Towards Norms for State Responsibilities Regarding Online Disinformation and Influence Operations

Brett van Niekerk[1] and Trishana Ramluckan[2,3]
[1] Durban University of Technology, South Africa
[2] University of KwaZulu-Natal, South Africa
[3] Educor Holdings, South Africa
brettv@dut.ac.za
ramluckant@ukzn.ac.za

Abstract: The Internet has provided a global mass communication system, and in particular social media technologies began a social revolution for the public sphere. However, these platforms have been exploited for the purposes of influence operations and disinformation campaigns to hinder or subvert national decision-making processes by affecting the policy makers, voters, or swaying general public opinion. Often this is achieved through manipulative means falling within a grey area of international and constitutional systems. Existing proposed normative frameworks for responsible state behaviour in Cyberspace have tended to focus on cyber operations. While online influence operations are recognised as a concern, they were not explicitly discussed in the frameworks, resulting in knowledge gaps related to countering influence operations and disinformation. There is a growing narrative that influence operations and disinformation campaigns are a cyber security issue and nations sometimes include legislation related to disinformation in cyber security. This indicates that existing cyber norms can be used to guide the development of norms for addressing disinformation and influence operations. This paper aims to propose a normative framework for state responsibility relating to influence operations emerging from thematic analysis of existing cyber norms and research on mitigating influence operations.

Keywords: Cyber diplomacy, Cyber norms, Disinformation, Influence operations, Information warfare

1. Introduction

Disinformation and coordinated influence operations, often with a strong online component, have been receiving attention since the alleged attempts to influence the 2016 US presidential election. During the COVID-19 pandemic there was an increase in disinformation. While disinformation had been recognised as a concern, limited coordinated international recognition of the problem with a view of mitigating disinformation have been established. Whilst the United Nations General Assembly (UNGA, 2021) passed a resolution in 2021, this does not have the maturity of 'traditional' cybersecurity, which has had multiple Groups of Governmental Experts (GGEs) with a second Open Ended Working Group (OEWG) at the time of writing and a proposed programme of action (Digital Watch, 2023).

Disinformation can be defined as the intentional spreading of false or misleading information, whereas misinformation is unintentional, and malinformation is the use of correct information used for misleading purposes (Khan, 2021). Influence operations are then considered the use of these techniques to manipulate and audience to support the goals of an actor, which may be a foreign actor seeking to subvert processes in another nation (Pamment, 2022).

This paper aims to begin the process of establishing norms for addressing disinformation and influence operations by analysing existing cyber norms and international documents on disinformation. Thematic analysis is used to identify key themes that the proposed normative framework should be based upon. The paper contributes an initial framework of norms for state responsibility in countering, conducting, and responding to online disinformation and influence campaigns, based upon the key themes identified in the document analysis.

Section 2 presents a background to facets of disinformation and influence operations, the relationship with cybersecurity, and cybersecurity norms. Frameworks and legal aspects for disinformation and influence operations at both international and national levels are discussed in Section 3. The proposed norms are presented in Section 4, followed by the conclusion in Section 5.

2. Literature Review

2.1 Major disinformation and influence operations

The major event that raised international concern regarding disinformation and foreign influence operations was the 2016 US presidential election, where allegations of interference resulted in a number of political investigations and processes. The actors used a variety of techniques, including using adverts on social media

platforms. Social data used to design manipulative messaging was linked to Cambridge Analytica and was also found to have influence the BREXIT referendum. Similarly, a private firm, Bell Pottinger, was at the centre of a scandal where disinformation was used to increase racial tensions in South Africa to distract from accusations of political wrongdoing against the president (Ramluckan, Wanless and van Niekerk, 2019; Ramluckan and van Niekerk, 2019).

Disinformation was prevalent in Ukraine during both the anti-government protests and the annexation of Crimea in 2014, and evolved during the 2022 Russo-Ukraine conflict to include the use of deepfakes (Roache *et al*., 2022). Disinformation again become an international concern due to its prevalence during the COVID-19 pandemic, with conspiracy theories and political blaming. In addition, a nation used 10 000 social media bots and compromised accounts for propaganda purposes (Allen-Ebrahimian, 2020).

Baines and Jones (2018) and Rid (2020) provide historical examples of election interference and influence operations, respectively. While they demonstrate that the concept is not new, the introduction of social media and cyber elements have evolved the practice dramatically. While many key examples focus on Europe, other areas are also affected by foreign influence operations. For example, approximately 60 percent of the over 50 recorded disinformation operations in Africa are considered to originate externally, with 16 attributed to Russia (Africa Center for Strategic Studies, 2022).

2.2 Influence operations versus cyber operations

Both influence operations and cyber operations both emerged out of the concept of information warfare, which encompassed military psychological operations and computer network operations; these were extended outside of a conflict situation to information operations. The terminology has evolved, where information operations and information warfare are often used synonymously with influence operations; for example, Stengel's (2019) book *Information Wars: How We Lost the Global Battle Against Disinformation and What We Can Do About It*, focuses purely on disinformation as the title suggests. Likewise, the term 'cyber operations' has become far more common and largely replaced the use of 'computer network operations'. Rid's (2020) book, *Active Measures: The Secret History of Disinformation and Political Warfare*, considers a cyber security incident as an influence operation: a group calling themselves the Shadow Brokers were selling cyber security tools stolen from the U.S. National Security Agency online; however, the narrative surrounding this appeared to be an influence operation to embarrass the agency as the tools were eventually released openly on the Internet.

Gleicher (20221; 2022b) indicates that there are similarities between cyber security and defending against disinformation and influence operations: both situations are adversarial, and the threats evolve to circumvent defences. Within the cyber security domain, cybercriminals offer specialist services for hire, and this has translated to services for hire for influence operations by relevant threat actors (Gleicher, 20221; 2022b). It is reported that a North African cybercriminal shifted the focus of operations from cybercrime to distributing online propaganda (Lunghi, 2017). Account compromises, a common occurrence in the cyber security domain, is used to gain access to platforms with which to distribute and amplify influence operation narratives (Gleicher, 20221; 2022b), illustrating a cross-over between cyber security.

The Disinformation Analysis & Response Measures (DISARM) Framework provides two key frameworks describing the disinformation creation tactics, techniques and procedures (TTPs) by threat actors and mitigation TTPs by defenders (DISARM Foundation, n.d.; Terp and Breuer, 2022). This is modelled on the MITRE ATT&CK and D3FEND frameworks for 'traditional' cyber security. In addition, a STIX (Structured Threat Information eXpression) capability is available for threat information sharing (DISARM Foundation, n.d.), as is common with cyber security. The European Union Agency for Cybersecurity and EEAS (ENISA and EEAS, 2022) explicitly maps components of the MTRE ATT&CK framework to the DISARM framework; however, it is indicated that the DISARM framework is aligned to describing campaigns, whereas the ATT&CK framework is better at describing a cyber security incident. ENISA and EEAS (2022) indicates important components of cyber security that are relevant to foreign information manipulation and interference: cyber operations can be used for developing content (deep fakes or hack and leak operations) to support a narrative, or can compromise accounts and infrastructure to provide a platform for influence operations and dissemination; from a defensive perspective, cyber security techniques can be used to investigate the technical aspects to influence operations and aid in attribution. These aspects echo that of Gleicher mentioned above.

One of the strategies proposed to mitigate influence operations and disinformation is a 'whole of society approach' (Gleicher, 20221; 2022b), similar to the multi-stakeholder approaches in cyber security and the UN OEWG. Wanless and Shapiro (2022) echo a collaborative approach, advocating for research into disinformation

and influence operations to be based on a model such as the European Centre for Nuclear Research (CERN); currently individual projects are funded, resulting in replication of infrastructure for each project, whereas a conglomerate of nations could support global researchers with a single large-scale project to maximise economy of scale.

Due to the volume of possible incidents, automation is required for detecting both cybersecurity and disinformation threats, with artificial intelligence (AI) and machine learning playing a major role. While AI and machine learning have been successful within the cybersecurity space (Crowley, 2022; Fein and Stocker, 2022), it has been reported that AI is limited in detecting disinformation and influence operations (U.S. Federal Trade Commission, 2022; Kelley, 2022). Currently AI is not able to adequately detect nuances of human speech, and the available datasets used to train algorithms are not yet diverse enough, and technological solutions alone is not enough to address the problem (Kelley, 2022; IEEE Computer Society, 2021). Therefore, it is important to have an international normative framework to guide a common approach to addressing disinformation and influence operations. The next section provides and overview of norms within the cybersecurity domain which is used in conjunction with the legal aspects covered in Section 3 as a basis to develop norms for mitigating disinformation in Section 4.

2.3 Norms for state responsibility in Cyberspace

The Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security in the 2013 report indicated that International Law applies to Cyberspace, and in a 2015 report proposed 11 norms for responsible state behaviour in Cyberspace (Brown, Esterhuysen, and Kumar, 2019; United Nations General Assembly, 2015):

a) "Consistent with the purposes of the United Nations, including to maintain international peace and security, States should cooperate in developing and applying measures to increase stability and security in the use of ICTs and to prevent ICT practices that are acknowledged to be harmful or that may pose threats to international peace and security" (Brown, Esterhuysen, and Kumar, 2019: 4).

b) "In case of ICT incidents, States should consider all relevant information, including the larger context of the event, the challenges of attribution in the ICT environment and the nature and extent of the consequences" (Brown, Esterhuysen, and Kumar, 2019: 5).

c) "States should not knowingly allow their territory to be used for internationally wrongful acts using ICTs" (Brown, Esterhuysen, and Kumar, 2019: 5).

d) "States should consider how best to cooperate to exchange information, assist each other, prosecute terrorist and criminal use of ICTs and implement other cooperative measures to address such threats. States may need to consider whether new measures need to be developed in this respect" (Brown, Esterhuysen, and Kumar, 2019: 5).

e) "States, in ensuring the secure use of ICTs, should respect Human Rights Council resolutions 20/8 and 26/13 on the promotion, protection and enjoyment of human rights on the Internet, as well as General Assembly resolutions 68/167 and 69/166 on the right to privacy in the digital age, to guarantee full respect for human rights, including the right to freedom of expression" (Brown, Esterhuysen, and Kumar, 2019: 5).

f) "A State should not conduct or knowingly support ICT activity contrary to its obligations under international law that intentionally damages critical infrastructure or otherwise impairs the use and operation of critical infrastructure to provide services to the public" (Brown, Esterhuysen, and Kumar, 2019: 6).

g) "States should take appropriate measures to protect their critical infrastructure from ICT threats, taking into account General Assembly resolution 58/199 on the creation of a global culture of cybersecurity and the protection of critical information infrastructures, and other relevant resolutions" (Brown, Esterhuysen, and Kumar, 2019: 6).

h) "States should respond to appropriate requests for assistance by another State whose critical infrastructure is subject to malicious ICT acts. States should also respond to appropriate requests to mitigate malicious ICT activity aimed at the critical infrastructure of another State emanating from their territory, taking into account due regard for sovereignty" (Brown, Esterhuysen, and Kumar, 2019: 6).

i) "States should take reasonable steps to ensure the integrity of the supply chain so that end users can have confidence in the security of ICT products. States should seek to prevent the proliferation of malicious ICT tools and techniques and the use of harmful hidden functions" (Brown, Esterhuysen, and Kumar, 2019: 6).

j) "States should encourage responsible reporting of ICT vulnerabilities and share associated information on available remedies to such vulnerabilities to limit and possibly eliminate potential threats to ICTs and ICT-dependent infrastructure" (Brown, Esterhuysen, and Kumar, 2019: 7).

k) "States should not conduct or knowingly support activity to harm the information systems of the authorized emergency response teams (sometimes known as computer emergency response teams or cybersecurity incident response teams) of another State. A State should not use authorized emergency response teams to engage in malicious international activity" (Brown, Esterhuysen, and Kumar, 2019: 7).

The wording of the above norms is generic enough that most of the norms are applicable, particularly as they consider malicious use of ICTs that could affect international security, and this is not necessarily limited to 'traditional' cyber-operations. The concepts of assistance, information exchange, protecting human rights, and not conducting malicious activity will all be applicable to disinformation as well as cyber-operations. Where the norms are particularly specific and technical, they may not be relevant to disinformation scenarios, such as protecting emergency response teams and core infrastructure. In addition to the GGEs an Open-Ended Working Group process was established with a key initiative being the inclusion of multi-stakeholder engagement, which has had challenges but has also resulted in beneficial engagement (Digital Watch, 2023).

The Global Commission on the Stability of Cyberspace (GCSC, 2019) proposed eight norms:

1. "State and non-state actors should neither conduct nor knowingly allow activity that intentionally and substantially damages the general availability or integrity of the public core of the Internet, and therefore the stability of cyberspace." (p. 21)
2. "State and non-state actors must not pursue, support or allow cyber operations intended to disrupt the technical infrastructure essential to elections, referenda or plebiscites." (p. 21)
3. "State and non-state actors should not tamper with products and services in development and production, nor allow them to be tampered with, if doing so may substantially impair the stability of cyberspace." (p. 21)
4. "State and non-state actors should not commandeer the general public's ICT resources for use as botnets or for similar purposes." (p. 21)
5. "States should create procedurally transparent frameworks to assess whether and when to disclose not publicly known vulnerabilities or flaws they are aware of in information systems and technologies. The default presumption should be in favour of disclosure." (p. 21)
6. "Developers and producers of products and services on which the stability of cyberspace depends should (1) prioritize security and stability, (2) take reasonable steps to ensure that their products or services are free from significant vulnerabilities, and (3) take measures to timely mitigate vulnerabilities that are later discovered and to be transparent about their process. All actors have a duty to share information on vulnerabilities in order to help prevent or mitigate malicious cyber activity." (pp. 21-22)
7. "States should enact appropriate measures, including laws and regulations, to ensure basic cyber hygiene." (p. 22)
8. "Non-state actors should not engage in offensive cyber operations and state actors should prevent such activities and respond if they occur." (p. 22)

These eight norms are more technical in nature, as the intention was to focus on the public core of the Internet, ensuring the correct technical functioning of the Internet. However, some of the norms contain concepts that can be adapted to the case of influence operations, such as election interference, using public ICT resources for botnets (which could be used for distributing disinformation), implementing appropriate measures for cyber hygiene, and restricting malicious behaviour by non-state actors.

The Paris Call (2018) comprises of 9 principles:

1. "Protect individuals and infrastructure: Prevent and recover from malicious cyber activities that threaten or cause significant, indiscriminate or systemic harm to individuals and critical infrastructure."
2. "Protect the Internet: Prevent activity that intentionally and substantially damages the general availability or integrity of the public core of the Internet."
3. "Defend electoral processes: Strengthen our capacity to prevent malign interference by foreign actors aimed at undermining electoral processes through malicious cyber activities."
4. "Defend intellectual property: Prevent ICT-enabled theft of intellectual property, including trade secrets or other confidential business information, with the intent of providing competitive advantages to companies or commercial sector."
5. "Non-proliferation: Develop ways to prevent the proliferation of malicious software and practices intended to cause harm."
6. "Lifecycle security: Strengthen the security of digital processes, products and services, throughout their lifecycle and supply chain."
7. "Cyber hygiene: Support efforts to strengthen an advanced cyber hygiene for all actors."
8. "No private hack back: Take steps to prevent non-State actors, including the private sector, from hacking-back, for their own purposes or those of other non-State actors."
9. "International norms: Promote the widespread acceptance and implementation of international norms of responsible behaviour as well as confidence-building measures in cyberspace."

Of the nine principles, some are directly relevant and some are partially relevant to influence operations. Concepts such as cyber hygiene and preventing election interference are again present. Non-proliferation can be applied to tools or capacity to conduct influence operations. Principle 9 should be of particular note: it explicitly supports normative processes, which therefore provides motivation for considering norms for mitigating influence operations.

3. Laws and Frameworks related to Disinformation and Influence Operations

3.1 International law and frameworks

In December 2021 the UNGA passed a resolution on *Countering disinformation for the promotion and protection of human rights and fundamental freedoms*, which recognises the harms disinformation can cause, and encourages civil society, media organisation (including social media) and states to take measures to mitigate the impacts of disinformation within the boundaries of human rights (UNGA, 2021). The resolution also recognises a special report to the UN Human Right Council on *Disinformation and freedom of opinion and expression* (Kahn, 2021). A key aspect that was noted in the report is that under human rights freedom of expression is only limited in that it should not infringe upon another's rights, therefore the spreading of disinformation is ultimately protected by human rights (Kahn, 2021); similarly, Schmitt (2017) indicates that "propaganda does not constitute a prohibited intervention as it is not coercive in nature", therefore, unless propaganda explicitly incites war or civil unrest in another country, it is unlikely to constitute a wrongful act.

Wanless and Shapiro (2021) point out that the co-sponsors of the above-mentioned UN resolution all have poor rankings in terms of corruption, media freedom and democracy; they suggest that authoritarian regimes may benefit from international norms that provide them a mechanism for controlling the information domain within their territories. Stengel (2019) indicates a similar dilemma: democracies allow for open debate; therefore a 'fringe' idea needs to have an opportunity to respond to critics, thereby giving the ideas a platform and a degree of credibility. By comparison, authoritarian regimes will simply sensor information. Consequently, democratic processes enable the spread on disinformation (Stengel, 2019).

Pamment (2022) in a publication by the NATO Strategic Communications Centre of Excellence considers both the capabilities and actions required to counter disinformation and influence operations. For the capability, four key considerations are proposed: objectives, indicators, process maturity and risk assessment; of note is the consideration of norms under the 'objectives' category. Mitigation measures include content correction (fact checking) and public resilience (media literacy), analysis and identification, strategic communication, intelligence, and security policies. In addition, the importance of partnerships and professional development are highlighted (Pamment, 2022). ENISA and EEAS (2022) as mentioned above provide a more technical focus, and

suggest partnerships between the cybersecurity and disinformation/influence operations communities, increasing the quality and quantity of disinformation incident information, and capacity building amongst the various groups. They proposed the DISARM framework (DISARM Foundation, n.d.; Terp and Breuer, 2022), which will give consistent reporting of key information on detected disinformation and influence campaigns to improve collaboration and monitoring. Gleicher (2022a; 2022b) proposes an 'Adversarial Design Cycle' to mitigate disinformation and influence campaigns, with four main stages: threat ideation to assess threats and trends to predict future threats and design countermeasures, threat disruption to monitor and mitigate threats, improved defences to increase resilience, and partnerships to facilitate a whole-of-society approach.

In the above-mentioned documents, key themes emerge: the need for partnerships and collaboration, and capacity building. The Carnegie Endowment for International Peace's (2023) Partnership on Countering Influence Operations (PCIO) is aimed at collaborative research and information sharing. The Atlantic Council's Digital Forensics Research Lab (DFRLab, 2023) runs the Digital Sherlocks programme, which is also research-focussed and has initiatives aimed at capacity building and collaboration.

3.2 National laws and processes

Gleicher (2022a; 2022b) suggests a whole-society approach and that governments should introduce regulations to help mitigate disinformation. In Africa some countries did this by introducing the criminalisation of COVID disinformation in 2020; however, there were concerns that national laws criminalising disinformation may allow totalitarian and authoritarian states to abuse the laws to censor information criticising the government or limit the spread of opinions and information that do not align to an official government position (Hodgson, Farise, and Mavedzenge, 2020; Moyo, 2020). For example, in 2017 a US journalist was arrested in Zimbabwe for allegedly insulting former president Robert Mugabe (Burke, 2017). In 2022 Russia outlawed the use of 'war' and 'invasion' in reference to what was termed a 'special military operation' in Ukraine (Simon, 2022).

The implementation of national legislation therefore exhibits a challenge in reducing disinformation while protecting the freedom of expression; as indicated in Section 3.1, the freedom of expression is protected even if the content expressed in false. Given the severity of disinformation and influence operations, there appears to be the need for alignment to international humanitarian law to strengthen the required responsibilities related to freedom of expression in order to limit the intentional spread of disinformation. In addition, nations put pressure on social media organisations to detect and limit the spread disinformation during the pandemic (de Wet, 2020; Gold, 2020). Other initiatives included fact checking, such as the Real411 platform by the Media Monitoring Africa (MMA, 2023), Viral Facts Africa (2023), and multiple agencies in Europe (López-Marcos and Vicente-Fernández, 2021). Kyriakidou, Cushion, Hughes, and Morani (2022) indicate that while there is a recognised need for fact checkers in the UK, there was low awareness amongst the participants of the existing fact checkers. López-Marcos and Vicente-Fernández (2021) indicate that fact checking in the UK was more like a business, whereas in Spain it was conducted by NGOs. This indicates multiple models that fact checking can be deployed in a nation, and also highlights the benefits of multi-stakeholder involvement in mitigating disinformation.

4. Proposed Normative Framework for State Responsibilities on Online Disinformation and Influence Operations

This section provides thematic analysis of the above norms, principles, and documents and indicates the relevance to disinformation and influence operations. These themes form the basis for proposing a normative framework for influence operations. Table 1 provides an indication of the relevance of existing cyber norms to disinformation.

Table 1: Relevance of Existing Norms and Principles to Disinformation and Influence Operations

Norm			Directly Relevant	Indirectly Relevant	Not Relevant
2015 GGE	a)	Cooperation for implementing measures		X	
	b)	Attribution		X	
	c)	Not allow wrongful acts in territory	X		
	d)	Information exchange		X	
	e)	Respect human rights	X		
	f)	Do not conduct or support ICT activity contrary to international obligations		X	
	g)	Protect critical infrastructure			X
	h)	Respond to requests for assistance		X	
	i)	Integrity of the supply chain			X
	j)	Reporting of vulnerabilities			X
	k)	Do not harm CERTs			X
GCSC	1	No activities that damage public core of the Internet			X
	2	No activities that damage the technical infrastructure for elections		X	
	3	Do tampering with products			X
	4	Do not commandeer ICT resources for botnets		X	
	5	Vulnerability disclosure		X	
	6	Developer obligations			X
	7	Implement measures and laws for cyber hygiene		X	
	8	No non-state offensive operations		X	
Paris Call	1	Protect individuals and infrastructure		X	
	2	Protect the Internet			X
	3	Defend electoral processes		X	
	4	Defend intellectual property			X
	5	Non-proliferation		X	
	6	Lifecycle security			X
	7	Cyber hygiene		X	
	8	No private hack back			X
	9	International norms		X	

Norm e) can apply directly to the case of disinformation, as the wording is broad enough that it does not necessarily include technical subversion of infrastructure, but the use of ICTs in general; norm c) can be incorporated into mitigating non-state activity within a territory. Due to the concerns around the implementation of legislation, this is excluded from the proposed norms. In addition, as disinformation and influence operations do not breach international law unless there is a specific impact (e.g. interference on national processes), a generic norm to refrain from such operations is not suitable, and can be covered by not interfering in national processes.

The proposed norms for responsible state behaviour regarding disinformation and influence operations are:

1. States "should respect Human Rights Council resolutions 20/8 and 26/13 on the promotion, protection and enjoyment of human rights on the Internet, as well as General Assembly resolutions 68/167 and 69/166 on the right to privacy in the digital age, to guarantee full respect for human rights, including the right to freedom of expression" (Brown, Esterhuysen, and Kumar, 2019: 5).
2. States should implement measures to detect, monitor, and mitigate disinformation and influence operations.
3. States should cooperate with information exchange on detected operations with other nations to aid global detection and monitoring.
4. States should respond to requests for assistance in mitigating disinformation and influence operations.
5. States should foster multi-stakeholder environment for addressing disinformation and influence operations.
6. States should institute media-literacy and capacity building programmes to mitigate the impact of disinformation and influence operations

7. States should not conduct or support disinformation or influence operations that will interfere with another states' national processes (e.g. elections).
8. States should not commandeer computing resources or accounts for distribution of disinformation or influence operations.
9. States should not knowingly allow their territory to be used for disinformation or influence operations, and should discourage non-state actors within their jurisdiction from participating in such activity operations.
10. States should not proliferate capabilities for conducting disinformation and influence operations.

Table 2 presents key themes from the relevant documents discussed above, and indicates which of the proposed norm they align to (the norms are presented after Table 2). As illustrated, the GGE and the Paris Call each aligned to 6 of the key themes and norms, with the GCSC aligning to 5 key themes and norms. ENISA & EEAS, together with Gleicher (2022a; 2022b) aligned to three of the key themes and norms, and Pamment (2022) to only two. The most prevalent alignment across documents is to Norm 2, which is the implementation of measure to mitigate disinformation.

Table 2: Alignment of key themes to documents and proposed norms

	Human rights	**Cooperation & partnerships**	**Capacity building**	**Assistance**	**Electoral processes**	**Implement legislation & policies**	**Refrain from malicious actions**	**Information sharing and reporting**	**Implement mitigating measures**	**Non-proliferation**	**Mitigate use of territory**	**Mitigate non-state activity**	**Commandeering of resources**
GGE	X	X		X			X	X			X		
GCSC					X	X		X	X			X	X
Paris Call		X	X		X				X	X		X	
Pamment (2022)			X			X			X				
ENISA & EEAS		X						X	X				
Gleicher (2022a; 2022b)		X				X		X	X				
Proposed norm	**1**	**3 & 5**	**6**	**4**	**7**	N/A	N/A	**3**	**2**	**10**	**9**	**9**	**8**

5. Conclusion

Disinformation and influence operations have been recognised as a growing concern for international security; however, the international discussion is not as mature as those for cyber operations which has multiple recognised international norms and principles. Discussions have emerged placing disinformation as a cybersecurity problem, and similarities between addressing disinformation and cybersecurity have been considered in research and official documents. This indicates that the concept of cyber norms for cybersecurity may be a basis for developing norms on state behaviour for addressing influence operations.

In this paper, thematic analysis is used to analyse existing cyber norms and disinformation documents in order to identify major themes related to responsible state behaviour and disinformation. Thirteen themes were identified, and from these a framework of 10 norms for responsible state behaviour for influence operations is proposed. Due to concerns over possible Human Rights abuses through the implementation of legislation, promoting human rights was favoured over the theme of legislation and policies. As with cybersecurity, multi-stakeholder and international cooperation is important, and nations should discourage non-state actors from contributing to influence operations and disinformation, and aid efforts to disrupt such operations when they are detected.

The significance of this paper is the application of the cybersecurity normative frameworks to the influence operations space, and to propose norms specific to disinformation. This further demonstrates an overlap between cybersecurity and influence operations, indicating a need for the two communities to engage. It is necessary to implement other measures to support a normative framework, such as confidence building

measures, which have been recognised in cybersecurity and international security fields. However, this is outside the scope of the current paper, and will be considered in future research.

Acknowledgements

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant no. 150381, received by the first author). The opinions, findings and conclusions or recommendations expressed in this paper are those of the authors, and not of the respective institutions or funding agencies.

References

Africa Center for Strategic Studies. (2022) *Mapping Disinformation in Africa*, 26 April, [online], accessed 18 January 2023, https://africacenter.org/spotlight/mapping-disinformation-in-africa/

Allen-Ebrahimian, B. (2020) Bots boost Chinese propaganda hashtags in Italy, Axios, 1 April, [online], accessed 12 January 2023, https://www.axios.com/bots-chinese-propaganda-hashtags-italy-cf92c5a3-cdcb-4a08-b8c1-2061ca4254e2.html

Brown, D., Esterhuysen, A., and Kumar, S. (2019) Unpacking the GGE's framework on responsible state behaviour: Cyber norms, Association for Progressive Communications and Global Partners Digital, 23 December, [online], accessed 4 January 2022, https://www.apc.org/en/pubs/unpacking-gges-framework-responsible-state-behaviour-cyber-norms.

Burke, J. (2017) "US woman charged over tweet allegedly insulting Robert Mugabe", *The Guardian*, 3 November, [online], accessed 14 January 2023, https://www.theguardian.com/world/2017/nov/03/martha-odonovan-arrested-zimbabwe-alleged-mugabe-goblin-tweet-harare

Carnegie Endowment for International Peace. (2023) Partnership on Countering Influence Operations, [online], accessed 12 January 2023, https://carnegieendowment.org/specialprojects/counteringinfluenceoperations/

Crowley, K. (2022) "5 Advantages of Deep Learning in Cybersecurity", Deep Instinct, 7 September, [online], accessed 30 December 2022, https://www.deepinstinct.com/blog/five-advantages-of-deep-learning-for-preventing-cybersecurity-threats

De Wet, P. (2020) "SA expects WhatsApp to 'immediately' remove fake coronavirus news under new rules", *Business Insider*, 27 March, [online], accessed 14 January 2023, https://www.businessinsider.co.za/ott-fake-news-rules-against-coronavirus-for-whatsapp-2020-3

DFRLab. (2023) 360/Digital Sherlocks, Carnegie Endowment for International Peace, [online], accessed 12 January 2023, https://www.digitalsherlocks.org/360os-digitalsherlocks

Digital Watch (2023). UN OEWG and GGE, Geneva Internet Platform, [online], accessed 12 January 2023, https://dig.watch/processes/un-gge

DISARM Foundation. (n.d.) DISARM Framework, [online], accessed 27 December 2022, https://www.disarm.foundation/framework

ENISA and EEAS. (2022). Foreign Information Manipulation and Interference (FIMI) and Cybersecurity – Threat Landscape, 8 December, [online], accessed 28 December 2022, https://www.enisa.europa.eu/publications/foreign-information-manipulation-interference-fimi-and-cybersecurity-threat-landscape

Federal Trade Commission. (2022) *Combatting Online Harms Through Innovation*, Report to Congress, 16 June, [online], accessed 30 December 2022, https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf

Fein, D., and Stocker, E. (2022) "Cyber AI Analyst: Cutting Through the Noise to Gain the Security Edge", Darktrace, 30 November, [online], accessed 30 December 2022, https://darktrace.com/blog/cyber-ai-analyst-cutting-through-the-noise-to-gain-the-security-edge

Gleicher, N. (2022a) "Innovating against Adversaries", Keynote, *14th International Conference on Cyber Conflict: Keep Moving*, 31 May to 3 June, [online], accessed 27 December 2022, https://www.youtube.com/watch?v=ULQ4RZedRlQ

Gleicher, N. (2022b) "Innovating Against Adversaries: Exploiting the Defenders' Advantage", Lightning Talk, 360/Open Summit: Contested Realities | Connected Futures, 6-7 June, [online], accessed 27 December 2022, https://youtu.be/ZOLkgoL17dE?t=18125

Gold, H. (2020) "YouTube tries to limit spread of false 5G coronavirus claims after cellphone towers attacked", *CNN*, 6 April, [online], accessed 14 January 2023, https://edition.cnn.com/2020/04/06/tech/5g-coronavirus-conspiracy/index.html

Hodgson, T.F., Farise, K., and Mavedzenge, J. (2020) "Southern Africa has cracked down on fake news, but may have gone too far", *Mail&Guardian*, 5 April, [online], accessed 14 January 2023, https://mg.co.za/analysis/2020-04-05-southern-africa-has-cracked-down-on-fake-news-but-may-have-gone-too-far/

IEEE Computer Society. (2021) *Mitigating Societal Harms in a Social Media World*, Lessons Learned: A Reflection on the IEEE Computer Society Tech Forum, 27 September, [online], accessed 30 December 2022, https://www.computer.org/publications/tech-news/research/misinformation-disinformation-resources#report

Kahn, I. (2021) *Disinformation and freedom of opinion and expression*, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Human Rights Council, 47th session, 21 June-9 July.

Kelley, A. (2022) "AI is 'No Magical Shortcut' FTC Says in Fighting Disinformation Online", *NextGov*, 17 June, [online], accessed 21 June 2022, https://www.nextgov.com/emerging-tech/2022/06/ai-no-magical-shortcut-ftc-says-fighting-disinformation-online/368341/

Kyriakidou, M., Cushion, S., Hughes, C., and Morani, M. (2022) "Questioning Fact-Checking in the Fight Against Disinformation: An Audience Perspective", *Journalism Practice*, DOI: 10.1080/17512786.2022.2097118

López-Marcos C., and Vicente-Fernández P. (2021) "Fact Checkers Facing Fake News and Disinformation in the Digital Age: A Comparative Analysis between Spain and United Kingdom", *Publications* 9(36), https://doi.org/10.3390/publications9030036

Lunghi, D. (2017) "From Cybercrime to Cyberpropaganda," *Trend Micro*, 16 October, [online], accessed 13 November 2017, http://blog.trendmicro.com/trendlabs-security-intelligence/from-cybercrime-to-cyberpropaganda/

Media Monitoring Africa. (2023) Real411, [online], accessed 14 January 2023, https://www.real411.co.za/

Moyo, T. (2020) "Covid-19 and the suppression of freedom of expression", *Daily Maverick*, 9 April, [online], accessed 14 January 2023, https://www.dailymaverick.co.za/article/2020-04-09-covid-19-and-the-suppression-of-freedom-of-expression-part-two/

Pamment, J. (2022) *A Capability Definition and Assessment Framework for Countering Disinformation, Information Influence, and Foreign Interference*, NATO Strategic Communications Centre of Excellence, [online], accessed 17 January 2023, https://stratcomcoe.org/publications/a-capability-definition-and-assessment-framework-for-countering-disinformation-information-influence-and-foreign-interference/255

Ramluckan, T., Wanless, A., and van Niekerk B. (2019) "Cyber-Influence Operations: A Legal Perspective", *Proceedings of the 18th European Conference on Cyber Warfare and Security* (ECCWS), Coimbra, Portugal, 4-5 July, pp. 379-388.

Ramluckan, T., and van Niekerk B. (2019) "International Humanitarian Law and Cyber-Influence Operations", *Journal of Information Warfare* 18(3), 67-82.

Roache, M, Tewa, S, Cadier, A, Labbe, C, Padovese, V, *et al*. (2022) "Russia-Ukraine disinformation tracking center: 303 websites spreading war disinformation and the top myths they publish", *NewsGuard*, 15 November, [online], accessed 16 November 2022, https://www.newsguardtech.com/special-reports/russian-disinformation-tracking-center/

Rid, T. (2020) *Active Measures: The Secret History of Disinformation and Political Warfare*, New York: Farrar, Straus and Giroux.

Schmitt, M.N. (2017) *Tallinn Manual 2.0: On the International Law Applicable to Cyber Operations*, Cambridge: Cambridge University Press.

Simon, S. (2022) "Russian law bans journalists from calling Ukraine conflict a 'war' or an 'invasion'", *NPR*, 5 March, [online], accessed 14 January 2023, https://www.npr.org/2022/03/05/1084729579/russian-law-bans-journalists-from-calling-ukraine-conflict-a-war-or-an-invasion

Stengel, R. (2019) *Information Wars: How we Lost the Global Battle Against Disinformation and what we can do about it*, New York City: Atlantic Monthly Press.

Terp, S.J., Breuer, P. (2022) "DISARM: a Framework for Analysis of Disinformation Campaigns", 2022 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), 6-10 June. Available at: https://ieeexplore.ieee.org/document/9830669

The Paris Call for Trust and Security in Cyberspace. (2018) The 9 Principles, 12 November, [online], accessed 5 January 2023, https://pariscall.international/en/principles

United Nations General Assembly. (2015) Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, 17th Session, UN Doc A/70/174, 22 July.

United Nations General Assembly. (2021) *Countering disinformation for the promotion and protection of human rights and fundamental freedoms*, Resolution 76/277, 24 December.

Viral Facts Africa. (2023) Viral Facts Africa, Twitter, [online], accessed 14 January 2023, https://twitter.com/viralfacts

Wanless, A., and Shapiro, J.N. (2021) Why Are Authoritarians Framing International Approaches to Disinformation? *LawFare Blog*, 28 December, [online], accessed 17 February 2022, https://www.lawfareblog.com/why-are-authoritarians-framing-international-approaches-disinformation

Wanless, A., and Shapiro, J.N. (2022) *A CERN Model for Studying the Information Environment*, Carnegie Endowment for International Peace, 17 November, [online], accessed 28 December 2022, https://carnegieendowment.org/2022/11/17/cern-model-for-studying-information-environment-pub-88408