' incidents?' BBC News, 23 February 2022

41 Bellingcat, 'Documenting and Debunking Dubious Footage from Ukraine's Frontlines'

42 Centre for Information Resilience

43 DFRLab, 'Russian War Report: Russian false-flag operation seeks to drag Belarus into Ukraine war'

44 Johns Hopkins University & Imperial College London, 'Countering cognitive warfare: awareness and resilience', *NATO Review*, 20 May 2021

45 Ellen Nakashima, 'U.S. Cyber Command operation disrupted Internet access of Russian troll factory on day of 2018 midterms', *Washington Post*, 27 February 2019

46 Matthew Holroyd, 'Russia Today row: EU says Moscow's expulsion of Deutsche Welle is 'unacceptable'', *euronews*, 4 February 2022

47 Pamment, James (2022) How the Kremlin circumvented the EU sanctions on Russian state media in the first weeks of the illegal invasion of Ukraine. *Journal of Place Branding & Public Diplomacy*

48 Le Monde, 'Fichier de comptes Twitter actifs lors de l'affaire Benalla : la CNIL saisie'

49 Stanford Internet Observatory, 'The Long Fuse'

50 https://twitter.com/katestarbird/status/1578065577146343424

51 Nathaniel Gleicher, 'Removing Coordinated Inauthentic Behavior from France and Russia'

52 Craig Timberg & Elizabeth Dwoskin, 'People affiliated with French military used Facebook to meddle in Africa', *Washington Post* 16 December 2020

53 Naomi Nix, 'Facebook, Twitter dismantle a U.S. influence campaign about Ukraine', *Washington Post,* 24 August 2022; Graphika & Stanford Internet Observatory, 'Unhear Voice'

54 Ellen Nakashima, 'Pentagon opens sweeping review of clandestine psychological operations', *Washington Post*, 19 September 2022

55 Tom Bateman, 'Russia could ban YouTube after it deleted two RT channels it said spread COVID misinformation', *euronews*, 1 October 2021

56 Rhett Jones, 'Facebook Just Made a Shocking Admission, and We're All Too Exhausted to Notice', *Gizmondo* 5 April 2018

57 Meta, 'An Update on Our Plans to Restrict Data Access on Facebook'

58 Tristan Hotham, 'Facebook risks starting a war on knowledge', *The Conversation*

59 Scholar.Google

60 Axel Bruns, 'Facebook shuts the gate after the horse has bolted, and hurts real research in the process', *Internet Policy Review*

61 Ad Observatory

62 Jeremy B. Merrill, 'What We Learned From Collecting 100,000 Targeted Facebook Ads', *ProPublica*, 26 December 2018

63 Kim Lyons, 'Facebook wants the NYU Ad Observer to quit collecting data about its ad targeting', *The Verge*, 24 October 2020

64 Jeremy B. Merrill, 'Facebook Moves to Block Ad Transparency Tools — Including Ours', *ProPublica*, 28 January 2019

65 David Gilbert, 'After Public Criticism, Facebook Will Allow Political Ad-Tracking Project to Continue', *Vice News*, 4 December 2020

66 Netgain Partnership, 'Facebook Must Not Block Research Into Disinformation'

67 Mike Clark, 'Research Cannot Be the Justification for Compromising People's Privacy'

68 Gilad Edelman, 'Facebook's Reason for Banning Researchers Doesn't Hold Up', *Wired*, 4 August 2021

69 Official Journal of the European Union, L277, Vol 65

70 European Commission, 'Digital Services Act: Delegated Regulation on independent audits now available for public feedback'

71 James Landale, 'Russia-linked hack 'bid to discredit' UK anti-disinformation campaign - Foreign Office', *BBC News*

72 Kit Klarenberg, 'Integrity Initiative: Foreign Office Funded, Staffed by Spies, Housed by MI5?' *Medium*; Paul McKeigue, David Miller, Jake Mason, Piers Robinson, 'Briefing note on the Integrity Initiative'

73 UK Parliament, UIN 196177, tabled on 27 November 2018

74 Wikispooks, 'Integrity Initiative'

75 James Landale, 'Russia-linked hack 'bid to discredit' UK anti-disinformation

campaign - Foreign Office', *BBC News*

76 CBS News, 'Russian SolarWinds hackers have launched new campaign, Microsoft says'

77 EU Disinfo Lab, 'Brussels needs to move beyond ready-made slogans in the fight against disinformation'

78 UK Parliament, UIN 196177, tabled on 27 November 2018

79 George Allison, '77 Brigade is countering Covid misinformation', *ukdj*

80 George Allison, 'Politician claims that the British Army's 77th Brigade is 'attacking' Scots online', ukdj

81 Oliver Carroll, 'Ukrainian journalist forced to flee following threats from far-right', *Independent*, 14 July 2020 |

82 Personal interview.

83 Brett Boudreau, 'The Rise and Fall of Military Strategic Communications at National Defence 2015-2021: A Cautionary Tale for Canada and NATO, and a Roadmap for Reform', *Canadian Global Affairs Institute*

84 David Pugliese, 'Forged letter warning about wolves on the loose part of Canadian Forces propaganda campaign that went awry', *National Post*

85 Murray Brewster & Ashley Burke, 'Military campaign to influence public opinion continued after defence chief shut it down', *CBC*; Mack Lamoureux, 'Conspiracy Theorists Are Salivating Over a Canadian Military Psy-Op Report', *Vice News*

86 EUvsDisinfo

87 European Council, EUCO 11/15

88 European Pravda, 'The EU trolls the ambassador of the Russian Federation for words about "ideological special forces" and an inferiority complex'

89 'Политизация экономических связей России и ЕС', The International Affairs

90 EUvsDisinfo, 'Removal of three cases further to complaints by Dutch media'

91 Arjen Nijboer, 'Why the EU must close EUvsDisinfo', *euobserver*

92 EUvsDisinfo, Disinfo review

93 Matt Apuzzo, 'Pressured by China, E.U. Softens Report on Covid-19 Disinformation', *New York Times*

94 Florian Elder, 'POLITICO Brussels Playbook: NATO hopes — Hogan's ambitions — EEAS' mob rule', *Politico*

95 Department of Homeland Security, 'Following HSAC Recommendation, DHS terminates Disinformation Governance Board'

96 'CSSF Programme Summary', UK Government

97 'Florence Parly présente la doctrine militaire de lutte informatique d'influence', *DICoD*

98 EU Disinfo Lab, 'Gender-Based Disinformation: Advancing Our Understanding and Response'

99 'Jessikka Aro', *International Women's Media Foundation*

100 Julie Posetti, Diana Maynard and Kalina Bontcheva, 'Maria Ressa: Fighting an Onslaught of Online Violence', *International Center for Journalists*

101 Martina Adami, 'Facing hate and abuse as a woman online: Nina Jankowicz on her latest book'

102 Martin Kragh & Sebastian Åsberg (2017) Russia's strategy for influence through public diplomacy and active

measures: the Swedish case, Journal of Strategic Studies, 40:6, 773-816, DOI: 10.1080/01402390.2016.1273830

103 Kragh, M. (2020). ”Martin Kragh är ett demokratiskt problem”: Hur Aftonbladet gav spridning åt en rysk påverkansoperation. *Statsvetenskaplig tidskrift*, *122*(3).

104 Luke Harding, **‘Roman Abramovich settles libel claim over Putin biography’**, *The Guardian*, 22 December 2021

105 Scottish PEN **‘Free expression groups call on Arron Banks to drop SLAPP lawsuit against Carole Cadwalladr’**

106 Anneli Ahonen & Martin Innes, **‘The Ghostwriter Campaign as a multi-vector operation’**

107 Mike Giglio, **‘The U.S. Government Keeps Too Many Secrets’**, *The Atlantic*

Prepared and published by the
NATO STRATEGIC COMMUNICATIONS
CENTRE OF EXCELLENCE

The NATO Strategic Communications Centre of Excellence (NATO StratCom COE) is a NATO accredited multi-national organisation that conducts research, publishes studies, and provides strategic communications training for government and military personnel. Our mission is to make a positive contribution to Alliance's understanding of strategic communications and to facilitate accurate, appropriate, and timely communication among its members as objectives and roles emerge and evolve in the rapidly changing information environment.

www.stratcomcoe.org | @stratcomcoe | info@stratcomcoe.org