**Article:**
Sander, B. and Tsagourias, N. (2020) The covid-19 infodemic and online platforms as intermediary fiduciaries under international law. Journal of International Humanitarian Legal Studies, 11 (2). pp. 331-347. ISSN 1878-1373

https://doi.org/10.1163/18781527-01102002

# The Covid-19 Infodemic and Online Platforms as Intermediary Fiduciaries under International Law

Barrie Sander
Fellow, Fundação Getúlio Vargas
barrie.sander@graduateinstitute.ch
Nicholas Tsagourias
Professor of International Law, University of Sheffield
Nicholas.tsagourias@sheffield.ac.uk

Moments of crisis, whether arising from terrorist attacks, financial meltdowns, or incipient pandemics, tend to trigger periods of heightened uncertainty and anxiety for affected communities. At such times, it is natural for people to come together in an attempt to make sense of their situation and to try to figure out how to respond. In the contemporary era, a major part of this process of 'collective sensemaking' takes place online.[1] The Covid-19 pandemic offers the most recent and arguably most striking illustration of the importance of online information during a period of crisis.[2] After all, human health depends not only on readily accessible health care, but also on 'access to accurate information about the nature of the threats and the means to protect oneself, one's family, and one's community'.[3]

During a speech delivered in mid-February 2020, the Director-General of the World Health Organization (WHO) observed that communities around the world were confronting not only the spread of the novel coronavirus, but also an 'infodemic' caused by an overabundance of information – some accurate, some not – that makes it challenging to identify trustworthy sources and reliable guidance about Covid-19.[4] At the epicentre of this infodemic are online platforms. Over the course of the past decade, a small number of platforms have grown to become dominant and essential channels of online communication for a wide range of services,

---

1 Kate Starbird, 'Reflecting on the Covid-19 Infodemic as a Crisis Informatics Researcher' (*Medium*, 9 March 2020) < https://onezero.medium.com/reflecting-on-the-covid-19-infodemic-as-a-crisis-informatics-researcher-ce0656fa4d0a> accessed 17 May 2020.

2 While the present paper focuses on challenges associated with the online information ecosystem, it is important to recognise that the COVID-19 crisis has amplified a wide range of well-established controversies associated with the online environment, ranging from Internet shutdowns and the digital divide to intrusive data surveillance and hostile cyberattack operations. See, for example, Barrie Sander and Luca Belli, 'COVID-19, Cyber Surveillance Normalisation and Human Rights Law' (*Opinio Juris*, 1 April 2020) <http://opiniojuris.org/2020/04/01/covid-19-symposium-covid-19-cyber-surveillance-normalisation-and-human-rights-law/> accessed 17 May 2020; Laura DeNardis and Jennifer Daskal, 'Society's dependence on the internet: 5 cyber issues the coronavirus lays bare' (*The Conversation*, 27 March 2020) < https://theconversation.com/societys-dependence-on-the-internet-5-cyber-issues-the-coronavirus-lays-bare-133679> accessed 17 May 2020.

3 United Nations Human Rights Office of the High Commissioner, 'COVID-19: Governments must promote and protect access to and free flow of information during pandemic' (*OHCHR*, 19 March 2020) < https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25729&LangID=E> accessed 17 May 2020.

4 World Health Organisation, 'Munich Security Conference' (*WHO*, 15 January 2020) <https://www.who.int/dg/speeches/detail/munich-security-conference> accessed 17 May 2020). See also, World Health Organisation, 'Managing Epidemics: Key Facts About Major Deadly Diseases' (*WHO*, 2018) 34 (defining an 'infodemic' as 'the rapid spread of information of all kinds, including rumours, gossip and unreliable information').

including marketplaces (e.g. Amazon), social networking (e.g. Facebook), search (e.g. Google), image-sharing (e.g. Instagram), video-sharing (e.g. YouTube), and microblogging (e.g. Twitter). Fuelled by surveillance-based business models, platforms are not passive conduits of online information, but active governors of user-generated content,[5] influencing the categories of content that are allowed and prohibited (*permissibility*), as well as how content is ranked, amplified, and organised (*visibility*).[6]

In this short reflection, we identify different dimensions of the Covid-19 infodemic (**1**), examine how platform governance has evolved in response to the crisis (**2**), and reflect on what the Covid-19 crisis reveals about the relationship between online platforms, international law, and the prospect of regulation (**3**), before offering some concluding remarks (**4**).

## 1 The Covid-19 Infodemic

Ushering in a world of social distancing and self-isolation, the global spread of Covid-19 has intensified societal reliance on the internet in general, and online platforms in particular. During this period of growing digital dependency, how online platforms govern user-generated content has taken on a heightened significance. When the Director-General of the WHO referred to the dangers posed by the Covid-19 infodemic, he failed to specify the different types of information challenges associated with online platforms during the crisis. Drawing on a conceptual framework developed by Claire Wardle and Hossein Derakhshan, it is possible to distinguish three types of 'information disorder':[7] *dis*information, *mis*information, and *mali*nformation.

*Disinformation* refers to the intentional creation and/or dissemination of verifiably false or misleading information, typically by organised state or non-state actors.[8] The motives underpinning disinformation campaigns tend to be varied, ranging from sowing discord or exploiting societal fears to interfering with public policies or securing an economic advantage – whether directly or indirectly.[9] In the context of the Covid-19 crisis, coordinated disinformation campaigns have sought to frame vulnerable minorities as the cause of the pandemic, and to fuel distrust in the ability of public health institutions to respond effectively

---

5 Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018) 131 *Harvard Law Review* 1598.

6 Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media* (Yale University Press 2018) 18.

7 Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making* (Council of Europe 2017) 20. For a different but equally useful typology, see Marko Milanovic, 'Viral Misinformation and the Freedom of Expression: Part II' (*EJIL Talk!*, 13 April 2020) < https://www.ejiltalk.org/viral-misinformation-and-the-freedom-of-expression-part-ii/> accessed 17 May 2020 (distinguishing viral misinformation in terms of content, source (state actors, organized non-state actors, individuals acting spontaneously or organically), target audience (in-groups and out-groups), and motives (sincere or insidious)).

8 European Commission, *Communication – Tackling Online Disinformation: A European Approach* (COM, 26 April 2018) 3-4.

9 AccessNow, 'Fighting Misinformation and Defending Free Expression During COVID-19: Recommendations for States' (*Access Now*, April 2020) 11 < https://www.accessnow.org/cms/assets/uploads/2020/04/Fighting-misinformation-and-defending-free-expression-during-COVID-19-recommendations-for-states-1.pdf> accessed 17 May 2020.

to the crisis.[10] Importantly, the precise narrative promoted as part of a disinformation operation will typically vary depending on the target audience.[11] Russian disinformation campaigns targeting domestic audiences, for example, have tended to describe the novel coronavirus as a form of foreign aggression, whereas those targeting international audiences have generally focused on conspiracy theories about 'global elites' weaponizing or exploiting the virus for their own ends.

Although disinformation remains an important concern with respect to the novel coronavirus, a leaked report by the European External Action Service concluded that 'the more pressing challenge' for public health in this context has been *misinformation* – namely, the unintentional spread of false or misleading information.[12] During fast-paced crisis situations, it is not uncommon for experts to take extra care with their public messaging in an effort to ensure accuracy and reduce misinterpretation. Paradoxically, this cautious approach may result in an 'information vacuum' into which false or misleading information is ready to fill.[13] For example, research by the Reuters Institute examining a sample of 225 pieces of misinformation rated false or misleading by fact-checkers from January through to the end of March 2020, found that 88% of the sample appeared on social media platforms, 59% of the sample involved forms of reconfiguration where true information had been spun, reworked, or recontextualised, and the largest category of false or misleading claims (appearing in 39% of the sample) concerned the actions or policies of public authorities, including government and international bodies like the WHO.[14] Importantly, the spread of Covid-19 misinformation has not been without consequence. Baseless claims linking next generation 5G mobile technology to the novel

---

10 EUvsDiSiNFO, 'EEAS Special Report Update: Short Assessment of Narratives and Disinformation around the COVID-19 Pandemic' (*EUvsDiSiNFO.eu*, 1 April 2020) <https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid-19-pandemic/> accessed 17 May 2020.

11 EUvsDiSiNFO, 'EEAS Special Report: Disinformation on the Coronavirus – Short Assessment of the Information Environment' (*EUvsDiSiNFO.eu*, 19 March 2020) <https://euvsdisinfo.eu/eeas-special-report-disinformation-on-the-coronavirus-short-assessment-of-the-information-environment/> accessed 17 May 2020. See also, Sean Martin McDonald and Xiao Mina, 'Coronavirus Crisis Pushes States to Quarantine Online Information' (*Foreign Policy*, 14 February 2020) <https://foreignpolicy.com/2020/02/14/wuhan-virus-censorship-coronavirus-crisis-pushes-states-quarantine-online-information/> accessed 17 May 2020 (distinguishing 'Nationalist (Consolidator)', 'Nationalist (Projector)' and 'Digital Influencer' behaviour).

12 Jennifer Rankin, 'Russian media "spreading Covid-19 disinformation"' (*The Guardian*, 18 March 2020) <https://www.theguardian.com/world/2020/mar/18/russian-media-spreading-covid-19-disinformation> accessed 17 May 2020.

13 S Harris Ali and Fuyuki Kurasawa, '#COVID19: Social media both a blessing and a curse during coronavirus pandemic' (*The Conversation*, 22 March 2020) <https://theconversation.com/covid19-social-media-both-a-blessing-and-a-curse-during-coronavirus-pandemic-133596> accessed 17 May 2020. Alternatively, an information vacuum may result from governments actively deploying broad and vague laws and/or more informal pressure to incentivise online platforms to collaterally censor content out of fear of legal or political liability. See, for example, Karman Lucero, 'China Responds to the Coronavirus with an Iron Grip on Information Flow' (*Lawfare*, 17 March 2020) <https://www.lawfareblog.com/china-responds-coronavirus-iron-grip-information-flow> accessed 17 May 2020 (discussing how China's structures of content control have hindered the flow of important information concerning the coronavirus and how to stem its spread, whilst also encouraging the active creation and dissemination of false or misleading information to fill the resulting information vacuum).

14 Scott Brennen et al., 'Types, Sources, and Claims of COVID-19 Misinformation' (*Reuters Institute for the Study of Journalism*, April 2020) <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation> accessed 17 May 2020.

coronavirus pandemic, for example, have contributed to real-world social harms, including petrol bomb attacks on telephone poles.[15]

Finally, *malinformation* refers to the intentional creation and/or dissemination of information that is threatening, abusive, discriminatory, harassing or disruptive, which aims to cause harm to a person, organisation or state.[16] Since the Covid-19 pandemic broke out, there have been reports of heightened racist and xenophobic sentiments in many parts of the world, with a proliferation of hate speech and stigmatization on online platforms.[17] An analysis conducted by Al Jazeera, for example, identified thousands of posts on Twitter employing racist terminology to describe the novel coronavirus.[18] The spread of malinformation can generate a number of harms; not only social stigma and the silencing of members of vulnerable groups in society, but also discriminatory treatment and acts of violence against them.[19]

While it is useful to distinguish different types of information disorder for analytical purposes, it is important to remember that in practice they tend to overlap and operate in tandem. False and misleading information about links between 5G mobile technology and Covid-19, for example, appears to have first emerged and spread organically as misinformation, before later being amplified as part of organised disinformation campaigns.[20]

## 2 The Online Platform Response to the Covid-19 Infodemic

In recent years, online platforms have witnessed a spate of controversies concerning issues ranging from data harvesting and surveillance to online censorship and influence operations. Given this increasingly hostile climate, it is notable that measures implemented by online platforms in response to the Covid-19 crisis have generated some rare positive headlines for the tech sector, even leading some to question whether the novel coronavirus has 'killed the techlash'.[21] Regardless of one's perspective on that question, it is undeniable that online platforms have responded to the Covid-19 infodemic by updating their content policies in various ways.

### *2.1 Partnerships and Collaboration*

Online platforms were quick to recognise the importance of partnerships and collaboration in responding to the Covid-19 crisis. For instance, a wide range of platforms have forged

---

15 Jim Waterson and Alex Hern 'How false claims about 56 health risks spread into the mainstream' (*The Guardian*, 7 April 2020) < https://www.theguardian.com/technology/2020/apr/07/how-false-claims-about-5g-health-risks-spread-into-the-mainstream> accessed 17 May 2020.

16 Chris Tenove et al, *Digital Threats to Democratic Elections: How Foreign Actors Use Digital Techniques to Undermine Democracy* (Center for the Study of Democratic Institutions 2018) 22-25.

17 Article19, *Viral Lies: Misinformation and the Coronavirus* (*Article19* March 2020) 4 < https://www.article19.org/wp-content/uploads/2020/03/Coronavirus-final.pdf> accessed 17 May 2020.

18 Eoghan Macquire, 'Anti-Asian hate continues to spread online amid COVID-19 pandemic' (*Al Jazeera*, 5 April 2020) <https://www.aljazeera.com/news/2020/04/anti-asian-hate-continues-spread-online-covid-19-pandemic-200405063015286.html> accessed 17 May 2020.

19 AccessNow (n 9) 16-18.

20 Milanovic (n 7).

21 Steven Levy, 'Has the Coronavirus Killed the Techlash?' (*WIRED*, 20 March 2020) < https://www.wired.com/story/plaintext-has-the-coronavirus-killed-the-techlash/> accessed 17 May 2020.

*partnerships with the WHO and other public health institutions* to promote authoritative and reliable information about Covid-19. Facebook and Instagram, for example, have been showing educational pop-ups connecting people to expert health organizations such as the WHO, as well as local health authorities, whenever anyone searches for information related to the novel coronavirus or taps on a Covid-19 related hashtag.[22] Facebook has also launched a Covid-19 Information Center, which sits at the top of the news feed in several countries and features real-time updates from national and global health authorities.[23] Similarly, Twitter has established a dedicated Covid-19 Event page and implemented a Covid-19 search prompt in partnership with the WHO and national public agencies in more than 70 countries to ensure that any search for information related to Covid-19 is met with credible and authoritative content.[24]

Online platforms have also *collaborated with each other*, though the details of such arrangements have remained relatively vague to date.[25] For instance, Facebook, Google, LinkedIn, Microsoft, Reddit, Twitter and YouTube, have released a joint industry statement confirming that they are 'working closely together on Covid-19 response efforts', including 'jointly combating fraud and misinformation about the virus, elevating authoritative content…, and sharing critical updates in coordination with government healthcare agencies around the world'.[26]

*2.2 Moderation Rules and Policies*

Beyond partnerships, online platforms have also revised their moderation rules and policies in an effort to address the Covid-19 infodemic. For instance, a range of platforms have *updated their moderation rules concerning organic content*. Twitter, for example, has broadened its definition of 'harm' in order to remove content that contradicts guidance from authoritative sources of global and local public health information, including tweets that encourage people not to social distance, promote harmful treatments or protection measures, or deny established scientific facts about transmission.[27] Similarly, Facebook has confirmed that it is removing false or misleading Covid-19 content 'as an extension of [its] existing policies to remove content that could cause physical harm'.[28] Concurrently, it continues to work with a network of over 60 fact-checking partners to reduce distribution and to show warning labels with more

---

22 Nick Clegg, 'Combating COVID-19 Misinformation Across Our Apps' (*Facebook Newsroom*, 25 March 2020) < https://about.fb.com/news/2020/03/combating-covid-19-misinformation/> accessed 17 May 2020.

23 Ibid.

24 Twitter Inc., 'Coronavirus: Staying safe and informed on Twitter' (*Twitter Blog*, 3 April 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 May 2020.

25 Inter-platform collaboration concerning the governance of online content is not new, with existing partnerships including varying degrees of cooperation concerning the removal of child exploitation material, terrorist and extremist content, and coordinated inauthentic behaviour. See generally, Evelyn Douek, 'The Rise of Content Cartels' (*Knight First Amendment Institute*, 11 February 2020) < https://knightcolumbia.org/content/the-rise-of-content-cartels> accessed 17 May 2020.

26 Nick Statt 'Major tech platforms say they're "jointly combating fraud and misinformation" about COVID-19' (*The Verge*, 16 March 2020) < https://www.theverge.com/2020/3/16/21182726/coronavirus-covid-19-facebook-google-twitter-youtube-joint-effort-misinformation-fraud> accessed 17 May 2020.

27 Twitter Inc (n 24).

28 Kang-Xing Jin, 'Keeping People Safe and Informed About the Coronavirus' (*Facebook Newsroom*, 9 April 2020) <https://about.fb.com/news/2020/05/coronavirus/> accessed 17 May 2020.

context in front of posts that are false but do not directly result in physical harm.[29] The impact of these policies was demonstrated when Twitter and Facebook decided to remove videos posted by Brazilian President Jair Bolsonaro in which he endorsed hydroxychloroquine as an effective treatment of Covid-19,[30] despite their long-standing reticence to take action against content posted by state leaders.[31]

Online platforms have also *updated their moderation rules concerning paid content*. For example, to protect against inflated prices and predatory behaviour Facebook has temporarily banned ads intended to create a panic or for products that claim to guarantee a cure or prevent people from contracting Covid-19.[32] The company has also temporarily banned ads and commerce listings for medical face masks, hand sanitizer, disinfecting wipes, and Covid-19 testing kits.[33] In addition, Facebook has committed to giving the WHO as many free ads as they need and millions in ad credits to other health authorities.[34] Similarly, Twitter has prohibited ads with distasteful references to Covid-19, sensational or panic-inducing content, inflated prices for products related to Covid-19, and certain products such as facemasks and alcohol hand sanitizers.[35] Google has also blocked hundreds of thousands of ads attempting to capitalize on the Covid-19 pandemic and announced a temporary ban on all ads for medical masks and respirators.[36]

### *2.3 Enforcement Challenges*

The stricter content measures implemented by online platforms appear to reflect not only the clear and present social harm that may result from information disorder concerning Covid-19, but also the fact that in many societies there seemed to be, initially at least,[37] less partisan disagreement than is typically the case concerning many political discussions and debates.[38]

---

29 Guy Rosen, 'An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19' (*Facebook Newsroom*, 16 April 2020) < https://about.fb.com/news/2020/04/covid-19-misinfo-update/> accessed 17 May 2020.

30 Jack Goodman and Christopher Giles, 'Coronavirus and chloroquine: Is there evidence it works?' (*BBC*, 28 April 2020) < https://www.bbc.com/news/51980731> accessed 17 May 2020.

31 Kim Lyons, 'Twitter removes tweets by Brazil, Venezuela presidents for violating COVID-19 content rules' (*The Verge*, 30 March 2020) < https://www.theverge.com/2020/3/30/21199845/twitter-tweets-brazil-venezuela-presidents-covid-19-coronavirus-jair-bolsonaro-maduro> accessed 17 May 2020.

32 Clegg (n 22).

33 Jin (n 28).

34 Ibid.

35 Twitter Inc (n 24).

36 Sundar Pichai, 'COVID-19: How we're continuing to help' (*The Keyword*, 15 March 2020) < https://www.blog.google/inside-google/company-announcements/covid-19-how-were-continuing-to-help/> accessed 17 May 2020.

37 See, for example, Frank Jordans and Elena Becatoros 'Many wary of virus reopenings as partisan divide grows in U.S.' (*CTV News*, 22 April 2020) < https://www.ctvnews.ca/health/coronavirus/many-wary-of-virus-reopenings-as-partisan-divide-grows-in-u-s-1.4906794> accessed 17 May 2020 (discussing how the US is beset with 'increasingly partisan disagreements over how and when to restart its economy'); Sam Adler-Bell, 'Facebook Is Removing Protest Pages. That's a Terrible Precedent' (*Medium*, 24 April 2020) <shorturl.at/ptL29 > accessed 17 May 2020 (discussing partisan disagreements in the US in response to Facebook's removal of 'certain event pages for in-person rallies against coronavirus lockdowns in California, New Jersey, and Nebraska'); and 'Coronavirus: Brazil's Bolsonaro joins anti-lockdown protests' (*BBC*, 20 April 2020) < https://www.bbc.com/news/world-latin-america-52351636> accessed 17 May 2020 (discussing anti-lockdown protests in Brazil).

38 Brennen et al (n 14) 7.

Notwithstanding these more stringent content policies, online platforms have nonetheless been confronted by two notable challenges concerning their enforcement.

First, online platforms have had to contend with *a significant reduction in the capacity of their human content moderators* due to the logistical and privacy challenges of moderators working from home as a result of the Covid-19 crisis. Consequently, platforms have had to temporarily increase their use of machine learning and automated systems to detect and remove violating content and disable accounts. Significantly, platforms have been relatively candid about the fact that their automated systems sometimes lack the ability to accurately assess the context of content compared to human content moderators, leading to a higher number of mistakes than usual.[39]

Second, online platforms have had to contend with the fact that *substantial numbers of users are turning to private or invite-only areas of their sites* to connect with the communities they care about – spaces that tend to be more difficult for platforms to moderate. Research by *POLITICO*, for example, has identified the spread of falsehoods across more than 30 invite-only Facebook groups dedicated to Covid-19, some of which have garnered tens of thousands of members.[40] Facebook has implemented a number of measures to address this challenge, including an educational pop-up directing group members to credible information from health organizations, prompts to group admins to share live broadcasts about Covid-19 from health authorities, and a curriculum that group admins can share with members to learn how to stay safe during the crisis. Nonetheless, false or misleading information about Covid-19 remains an ongoing challenge within these more private spaces of platforms.[41]

Taken together, these developments reveal not only the reactive nature of online platform governance, but also the complexity and impossibility of content moderation at scale. As Evelyn Douek has observed, confronted by the inevitability of error in addressing the Covid-19 infodemic, online platforms have chosen 'to err on the side of false positives and removing more content'.[42] In making this choice about error preference, platforms reveal 'the trade-offs between accuracy, comprehensive enforcement and speed [that] are inherent in *every* platform rule and not just in these exceptional moments'.[43] It is the capacity of platforms to make such choices, to determine how different interests should be balanced, that constitutes their power over how information circulates in the public domain and across the world.

## 3 Online Platforms as Intermediary Fiduciaries under International Law

---

39 See, for example, Jin (n 28); Twitter Inc (n 24).

40 Mark Scott, 'Facebook's private groups are abuzz with coronavirus fake news' (*POLITICO*, 30 March 2020) <https://www.politico.eu/article/facebook-misinformation-fake-news-coronavirus-covid19/> accessed 17 May 2020.

41 Jin (n 28).

42 Evelyn Douek, 'COVID-19 and Social Media Content Moderation' (*Lawfare*, 25 March 2020) <https://www.lawfareblog.com/covid-19-and-social-media-content-moderation> accessed 17 May 2020.

43 Ibid (emphasis in original).

The centrality of online platforms to global information flows, both during moments of crisis like Covid-19 and everyday life, raises the question of how their role and behaviour may be understood from an international legal perspective, and consequently what standards and regulatory schemes should inform their actions. In this final section, we seek to contribute to the debate currently taking place concerning whether online platforms should be regulated, how they should be regulated, and what form and content such regulation might take.

In our opinion, online platforms are intermediary fiduciaries of the international public good,[44] and for this reason regulation should be informed by relevant standards that apply to fiduciary relationships.[45]

*3.1 Fiduciary Relationships: States, People and the Public Good of Health*

A fiduciary relationship arises when a party is entrusted by another party, the entrustor, to serve her needs and deliver goods for her benefit. Fiduciary relations emerge because of certain social conditions, such as status, dependency, differentiated resources, or expertise.[46] Fiduciary relations can take various forms and may arise in different legal settings. They can be broad in nature or served by an agent entrusted with wide powers. Alternatively, they can be more limited, confined to a particular service or good provided by a specialised agent. The ultimate and broadest fiduciary relationship is that between the state and its people.

According to social contract theory,[47] individuals entrust the state with powers to pursue public goods such as security, health, and welfare, because they do not have the ability and resources to enjoy and share the benefits of these goods in a constant and non-exclusionary manner. Therefore, they enter into a fiduciary relationship with the state, such that the state's *raison d'être* and sovereign authority are linked to serving the people and delivering public goods. As Vatel wrote, 'the government was intrusted to him [the sovereign] only for the happiness of society, … [and] he uses the public power only with a view to the public welfare'.[48] This also means that the exercise of state power should be subject to certain standards and rules that derive from and aim at maintaining that fiduciary relationship. More specifically, the exercise of state power should not be self-interested, should take a holistic view of the public good, and should not be abused in view of the fact that the state is the ultimate fiduciary and power holder and individuals depend on the state for satisfying their needs. For this reason, the state accepts

---

44 For a definition of public goods, see Raymond Guess, *Public Goods, Private Goods* (Princeton University Press 2001).

45 Our perspective complements the work of Jack Balkin on 'information fiduciaries' in the domestic sphere. See, for example, Jack Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation' (2018) 51 UCDLR 1149, 1160-1163; Jack Balkin, 'Free Speech is a Triangle' (2018) 118 CLR 2040, 2054; Jack Balkin, 'Information Fiduciaries and the First Amendment' (2016) 49 UCDLR (2016) 1183, 1205-1209, 1221-1230. On fiduciary relations in domestic law see L.S. Sealy, 'Fiduciary Relationship' (1962) 20 CLJ 69, 81 and Tamar Frankel, 'Fiduciary Law' (1983) 71 CLR 795, 836.

46 *Plowright v Lambert* (1885) 52 LT 646, 652

47 Jean Jacques Rousseau, *The Social Contract* (Penguin Books 1968); John Locke, *Two Treatises of Government* (Everyman 1993) ch 2, 7 and 8.

48 Emer de Vattel, *The Law of Nations, or, Principles of the Law of Nature Applied to the Conduct and Affairs of Nations and Sovereigns* (Liberty Fund 2008) ch IV [39].

limitations to the exercise of its power, with human rights law providing an example of limitations imposed on the state as a fiduciary.[49]

In the absence of a global sovereign, States – individually or collectively – also become fiduciaries of humanity in delivering international public goods.[50] Although international law recognises and accepts the national social contract and the right of States to pursue the public good as defined by their people, it also caters for the international pubic good because the national and international public good are interconnected and interdependent. States act in this instance as fiduciaries of the global community of peoples and by promoting the international public good, they also satisfy the national public good. One method of promoting the international public good is through an international organisation which acts in that instance as an intermediary fiduciary.

Health is an international public good in the sense that it is a general and non-exclusive good with respect to which everyone is both a stakeholder and beneficiary.[51] Global health requires collective action which is pursued through an international organisation, the WHO, for the benefit of all human beings and, distinctly, for the benefit of States and for each state's population.[52] Information disorder – whether in the form of disinformation, misinformation, or malinformation – can hinder or thwart the delivery of the international public good of health by creating confusion, doubt, division, insularity, or exclusion, as well as by preventing action or undermining the state and the international institutions whose mandate is to serve this public good. In essence, information disorder has the potential to undermine the fiduciary relationships outlined above.

*3.2 Online Platforms as Intermediary Fiduciaries of the International Public Good of Health*

By adopting a range of measures to address information disorder, online platforms become intermediary fiduciaries. They become intermediary fiduciaries because they interpose themselves between States on the one hand, which are fiduciaries of their own people and the national public good of health and, collectively, fiduciaries of humanity and the international public good of health, and on the other, the global community of peoples. In this way, online platforms claim for themselves a distinct and indeed important fiduciary role in the pursuit of the international public good of health, a role that cannot otherwise be justified or legitimised.

---

49 On state disinformation concerning Covid-19 and human rights law, see generally, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, 'Disease Pandemics and the Freedom of Opinion and Expression' (23 April 2020) UN Doc A/HRC/44/49 [44]-[50]; Milanovic (n 7); AccessNow (n 9); Article19 (n 17).

50 Eyal Benvenisti, 'Sovereigns as Trustees of Humanity: On the Accountability of States to Foreign Stakeholders' (2013) 107 AJIL 295; Evan Criddle and Evan Fox-Decent, *Fiduciaries of Humanity: How International Law Constitutes Authority* (Oxford University Press 2016) ch 1; Eyal Benvenisti, 'Foreword: Upholding Democracy amid the Challenges of New Technology: What Role for the Law of Global Governance?' (2018) 29 *EJIL* 9.

51 Inge Kaul, Isabelle Grunberg, and Marc Stern, *Global Public Goods: International Cooperation in the 21st Century* (Oxford University Press 1999).

52 Constitution of the World Health Organization (signed 22 July 1946) 14 UNTS 185 (entered into force 7 April 1948) (Protocol).

Although there is no formal delegation, their characterisation as intermediary fiduciaries is justified by the fact that platforms have assumed such power in conjunction with the fact that States and people entrust them with such power.[53] However, it is not only how platforms project themselves or how they are perceived but, more crucially, it is the nature of their relationship with States and people that call for such a characterisation.[54] As it has been opined, 'discretion, power to act and vulnerability' can class a relationship as fiduciary and justify its regulation.[55]

Applying these conditions to online platforms, we can say that, first, they own and operate critical resources which individuals or States do not own or operate. These resources are indispensable for the functioning of the State and for the fulfilment of public goods. Accordingly, States and peoples are in a position of *vulnerability* (as far as the operation and management of these resources is concerned) due to the power differentials and the dependency relations that are created. This is even more so when online platforms rely upon elaborate technologies of data surveillance.

Secondly, online platforms *exercise power* in the sense that they can unilaterally modify human behaviour, adversely affect individuals' interests, and alter their factual and often legal relations and circumstances. Governing the permissibility and visibility of content amounts to an exercise of power because it affects participation and information which are important conditions for making autonomous decisions, participating in an equal and informed manner in public life, and the formulation and realisation of the public good. Even if their power may not be formally public – in the sense of authorised by law or enforced by public sanctions – it is nonetheless power in a functional sense: it is a means for attaining a certain goal.

Thirdly, the power of online platforms is *discretionary* in the sense that they make unilateral and individualised decisions on how the public good can be secured against a very broadly defined concept of the public good and without the participation of those affected by their decisions. It is also discretionary because of the relations of dependency and vulnerability mentioned above. Individuals are thus subject to the exercise of discretionary power and to the extent that decisions are automated, they are subject to the constant exercise of algorithmic power. This inflates even more the discretionary power of online platforms. It can also detach it from the pursuit of the public good when algorithms are used to make decisions that are not contextualised, individualised, or accounted for.

It becomes apparent from the above that fiduciary relations can be abused. The danger of abuse is even more serious in the case at hand. First, there are power differentials because, as

---

53 Tony Romm, 'White House ask Silicon Valley to help to combat coronavirus' (*The Washington Post*, 11 March 2020) <https://www.washingtonpost.com/technology/2020/03/11/white-house-tech-meeting-coronavirus/> accessed 17 May 2020; Charlotte Tobitt, 'NHS enlist help of social media platforms to tackle coronavirus fake news' (*Press Gazette*, 10 March 2020) < https://www.pressgazette.co.uk/nhs-enlists-help-of-social-media-platforms-to-tackle-coronavirus-fake-news/> accessed 17 May 2020.

54 *Chirnside v Fay* [2006] NZSC 68 [75].

55 United Kingdom Law Commission, *Fiduciary Duties and Regulatory Rules* (Law Commission Consultation Paper No 124, 1992) [2.4.6].

explained above, online platforms own the infrastructure and possess the technical expertise. Second, online platforms have their own interests and their own narrowly defined fiduciaries in the person of their shareholders – which do not necessarily align with the pursuit of the international public good. Third, the use of algorithms may produce biases and make the exercise of power discriminatory or unequal. Fourth, although it is a general principle of law that discretionary power should not be delegated, online platforms routinely delegate power to algorithms, a particularly dangerous practice when their decision-making processes cannot be explained or understood which is necessary for review and accountability.[56] As noted earlier, this practice has become even more prevalent during the Covid-19 infodemic due to reductions in the availability of human moderators. However, even if human moderators are available, questions may be asked about standards of training, the legal, social, cultural standards that apply, as well as levels of procedural transparency and accountability.[57] Finally online platforms can expand their power by instrumentalising the pursuit of the public good of health, causing collateral detriment to other public goods.

*3.3 Online Platforms and International Law*

Treating online platforms as intermediary fiduciaries of the international public good has important regulatory implications. It reveals that market regulation, contract regulation, and self-regulation are inadequate models of regulation in this context because they cannot address adequately the problems that arise from fiduciary relationships, particularly the problem of abuse. For this reason, regulation should be legal.

Regulation should be legal for four key reasons. Foremost, because law can authoritatively determine the structure and content of the fiduciary relationship. Second, it can moderate and balance often contradictory interests. Third, it can identify the goods to be served and how they will be served. Finally, regulation can provide mechanisms to prevent or address abuses of power, thus establishing accountability. In fact, individuals already resort to legal institutions to obtain protection if their status or rights have been abused by online platforms and when they feel that platform procedures are unsatisfactory. Even so, existing legal impediments confirm the need for smarter legal regulation.

Such regulation may be comprehensive or limited to specific services and goods, such as health, but it should ensure that online platforms exercise their entrusted power in good faith, with due care, and within the bounds of the aims for which it has been entrusted. Regulation should also ensure that in exercising their fiduciary powers, online platforms respect the human dignity and needs of individuals and the personality and needs of States.

---

[56] For discussion of the concerns raised by the use of automation in content moderation, see generally: Hannah Bloch-Wehba, 'Automation in Moderation,' [2020] CILJ (forthcoming); Robert Gorwa, Reuben Binns, and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' [2020] *Big Data & Society* 7; Emma Llansó et al, 'Artificial Intelligence, Content Moderation, and Freedom of Expression' (*Transatlantic Working Group*, 26 February 2020) < https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf> accessed 17 May 2020.

[57] For an ethnographic study of the commercial content moderation industry, see Sarah Roberts. *Behind the Screen: Content Moderation in the Shadows of Social Media* (Yale University Press 2019).

Since the role of online platforms as intermediaries concerns the international public good, it is the community of States, which should lay down these rules. This means that regulation should be international rather than national and should comprise of a minimum international law framework.[58] Equally, however, States should be able to contextualise domestic regulation with reference to domestic circumstances and needs, provided they respect the minimum international law standard. And, online platforms should be able to introduce their own community standards so long as they align with the minimum international and the relevant national regulatory standards. In this regard, it is interesting to note that certain online platforms already look to international human rights standards to make content-related decisions.[59]

Although the scope and content of such a regulatory framework requires more detailed analysis – which is beyond the confines of this short piece – the main takeaway is that applying a fiduciary framework to online platforms can recast the relationship between online platforms, peoples, and States under a different light. Doing so provides an indication of the ends, modes and content of regulation, as well as insight into how their power should be disciplined, and how online platforms should be made accountable.

## 4 Conclusion: Beyond Platform Responsibility

Nearly twenty years ago, Hilary Charlesworth observed how international lawyers 'revel in a good crisis', which often provides 'a focus for the development of the discipline and … also allows international lawyers the sense that their work is of immediate, intense relevance'.[60] Writing in the midst of the Covid-19 pandemic, we believe that the present crisis does indeed provide an opportunity to catalyse regulation of online platforms under international law.[61] With this in mind, we have put forward a fiduciary model of regulation, which, we argue, provides a sound basis for determining how power can be distributed and how interests can be balanced among states, online platforms, and peoples in order to avoid abuse, attain the common good, and protect human dignity.

---

58 See generally, Nicholas Tsagourias, 'The Rule of Law in Cyberspace: a Hybrid and Networked Concept?' [2020] ZaöRV (forthcoming). For discussion of human rights-based approaches to online platform regulation, see generally, Barrie Sander, 'Democratic Disruption in the Digital Age: Social Media Platforms, Cyber Governance and Human Rights Law' (forthcoming); and Barrie Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation' [2020] FIJL (forthcoming).

59 See, for example, Monika Bickert, 'Updating the Values That Inform Our Community Standards' (*Facebook Newsroom* 12 September 2019) <https://about.fb.com/news/2019/09/updating-the-values-that-inform-our-community-standards/> accessed 17 May 2020 ('In some cases, we allow content which would otherwise go against our Community Standards – if it is newsworthy and in the public interest. We do this only after weighing the public interest value against the risk of harm, and *we look to international human rights standards to make these judgments*') (emphasis added).

60 Hilary Charlesworth, 'International Law: A Discipline of Crisis', (2002) 65 *The Modern Law Review* 377, 377.

61 See, in this regard, Sundhya Pahuja and Jeremy Baskin, 'Never Waste a Crisis: A Practical Guide?' (*Critical Legal Thinking*, 20 March 2020) < https://criticallegalthinking.com/2020/03/20/never-waste-a-crisis-a-practical-guide/> accessed 17 May 2020 ('We know we should never waste a crisis […] Now is a good time for progressives to think about what we should encourage and what we should resist during the current crisis and in the reconstruction which will follow […] Whilst supporting measures to contain the virus, progressives need to ensure the longer-term outcomes from this crisis are better').

At the same time, we are also mindful of Charlesworth's warning that international law may become a mask for 'issues of structural justice that underpin everyday life'.[62] Reflecting on the online platform ecosystem, it is apparent that many of the existing concerns that arise online are symptoms of deeper structural problems – including social, economic and political inequalities that have been confronting societies around the world for generations. We, therefore, wish to conclude by emphasising that while the recognition of online platforms as intermediary fiduciaries under international law constitutes an important step in addressing contemporary challenges associated with the digital public sphere, it is essential that attention not be diverted from the broader distributional and societal divisions that underpin these challenges across the world.

[62] Charlesworth (n 60) 391.