2021

Friction to Fight Misinformation: Content-Neutral Frictive Measures Under International Law

Robert J. Garcia

FRICTION TO FIGHT MISINFORMATION:

CONTENT-NEUTRAL FRICTIVE MEASURES UNDER INTERNATIONAL LAW

INTRODUCTION

The Special Rapporteur on the Freedom of Opinion and Expression[1] (Special Rapporteur) is the position created by the U.N. Human Rights Council to address violations of the freedoms of opinion and expression, undertake fact-finding visits to countries, and publish annual reports relating to the freedom of opinion and expression.[2] In the Special Rapporteur's upcoming report on Disinformation and the freedom of expression,[3] they should advocate the use of content neutral frictive measures in combating misinformation on social media.

The first section of this paper provides a synopsis of the Special Rapporteur's six reports on the freedom of opinion and expression in the Information, Communication, Technology sector, including Social Media Companies. The first section also discusses the Special Rapporteur's most recent report on disease pandemics. The second section details the international legal framework based on the International Covenant on Civil and Political Rights and the United Nations Guiding Principles on Business which is reiterated in the Special Rapporteur's reports. The third section defines frictive measures and what the Special Rapporteur has said about them in their reports. The fourth section argues that content neutral frictive measures comply international law, despite

[1]United Nations Commission on Human Rights, *QUESTION OF THE HUMAN RIGHTS OF ALL PERSONS SUBJECTED TO ANY FORM OF DETENTION OR IMPRISONMENT*, U.N. Doc. E/CN.4/1993/L.48 (1993) (establishing the mandate of the Special Rapporteur on the and protection of the right to freedom of opinion and expression.).

[2] Office of the High Commissioner for Human Rights, https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/mandate.aspx (last visited May 2, 2021)

[3] Irene Kahn (Special Rapporteur on the promotion and protection of freedom of opinion and expression), *Disinformation and freedom of opinion and expression*, U.N. Doc. A/HRC/47/25 (Forthcoming June, 2021).

the potential for bias and present a viable method of combating disinformation which the Special Rapporteur should advocate in their upcoming report.

I. SYNOPSIS OF THE SPECIAL RAPPORTEUR REPORTS

Since 2015, the Special Rapporteur has published six reports on freedom of opinion and freedom of expression in the Information and Communication Technology sector.[4] The Information and Communication Technology sector consists of the private actors involved in "organizing, accessing, populating and regulating the Internet."[5] Social media companies make up part of the ICT sector.[6] The Special Rapporteur has written reports on the freedoms of opinion and expression in the ICT sector pertaining to: encryption and anonymity;[7] role of states and the private sector in the digital age;[8] the role of digital access providers;[9] online content regulation;[10] artificial intelligence;[11] and online hate speech.[12] In response to the COVID-19 pandemic, the Special

[4] David Kaye (Special Rapporteur on the promotion and protection of freedom of opinion and expression), *Online Hate Speech*, ¶ 3, U.N. Doc. A/74/486 (Oct. 19, 2019).
[5] David Kaye (Special Rapporteur on the promotion and protection of freedom of opinion and expression), *Freedom of Expression, States and the Private Sector in the Digital Age*, ¶ 15, U.N. Doc. A/HRC/32/38 (May 11, 2016) [hereinafter *States and the Private Sector in the Digital Age*].
[6] Kaye, *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 1.
[7] David Kaye (Special Rapporteur on the promotion and protection of freedom of opinion and expression), *The Use of Encryption and Anonymity to Exercise the Rights to Freedom of Opinion and Expression in the Digital Age*, U.N. Doc. A/HRC/29/32 (May 22, 2015) [hereinafter *Encryption and Anonymity*]
[8] Kaye, *States and the Private Sector in the Digital Age*, *supra* note 5.
[9] David Kaye (Special Rapporteur on the promotion and protection of freedom of opinion and expression), *The Role of Digital Access Providers*, U.N. Doc. A/HRC/35/22 (May 30, 2017).
[10] David Kaye (Special Rapporteur on the promotion and protection of freedom of opinion and expression), *Online Content Regulation*, U.N. Doc. A/HRC/38/35 (Apr. 6, 2018).
[11] David Kaye (Special Rapporteur on the promotion and protection of freedom of opinion and expression), *Artificial Intelligence Technologies and Implications for the Information Environment*, U.N. Doc. A/73/348 (Aug. 29, 2018) [hereinafter *Artificial Intelligence*].
[12] David Kaye (Special Rapporteur on the promotion and protection of freedom of opinion and expression), *Online Hate Speech*, U.N. Doc. A/74/486 (October 9, 2019).

Rapporteur published in April of 2020 on disease pandemics which discussed public health disinformation online.[13]

Encryption and Anonymity:

The Special Rapporteur's report on Encryption and Anonymity considers whether the rights to privacy, freedom of opinion and expression, include the ability to encrypt or anonymize.[14] The Rapporteur defines encryption as "a mathematical 'process of converting messages, information, or data into a form unreadable by anyone except the intended recipient…'"[15] Anonymity is defined as the "condition of avoiding identification."[16] The report addressed concerns that encryption and anonymity would hide criminal activity online.[17] While the report acknowledges that corporations have a duty to operate under the UN Guiding Principles on Business, the report specifically "is focused on State obligations."[18] Ultimately, encryption and anonymity are deemed "necessary for the exercise of the right to freedom of opinion and expression in the digital age."[19]

States and the Private Sector in the Digital Age:

In A/HRC/32/38, the Special Rapporteur acknowledges the influence of private actors on freedom of expression through the internet and social media.[20] The report raises questions concerning the protection of freedom of opinion and expression as relating to the private sector

[13] David Kaye (Special Rapporteur on the promotion and protection of freedom of opinion and expression), *Disease Pandemics and the Freedom of Opinion and Expression*, U.N. Doc. A/HRC/44/49 (Apr. 23, 2020) [hereinafter *Disease Pandemics*].
[14] *Encryption and Anonymity*, *supra* note 7, at ¶ 3.
[15] *Encryption and Anonymity*, *supra* note 7, at ¶ 7.
[16] *Encryption and Anonymity*, *supra* note 7, at ¶ 9.
[17] *Encryption and Anonymity*, *supra* note 7, at ¶ 13.
[18] *Encryption and Anonymity*, *supra* note 7, at ¶ 17.
[19] *Encryption and Anonymity*, *supra* note 7, at ¶ 56.
[20] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 3.

online.[21] Finally, the report concludes with the Special Rapporteur concludes by offering some "normative guidance" in the most needed areas of the ICT sector.[22]

The Special Rapporteur questions companies' terms of service,[23] and raises derivative concerns from the lack of clarity in terms of service including "inconsistent enforcement",[24] "overzealous censorship",[25] "lack of an appeals process",[26] and states' opportunistically using such ambiguity to remove objectionable content.[27] For private actors, the primary source of regulation on their platforms is their terms of service.[28] These terms of service are formulated in such a way that makes it difficult for users to predict what content is restricted on the private actor's platform.[29] Moving to content regulation, the Special Rapporteur examines the "design and engineering choices" of social media, recognizing that social media companies curation of content affects user's access to information.[30] Such inconsistent enforcement and limitation of information poses threats to the freedom of opinion and expression.[31]

The Role of Digital Access Providers:

Beginning in 2016, the Special Rapporteur began the process of detailing the different facets of the "information and communications technology (ICT) sector."[32] The report outlines state obligations, under international law, to protect expression online.[33] The Rapporteur condemns

[21] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 3.
[22] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 3.
[23] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 37.
[24] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 52.
[25] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 52.
[26] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 52.
[27] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 53.
[28] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 52.
[29] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 52.
[30] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 55.
[31] *States and the Private Sector in the Digital Age*, *supra* note 5, at ¶ 52.
[32] *The Role of Digital Access Providers*, *supra* note 9, at ¶ 4.
[33] *The Role of Digital Access Providers*, *supra* note 9, at ¶ 5.

internet and telecommunications shutdowns[34] and cautions that accessing online user information can interfere with privacy rights.[35] The role of digital access providers is also discussed, and the Rapporteur recognizes that access as critical to the freedom of expression.[36]

Online Content Regulation:

The Special Rapporteur's report on online content regulation "…focuses on the regulation of user-generated content, principally by States and social media companies…"[37] Moderation describes "the process by which Internet companies determine whether user-generated content meets the standards articulated in their terms of service and other rules."[38] The report expresses concerns about national laws restricting speech across borders.[39] Because many social media companies have an international user base, "national laws are inappropriate for companies that seek common norms."[40] In lieu of national laws, the Special Rapporteur advocates compliance with the UN Guiding Principles on Business as an international framework for the content moderation policies of social media companies.[41]

The UN Guiding Principles on Business "and their accompanying body of 'soft law'",[42] inform the Rapporteur's substantive standards for content moderation.[43] When developing standards, companies should seek policy that allows platforms "…for users to develop opinions, express themselves freely and access information of all kinds in a manner consistent with human

[34] *The Role of Digital Access Providers, supra* note 9, at ¶ 8.
[35] *The Role of Digital Access Providers, supra* note 9, at ¶ 17.
[36] *The Role of Digital Access Providers, supra* note 9, at ¶ 29.
[37] *Online Content Regulation, supra* note 10 at, ¶ 3.
[38] *Online Content Regulation, supra* note 10 at, ¶ 3.
[39] *Online Content Regulation, supra* note 10 at, ¶ 18.
[40] *Online Content Regulation, supra* note 10 at, ¶ 41.
[41] *Online Content Regulation, supra* note 10 at, ¶ 6.
[42] *Online Content Regulation, supra* note 10 at, ¶ 42.
[43] *Online Content Regulation, supra* note 10 at, ¶ 44,

rights law."[44] Part of content moderation policy should include a standard of non-discrimination which requires companies to "transcend formalistic approaches" and "take into account the concerns of communities historically at risk of censorship and discrimination."[45]

The report ends with the recommendations of Special Rapporteur, including recognition that human rights law is "the authoritative global standard for ensuring freedom of expression" on social media,[46] and "smart" content moderation as opposed to "heavy-handed viewpoint-based regulation."[47]

AI and Free Speech:

The Special Rapporteur's report on AI and Free Speech has three goals: "define key terms essential to a human rights discussion about AI; identify the human rights legal framework relevant to AI; and present some preliminary recommendations to ensure that, as the technologies comprising AI evolve, human rights considerations are baked into that process."[48] Artificial Intelligence is defined as "a 'constellation' of processes and technologies enabling computers to complement or replace specific tasks otherwise performed by humans, such as making decisions and solving problems."[49] The Rapporteur urges caution against removing human intervention from content moderation through the promotion of artificial intelligence.[50]

[44] *Online Content Regulation, supra* note 10 at, ¶ 45.
[45] *Online Content Regulation, supra* note 10 at, ¶ 48.
[46] *Online Content Regulation, supra* note 10 at, ¶ 70.
[47] *Online Content Regulation, supra* note 10 at, ¶ 66.
[48] *Artificial Intelligence, supra* note 11 at, ¶ 2.
[49] *Artificial Intelligence, supra* note 11 at, ¶ 3.
[50] *Artificial Intelligence, supra* note 11 at, ¶ 6.

Concerning AI and Social Media, the Special Rapporteur notes areas of concern in content display, personalization,[51] content moderation and removal.[52] Because Social media companies tailor content based on user preferences, users might experience a total absence of "diverse views, interfering with individual agency to seek and share ideas and opinions across ideological, political or societal divisions."[53] The Special Rapporteur reaffirms their advocacy for an international legal framework for Social Media companies to follow, calling attention to the right of freedom of opinion which "requires freedom from undue coercion."[54]

Online Hate Speech:

The Special Rapporteur's report on Hate Speech Online discusses "Governments considering regulatory options and companies determining how to respect human rights online."[55] While the report offers no concrete definition of Hate Speech, the Rapporteur cites ICCPR Article 20 Section, which prohibits "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence".[56] Supplementing the definition, the report cites the 2013 Rabat Plan definitions of "'Hatred' and "hostility" refer to intense and irrational emotions of opprobrium, enmity and detestation towards the target group."[57] The Rapporteur clarifies that hateful expression does not always "constitute advocacy or incitement."[58]

Social Media companies are criticized in the report for operating with a business model that values the spread of hateful content.[59] Simultaneously, the Special Rapporteur warns of

[51] *Artificial Intelligence, supra* note 11 at, ¶ 10.
[52] *Artificial Intelligence, supra* note 11 at, ¶ 13.
[53] *Artificial Intelligence, supra* note 11, at ¶ 12.
[54] *Artificial Intelligence, supra* note 11, at ¶ 23.
[55] *Online Hate Speech, supra* note 12, at ¶ 2.
[56] *Online Hate Speech, supra* note 12, at ¶ 8.
[57] *Online Hate Speech, supra* note 12, at ¶ 6.
[58] *Online Hate Speech, supra* note 12, at ¶ 20.
[59] *Online Hate Speech, supra* note 12, at ¶ 40.

ambiguous definitions of hate speech in Social Media policy and of automated content moderation that "is notoriously bad at evaluating context."[60] To improve their moderation policies, the Rapporteur recommends "de-amplification, de-monetization, education, counter-speech."[61]

Disease Pandemics:

In the wake of the COVID-19 Pandemic, the Special Rapporteur issued a report detailing five challenges to freedom of opinion and expression.[62] One of the five challenges discussed is the spread of Public Health Disinformation.[63] The Special Rapporteur recognized the need to engage "rumors in order to correct them" whilst cautioning against disproportionate punishment for sharing disinformation.[64] The Special Rapporteur also acknowledged Social Media's "enormous impact on public discourse and the rights of individuals on and off their platforms."[65] To assist in stemming the flow of disinformation, Social Media companies "should aim towards maximum transparency of their policies and engage" with both public authorities and affected communities.[66]

II.INTERNATIONAL LAW FRAMEWORK

The Special Rapporteur calls upon social media companies to tailor their content moderation policy based on an international law framework. National laws are an insufficient basis for content moderation policy because Social Media Companies involve a "geographically and culturally diverse user base."[67] The Special Rapporteur presents an international law framework

[60] *Online Hate Speech, supra* note 12, at ¶ 50.
[61] *Online Hate Speech, supra* note 12, at ¶ 58.
[62] *Disease Pandemics, supra* note 13, at ¶ 6.
[63] *Disease Pandemics, supra* note 13, at ¶ 41.
[64] *Disease Pandemics, supra* note 13, at ¶ 42.
[65] *Disease Pandemics, supra* note 13, at ¶ 52.
[66] *Disease Pandemics, supra* note 13, at ¶ 52.
[67] *Online Content Regulation, supra* note 10 at, ¶ 41.

in their reports based on the International Covenant of Civil and Political Rights,[68] and the United Nations Guiding Principles on Business and Human Rights.[69] The Special Rapporteur uniformly cites ICCPR Art. 19, General Comment 34 to Article 19, and UN Guiding Principles on business for social media companies and the ICT sector in general.[70]

International Covenant of Civil and Political Rights, Art. 19

The purpose of the ICCPR, explained in its preamble, is to promote "civil and political freedom…"[71] The ICCPR was adopted by the United Nations General Assembly in 1966.[72] Article 19 of the ICCPR protects the rights to freedom of expression and freedom of opinion.[73]

As part of the rights to hold opinions and freedom of expression, Article 19 delineates important qualifications clarifying the breadth of those rights.[74] Not only does the Art. 19 protect the right to hold opinions, but to "hold opinions without interference."[75] The right to freedom of expression under Art. 19, includes "freedom to seek, receive and impart information and ideas of all kinds…through any other media of his choice."[76] However, Art. 19 also delineates when restrictions to the freedom of expression are acceptable.

Art. 19 allows derogation to protect the rights and reputation of others, "national security…public order…public health or morals."[77] Under ICCPR Art. 4, states may derogate from

---

[68] UN General Assembly, *International Covenant on Civil and Political Rights*, 16 December 1966, United Nations, Treaty Series, vol. 999, p. 171 [hereinafter ICCPR].
[69] UN Office of High Commissioner on Human Rights, *Guiding Principles on Business and Human Rights*, U.N. Doc. HR/PUB/11/04 (June 16, 2011) [hereinafter Guiding Principles].
[70] *See States and the Private Sector in the Digital Age*, *supra* note 2, at ¶ 5 and ¶ 9; *The Role of Digital Access Providers*, *supra* note 6, at ¶ 5 and ¶ 45; *Online Hate Speech*, *supra* note 9 at ¶ 5 and ¶ 47.
[71] *ICCPR*, *supra* note 68, at 1.
[72] *ICCPR*, *supra* note 68, at 1.
[73] *ICCPR*, *supra* note 68, at art. 19.
[74] *ICCPR*, *supra* note 68, at art. 19.
[75] *ICCPR*, *supra* note 68, at art. 19.
[76] *ICCPR*, *supra* note 68, at art. 19.
[77] *ICCPR*, *supra* note 68, at art. 19(3)(a)-(b).

their obligations under the treaty excepting some articles.[78] Such derogation cannot be inconsistent with international law and cannot "involve discrimination solely on the ground of race, colour, sex, language, religion or social origin."[79] Restrictions under Art. 19 must be "provided by law and…necessary."[80] Restrictions can be applied to the freedom of expression, but not the freedom of opinion.[81] To further clarify the extent of protections under Art. 19 and the requirements for restrictions on the freedom of expression, the Human Rights Committee of the United Nations issued General Comment No. 34.[82]

General Comment No. 34

General Comment No. 34 was published in July 2011.[83] While the ICCPR binds state parties, the obligations under Art. 19 require states to protect people "from any acts by private persons or entities" infringing upon the freedoms of opinion and expression.[84]

Comment No. 34 addresses the right to freedom of expression in media,[85] to access information,[86] to political rights,[87] and the scope of restrictions to expression.[88] "Internet-based modes of expression" are protected by Art. 19. The Comment reiterates that no restriction upon the holding of an opinion nor impairments on other rights because one holds an opinion are

[78] *ICCPR*, *supra* note 68, at art. 4(1).
[79] *ICCPR*, *supra* note 68, at art. 4(1).
[80] *ICCPR*, *supra* note 68, at art. 19.
[81] *See* U.N. Human Rights Committee, *General Comment No. 34, Article 19 Freedom of Opinion and Expression*, ¶ 9, U.N. Doc. CCPR/C/GC/34 (Sept. 12, 2011) [hereinafter *General Comment No. 34*]; *States and the Private Sector in the Digital Age*, *supra* note 2 at ¶ 7.
[82] *General Comment No. 34*, *supra note* 81, at ¶ 5.
[83] *General Comment No. 34*, *supra note* 81, at ¶ 5.
[84] *General Comment No. 34*, *supra note* 81, at ¶ 7..
[85] *General Comment No. 34*, *supra note* 81, at ¶ 13.
[86] *General Comment No. 34*, *supra note* 81, at ¶ 18.
[87] *General Comment No. 34*, *supra note* 81, at ¶ 20.
[88] *General Comment No. 34*, *supra note* 81, at ¶ 37.

permissible.[89] Part of the freedom to hold opinions is the prohibition on coercion to hold or not hold an opinion.[90]

Comment No. 34 clarifies the requirements of Art. 19(3) restrictions of the freedom of expression.[91] In addition to the requirements of legality, for the legitimate purposes listed in 19(3)(a-b), and the restrictions must be necessary and proportional.[92]

The Comment goes into detail about what it means for a restriction to be required by law.[93] To be a law, means that a regulation is not merely discretionary, is sufficiently precise, and publicly available to allow people to tailor their conduct accordingly.[94]

The extent of legitimate grounds—or legitimacy—for the restrictions to the freedom of expression are also clarified in the Comment.[95] Restrictions are permissible to protect rights under international law, both to individuals and members of a community.[96] While national security is a legitimate purpose for a restriction on expression, states must take care to ensure such restrictions take "extreme care" to conform with the requirements of Art. 19.[97] Finally, the comment explains that the legitimate grounds must no stem from any single social, philosophical or religious tradition.[98]

[89] *General Comment No. 34, supra note* 81, at ¶ 9.
[90] *General Comment No. 34, supra note* 81, at ¶ 10.
[91] *General Comment No. 34, supra note* 81, at ¶ 35.
[92] *General Comment No. 34, supra note* 81, at ¶ 13.
[93] *General Comment No. 34, supra note* 81, at ¶ 24.
[94] *General Comment No. 34, supra note* 81, at ¶ 25.
[95] *General Comment No. 34, supra note* 81, at ¶ 28.
[96] *General Comment No. 34, supra note* 81, at ¶ 28.
[97] *General Comment No. 34, supra note* 81, at ¶ 30.
[98] *General Comment No. 34, supra note* 81, at ¶ 32.

For a restriction to be proportionate, it must not be overbroad.[99] Considering the form of expression restricted along with its method of dissemination, the restriction must be the least intrusive option to be proportionate.[100]

To show the necessity of a restriction on the freedom on expression, there must be "a direct and immediate connection between the expression and the threat."[101]

The Comment also outlines "certain specific areas" where restrictions should have a limited scope and are subject to greater scrutiny.[102] The value placed on uninhibited expression relating to political discourse is particularly high by the ICCPR.[103] Accordingly, the Comment expresses the Committee's concern for restrictions relating to political discourse.[104] Concerning websites and blogs, the Committee explains that restrictions by the state need to meet the Art. 19 requirements and cannot be based on prohibiting material critical of the government.[105] While Art.19 and its subsequent commentary detail state obligations including protections against private actors infringing upon the freedom of expression, the UN Guiding Principles directly address business' role in protecting human rights.

UN Guiding Principles on Business and Human Rights

The UN Guiding Principles outline "The State Duty to Protect Human Rights", "The Corporate Responsibility to Respect Human Rights", and "Access to Remedy."[106] In addition to reaffirming states' duty to protect against human rights abuse,[107] when business enterprises are

[99] *General Comment No. 34, supra note* 81, at ¶ 34.
[100] *General Comment No. 34, supra note* 81, at ¶ 34.
[101] *General Comment No. 34, supra note* 81, at ¶ 34.
[102] *General Comment No. 34, supra note* 81, at ¶ 37.
[103] *General Comment No. 34, supra note* 81, at ¶ 38.
[104] *General Comment No. 34, supra note* 81, at ¶ 38.
[105] *General Comment No. 34, supra note* 81, at ¶ 43.
[106] *Guiding Principles, supra note 69* at iii.
[107] *Guiding Principles, supra note 69* at 1.

owned or controlled by the states, additional steps to protect against human rights abuse and oversight are encouraged.[108] Although the principles do not have the force of law, they provide a method of enhancing business practices to protect human rights.[109]

The UN Guiding Principles state that "business enterprises should respect human rights."[110] For business to respect human rights at a minimum, they must not violate the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights.[111] Respecting such rights includes preventing negative impacts to human rights through their business activities as well as from business relationships.[112] The responsibility to protect human rights regardless of the size of the business enterprise.[113] However, businesses can craft their means of addressing and protecting human rights according to their size and circumstances.[114]

The Guiding Principles, recommend a policy commitment by businesses to protect human rights.[115] The policy should be holistic, recognized at "the most senior level of the business", informed by expertise, made publicly available and reflected through operational policies.[116] Part of the policy should include a due diligence assessment to determine the potential adverse human rights impact of business practices.[117] Upon assessment, business should implement the findings "across relevant and internal functions" affecting decision-making.[118] The effectiveness of

[108] *Guiding Principles, supra note 69* at 4.
[109] *Guiding Principles, supra note 69* at 1.
[110] *Guiding Principles, supra note 69* at princ. 11.
[111] *Guiding Principles, supra note 69* at princ. 12.
[112] *Guiding Principles, supra note 69* at princ. 13.
[113] *Guiding Principles, supra note 69* at princ. 14.
[114] *Guiding Principles, supra note 69* at princ. 15.
[115] *Guiding Principles, supra note 69* at princ. 16.
[116] *Guiding Principles, supra note 69* at princ. 16.
[117] *Guiding Principles, supra note 69* at princ. 17.
[118] *Guiding Principles, supra note 69* at princ. 19.

response should be tracked by the business.[119] Communication between stakeholders effected by the human rights impact of businesses is expected.[120] When businesses come to learn that their practices have had adverse impacts on human rights, "they should provide for or cooperate in their remediation.[121]

Finally, the UN Guiding Principles outline principles to provide access to remedy for persons experiencing an adverse impact on human rights from business practices.[122] States have duty to provide remedy "through judicial, administrative, legislative, or other appropriate means."[123] Businesses are recommended to have "effective operational-level grievance mechanisms" so effected persons and their communities can have direct access to remedy.[124] Operational-level grievance mechanisms help identify potential adverse impacts of business practices along with allowing grievances to be addressed and remedy before they escalate.[125] These mechanisms need not require a grievance that reaches the level of a human rights abuse, but rather should seek to act when legitimate concerns are identified.[126]

Special Rapporteur on International Law

The Special Rapporteur identifies the freedoms protected in Article 19 as the basis for a wide range of human rights and are the foundation for free and democratic society.[127] The language

[119] *Guiding Principles, supra note 69* at princ. 20.
[120] *Guiding Principles, supra note 69* at princ. 21.
[121] *Guiding Principles, supra note 69* at princ. 22.
[122] *Guiding Principles, supra note 69* at princ. 25.
[123] *Guiding Principles, supra note 69* at princ. 25.
[124] *Guiding Principles, supra note 69* at princ. 29.
[125] *Guiding Principles, supra note 69* at princ. 29.
[126] *Guiding Principles, supra note 69* at princ. 29.
[127] *See Online Hate Speech*, *supra* note 4, at ¶ 5; *Encryption and Anonymity*, *supra* note 4, at ¶ 22.

of Art. 19 is consciously devoid of a list of relevant media because the rights of freedom of opinion and expression accommodated future technological advances.[128]

The Special Rapporteur uses the framework of Article 19 of the ICCPR to determine the conditions for restrictions on the freedom of expression to comply with international law.[129] The conditions are legality, necessity and proportionality, and Legitimacy.[130] The usefulness, reasonableness, and desirability of a restriction is neither dispositive nor sufficient to demonstrate any of requirements for the restriction under Article 19.[131] The requirements for the restrictions are "to be applied strictly and in good faith, with robust and transparent oversight.[132] The burden of justifying the restriction falls on the authority imposing the restriction rather than the speakers subject to it.[133]

A/HRC/29/32 includes a discussion on how the storage, transmission, and security of information in the digital age effect the right to freedom of opinion.[134] Because State and non-state actors have control over the storage, transmission, and security of information online, they both have the ability to interfere with rights in Art. 19 of the ICCPR. The right to freedom of opinion includes the ability to come to an opinion through reasoning, prohibiting undue coercion.[135] Offering preferential treatment to induce acceptance of an opinion could also constitute impermissible coercive conduct violating Art. 19.[136] The Special Rapporteur recognized in

[128] *Encryption and Anonymity, supra* note 4, at ¶ 26.
[129] *Online Content Regulation, supra* note 10 at, ¶ 66; *Artificial Intelligence, supra* note 11, at ¶ 28; *Online Hate Speech, supra* note 12, at ¶ 6; *Disease Pandemics, supra* note 13, at ¶ 11.
[130] *Online Content Regulation, supra* note 10 at, ¶ 7; *Artificial Intelligence, supra* note 11, at ¶ 28; *Online Hate Speech, supra* note 12, at ¶ 6; *Disease Pandemics, supra* note 13, at ¶ 11.
[131] *Disease Pandemics, supra* note 13, at ¶ 15.
[132] *Online Hate Speech, supra* note 12, at ¶ 7.
[133] *Online Hate Speech, supra* note 12, at ¶ 6.
[134] *Encryption and Anonymity, supra* note 7, at ¶ 12.
[135] *Artificial Intelligence, supra* note 11, at ¶ 23.
[136] *Artificial Intelligence, supra* note 11, at ¶ 23.

A/73/348 that content curation effects the ability to formulate opinions and "raises novel questions about the types of coercion or inducement that may be considered an interference."[137]

In recognizing the impact of ICT's, the Special Rapporteur has consistently reiterated that companies should apply international law and "apply human rights principles in their operations."[138] Rather than applying throughout business practices, most companies apply human rights principles in response to demands from the states.[139] In the A/HRC/38/35, the Special Rapporteur succinctly articulates the UN Guiding Principles' minimum standards for business practices as an operative framework for companies to use:

> *(a) Avoid causing or contributing to adverse human rights impacts and seek to prevent or mitigate such impacts directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts (principle 13);*
>
> *(b) Make high-level policy commitments to respect the human rights of their users (principle 16);*
>
> *(c) Conduct due diligence that identifies, addresses and accounts for actual and potential human rights impacts of their activities, including through regular risk and impact assessments, meaningful consultation with potentially affected groups and other stakeholders, and appropriate follow-up action that mitigates or prevents these impacts (principles 17–19);*
>
> *(d) Engage in prevention and mitigation strategies that respect principles of internationally recognized human rights to the greatest extent possible when faced with conflicting local law requirements (principle 23);*
>
> *(e) Conduct ongoing review of their efforts to respect rights, including through regular consultation with stakeholders, and frequent, accessible and effective communication with affected groups and the public (principles 20–21);*
>
> *(f) Provide appropriate remediation, including through operational-level grievance mechanisms that users may access without aggravating their "sense of disempowerment" (principles 22, 29 and 31).*[140]

---

[137] *Artificial Intelligence*, *supra* note 11, at ¶ 24.
[138] *Online Content Regulation*, *supra* note 10 at, ¶ 9.
[139] *Online Content Regulation*, *supra* note 10 at, ¶ 10.
[140] *Online Content Regulation*, *supra* note 10 at, ¶ 11.

While this framework is not meant to carry the weight of law, it does provide an aspirational standard for businesses.[141] One area that social media companies can assess their impact on the right to freedom of expression is the effect of frictionless sharing,[142] on their platforms and how it leads to misinformation.[143]

III.FRICTIVE MEASURES DEFINED

In media policy, there exists a concept called the "signal-to-noise ratio."[144] Signal represents truthful information that supports democratic discourse.[145] Noise represents content that misinforms and undermines discourse.[146] As various kinds of content are amplified by algorithms on social media, it can become difficult to distinguish between signal and noise.[147] Friction can be introduced to help distinguish between the two.

Friction can be defined as the force that resists between two bodies.[148] To resist can be defined as "to exert force in opposition."[149] Taken together, friction can be understood as the force exerted in opposition to the relative motion between two bodies in contact. In digital platforms seeking to streamline the flow of information, friction is an obstacle to be overcome.[150] However, whether friction is good or bad depends upon the force it is acting against.[151] Friction can act "to disincentivize and disrupt practices that addict, surveil, and dull critical functions."[152] Frictive

---

[141] *Guiding Principles, supra note 69* at pg. 1.
[142] William McGeveran, *The Law of Friction*, 2013 U. CHI. LEGAL F. 15 at 19 (2013).
[143] Ellen P. Goodman, *Digital Information Fidelity and Friction*, Knight First Amendment Institute, 2, (2020) https://s3.amazonaws.com/kfai-documents/documents/c5cac43fec/2.27.2020_Goodman-FINAL.pdf.
[144] *Digital Information Fidelity and Friction*, *supra* 143 at 2.
[145] *Digital Information Fidelity and Friction*, *supra* 143 at 2.
[146] *Digital Information Fidelity and Friction*, *supra* 143 at 2.
[147] *Digital Information Fidelity and Friction*, *supra* 143 at 2.
[148] *The Law of Friction*, *supra* 142 at 15.
[149] Meriam-Webster, https://www.merriam-webster.com/dictionary/resisting (last visited May 5, 2021).
[150] *Digital Information Fidelity and Friction*, *supra* 143 at 20; *see also The Law of Friction*, *supra* 142 at 56.
[151] *The Law of Friction*, *supra* 142 at 56.
[152] *Digital Information Fidelity and Friction*, *supra* 143 at 3-4.

measures are forms of creating such friction and their goal is to "open pathways for reflection."[153] While not exhaustive, the Knight Institute has identified three forms of friction relevant to digital platforms: "communication delays, virality disruptions, and taxes."[154]

Communication delays are a form of frictive measures that seek to systematize a pause.[155] There is some research to suggest that people are "more likely to resist manipulative communications" when they have the time to "raise cognitive defenses."[156] In live broadcast media, short delays are already implemented for the sake of quality control.[157] In the stock market, the IEX ensure a degree of friction by running all trades through extra cable to avoid any trader from having an advantage by getting their information first.[158] A pause by way of communication delay gives people the time to think about the fidelity of information they are considering, to distinguish between signal and noise.[159]

Virality Disruptors are a form of frictive measure that disrupt traffic once "a certain threshold of circulation" is reached.[160] Virality can be understood as the quality of "triggering quick, emotionally, intense responses."[161] Disrupting virality would involve a pause both for the user to process the incoming information and also, to allow content moderation through human review to ensure compliance of the disrupted communication.[162] Moderation is defined as "the process by which internet companies determine whether user-generated content meets the

[153] *Digital Information Fidelity and Friction*, *supra* 143 at 3.
[154] *Digital Information Fidelity and Friction*, *supra* 143 at 21.
[155] *Digital Information Fidelity and Friction*, *supra* 143 at 21.
[156] *Digital Information Fidelity and Friction*, *supra* 143 at 21.
[157] *Digital Information Fidelity and Friction*, *supra* 143 at 21.
[158] *Digital Information Fidelity and Friction*, *supra* 143 at 21.
[159] *Digital Information Fidelity and Friction*, *supra* 143 at 21.
[160] *Digital Information Fidelity and Friction*, *supra* 143 at 22.
[161] Anthony Nadler, Matthew Crain, and Joan Donovan, *Weaponizing the Digital Influence Machine: The Political Perils of Online Ad Tech*, Data & Society, 32 (2018) https://datasociety.net/wp-content/uploads/2018/10/DS_Digital_Influence_Machine.pdf
[162] *Digital Information Fidelity and Friction*, *supra* 143 at 22.

standards articulated in their terms of service and other rules.[163] In the financial markets, a similar disruptive measure called a circuit-breaker is used to stop the flow of information that could overwhelm traders and contribute to instability.[164] When information flows result in a certain threshold of volatility in financial markets, regulators like the U.S. Securities and Exchange commission and the New York Stock Exchange can activate a circuit-breaker to cause a disruption. This disruption's purpose is to allow time to process information and make informed decisions, "to create the space for the exercise of cognitive autonomy."[165] Disruption to virality on digital platforms, allowing for time to process and, also for content moderation can limit the spread of noise.[166]

Taxes can act as frictive measures to encourage businesses to avoid boosting noise over signal.[167] Taxes can act as friction by making companies suffer a financial cost for monetizing the virality of communication.[168] The revenue from such taxes could be put to use by supporting the production of signal.[169]

Frictive Measures in the Special Rapporteur's Reports

The Special Rapporteur recommends that companies in the ICT sector recognize the standard for freedom of expression is not based on the law of any state nor private interest, but rather human rights law.[170] Additionally, they recommend that companies engage in "smart regulation" focused on transparency and remediation, allowing people to choose whether to engage

[163] *Online Content Regulation*, *supra* note 10 at, ¶ 3.
[164] *Digital Information Fidelity and Friction*, *supra* 143 at 22.
[165] *Digital Information Fidelity and Friction*, *supra* 143 at 23.
[166] *Digital Information Fidelity and Friction*, *supra* 143 at 22.
[167] *Digital Information Fidelity and Friction*, *supra* 143 at 23.
[168] *Digital Information Fidelity and Friction*, *supra* 143 at 23.
[169] *Digital Information Fidelity and Friction*, *supra* 143 at 23.
[170] *Online Content Regulation*, *supra* note 10 at, ¶ 70.

in digital platforms.[171] Companies should not engage in viewpoint-based regulation.[172] At a minimum content moderation of digital platforms should be based on the UN Guiding Principles on business and restrictions to the freedom of expression should conform to the requirements of: Legality, Necessity and Proportionality, and legitimacy.[173] The Special Rapporteur also recognized a standard of non-discrimination in content moderation which required transcending formalistic approaches and considering "communities historically at risk of censorship and discrimination."[174]

The Special Rapporteur has acknowledged that digital platforms of Internet companies operate with a business model that benefits from attention and virality.[175] Artificial Intelligence and algorithmic personalization are "optimized for engagement and virality at scale."[176] However, this preference for virality can threaten individual's ability to find some content.[177] Because digital platforms value virality, content with lower levels of engagement can be deprioritized making some content obscure.[178]

The Special Rapporteur recognizes that companies engage in content moderation and predicate access to their digital platforms upon compliance with user agreements and terms of service.[179] A major flaw in companies' content moderation practices is a lack of transparency.[180] Automated content moderation creates risks that content moderation might violate human rights

[171] *Online Content Regulation*, *supra* note 10 at, ¶ 66.
[172] *Online Content Regulation*, *supra* note 10 at, ¶ 66.
[173] *Online Content Regulation*, *supra* note 10 at, ¶ 45-49; *Online Hate Speech, supra* note 12, at ¶ 18-20.
[174] *Online Content Regulation*, *supra* note 10 at, ¶ 48.
[175] *Online Hate Speech, supra* note 12, at ¶ 40.
[176] *Artificial Intelligence, supra* note 11, at ¶ 12.
[177] *Artificial Intelligence, supra* note 11, at ¶ 12.
[178] *Artificial Intelligence, supra* note 11, at ¶ 12.
[179] *Online Content Regulation*, *supra* note 10 at, ¶ 12.
[180] *Online Hate Speech, supra* note 12, at ¶ 45.

law.[181] Some methods of company moderation that the Special Rapporteur has touched on include: automated flagging; automated removal; automated pre-publication filtering; user flagging of impermissible content; trusted flagging; human evaluation; action by the company against the account or content; notification; appeals and remedies.[182] Additionally, the Special Rapporteur expressed concern at the prospect of promoting counter-narrative to user communications on digital platforms.[183] Use of such counter-narratives could transform digital platforms into propaganda carriers.[184] While the Special Rapporteur has acknowledged a myriad of content moderation tools, some which might even have the effect of creating friction, they have not directly advocated the use of frictive measures.

Despite not expressly using the term frictive measures, the Special rapporteur has recognized that companies have content moderation tools that can restrict the virality of communications and include a "range of options short of deletion."[185] In countering the spread of disinformation—noise—on digital platforms, particularly in response to the COVID-19 pandemic, rumors must be addressed in order to correct them.[186] However, state penalization of disinformation is not proportionate and can deter communication of valuable information.[187]

IV. FRICTIVE MEASURES UNDER THE INTERNATIONAL LAW FRAMEWORK

Although states, not business, are parties to the International Covenant of Civil and Political Rights,[188] the international law framework outlined in the special rapporteur's reports can

[181] *Online Content Regulation*, *supra* note 10 at, ¶ 56.
[182] *Online Content Regulation*, *supra* note 10 at, ¶ 32-37.
[183] *Online Content Regulation*, *supra* note 10 at, ¶ 21.
[184] *Online Content Regulation*, *supra* note 10 at, ¶ 21.
[185] *Online Hate Speech, supra* note 12, at ¶ 51.
[186] *Disease Pandemics, supra* note 13, at ¶ 42.
[187] *Disease Pandemics, supra* note 13, at ¶ 42.
[188] *ICCPR, supra* note 65, at 1.

be applied to assess restrictions on the freedom of expression by social media companies. Frictive measures would have to satisfy the requirements of: Legality; Necessity and Proportionality; and Legitimacy.[189] Additionally, social media companies are covered under the UN Guiding Principles of Business and Human Rights.[190] Accordingly, frictive measures can be assessed under the framework of assessing restrictions to the freedom of expression under the ICCPR and recommendations of the UN Guiding Principles on Business.

Communication Delays

Systematizing a pause would be unlikely to violate either he freedoms under Art. 19 of the ICCPR or the Guiding Principles.

Legality

To be legal, a regulation must be publicly available and sufficiently precise for people to conform to the regulation.[191] Nothing about a communication delay would be illegal under the law because it would be content neutral meaning that people subject to the restriction would not have to change their behavior in a meaningful. So long as companies publicly inform users that they implement communication delays, such regulation would be publicly available.

Necessity and Proportionality

Communication delays comply with the requirement of necessity because of their benefit in reduction of noise on social media.[192] To demonstrate necessity, the body issuing the restriction

[189] *Online Content Regulation*, *supra* note 10 at, ¶ 7; *Artificial Intelligence, supra* note 11, at ¶ 28; *Online Hate Speech, supra* note 12, at ¶ 6; *Disease Pandemics, supra* note 13, at ¶ 11.
[190] *Guiding Principles, supra note 69* at princ. 11.
[191] *General Comment No. 34*, *supra* note 81, at ¶ 24.
[192] *Digital Information Fidelity and Friction*, *supra* 143 at 22.

to freedom of expression needs to show a direct and immediate connection between the kind of expression and the threat.[193] In the context of digital platforms, the threat is disproportionate noise in the signal to noise ratio.[194] Frictionless sharing degrades the quality of information being shared.[195] The immediacy connection comes with the virality of the communication that is promoted through business models thriving on such virality.[196] On most social media, publication of communication is almost instantaneous.[197] The restriction of a communication delay allows for the creation of time for cognitive processing before information is shared.[198]

Because delaying communication falls short of the most extreme restriction of deletion, it would likely qualify as proportionate. Considering the form of expression restricted along with its method of dissemination, the restriction must be the least intrusive option to be proportionate.[199] Because social media companies have the capacity to delete content from their platforms, any restriction short of deletion would be short of the most restrictive option.[200] While it is likely untenable to create an exhaustive list of potential restrictions and assign a level to intrusiveness to each, simply creating a delay in the message does not compare to silencing the message in its entirety through deletion.

Legitimacy

[193] *General Comment No. 34, supra* note 81, at ¶ 34.
[194] *Digital Information Fidelity and Friction, supra* 143 at 13.
[195] *The Law of Friction, supra* 142 at 46.
[196] *The Law of Friction, supra* 142 at 48.
[197] Catherine O'Regan, *Hate Speech Online: an (Intractable) Contemporary Challenge?*, Vo1. 71 Current Legal Problems 403, 416 (2018).
[198] *Digital Information Fidelity and Friction, supra* 143 at 21.
[199] *General Comment No. 34, supra* note 81, at ¶ 34.
[200] *Online Hate Speech, supra* note 12, at ¶ 51.

Communication delays qualify as legitimate restrictions to the freedom of expressions because they help protect rights under international law.[201] Some legitimate purposes for restrictions on the freedom of expression included in the language of Art. 19 are, "national security…public order…public health or morals." (Art. 19, pg 11, 3(a)(b)). The purpose of communication delays is to boost signal, truthful information that supports democratic discourse. (Digital Information Fidelity and Friction, pg 2). The promotion of truthful, democratic discourse would likely be considered furtherance of public order and accordingly, legitimate.

Virality Disruptors

For much of the same reasons as Communication Delays, Virality Disruptors would likely be permissible under Art. 19's international legal framework. However, because virality disruptors can be used to allow time for content moderation, there is increased likelihood for viewpoint discrimination. For this reason, it is valuable to distinguish between disrupting of the virality of content and disrupting virality for the purposes of determining whether a communication should be deleted. In this context, I choose to focus on virality disruptors "to create the space for the exercise of cognitive autonomy."[202]

Legality

So long as the implementation of virality disruptors on social media platforms is made publicly available and is sufficiently precise to allow individuals to tailor their conduct,[203] virality disruptors are legal under international law.

[201] *General Comment No. 34, supra* note 81, at ¶ 28.
[202] *Digital Information Fidelity and Friction*, *supra* 143 at 23.
[203] *General Comment No. 34, supra* note 81, at ¶ 24.

Necessity and Proportionality

Virality disruptors seek the same goal as communication delays, to reduce noise.[204] Whereas frictionless sharing increases noise,[205] the introduction of a pause at a certain threshold of circulation allows for time to "exercise of cognitive autonomy."[206] The expression of frictionless threatens to drown out signal with noise,[207] but a virality disruptor introduces friction in the hopes of increasing signal by allowing time to process. The immediate connection between the threat of noise and the expression of frictionless sharing should be sufficient to fulfill the necessity requirement of Art. 19.[208]

Because slowing the spread of information by way of virality disruptors falls short of deleting the content or censorship, there is a strong likelihood that such frictive measures would be proportional for the same reasons as communication delays.

Legitimacy

Virality disruptors help protect rights under international law by increasing the amount of signal, thereby supporting democratic discourse and public order. Instead of allow discourse on social media to be dominated by frictionless, quick, emotional responses, disrupting the virality of a communication allows more time for processing and increases the ability for cognitive choice. Cognitive choice helps democratic discourse by allowing people to make their intent a conscious

[204] *Digital Information Fidelity and Friction*, *supra* 143 at 22.
[205] *The Law of Friction*, *supra* 142 at 46.
[206] *Digital Information Fidelity and Friction*, *supra* 143 at 23.
[207] *The Law of Friction*, *supra* 142 at 46.
[208] *General Comment No. 34*, *supra* note 81, at ¶ 34.

decision as opposed to an assumed intent based on a frictionless architecture.[209] Supporting democratic discourse would benefit public order and sufficiently legitimize virality disruptors.[210]

Frictive Measures under the UN Guiding Principles

Whereas Art. 19 has an analytical framework which can be readily applied to restrictions on freedom of expression to determine if they comply with international law, the UN Guiding Principles provide for business to better comply with human rights law.[211] Frictive measures that are permissible under international law can be used to help businesses including social media companies comply with the UN Guiding Principles.

Frictive measures provide an opportunity for social media companies engage in "prevention and mitigation strategies" of the adverse impact their business can have on human rights[212] Frictionless sharing is built into the architecture of social media companies.[213] Frictionless sharing results in noise,[214] and noise undermines the discursive potential of democratic discourse.[215] Frictive measures prevent and mitigate noise overwhelming signal on social media.[216] Specifically, Communication Delays and frictive measures can prevent and mitigate noise on social media,[217] without viewpoint discrimination.[218]

[209] *The Law of Friction*, *supra* 142 at 47.
[210] *Digital Information Fidelity and Friction*, *supra* 143 at 2.
[211] *Guiding Principles, supra note 69* at 1.
[212] *Guiding Principles, supra note 69* at princ. 13 and 23; *see also Online Content Regulation*, *supra* note 10 at, ¶ 11.
[213] *The Law of Friction*, *supra* 142 at 19.
[214] *The Law of Friction*, *supra* 142 at 46.
[215] *Digital Information Fidelity and Friction*, *supra* 143 at 2.
[216] *Digital Information Fidelity and Friction*, *supra* 143 at 13.
[217] *Digital Information Fidelity and Friction*, *supra* 143 at 21-22.
[218] *Digital Information Fidelity and Friction*, *supra* 143 at 21.

While frictive measures will not satisfy all the UN Guiding Principles,[219] they provide an opportunity for social media companies to promote democratic discourse on their platforms and mitigate adverse impacts on human rights.

Concerns about AI

Despite the benefits of frictive measures, they are likely considered a form of artificial intelligence and are susceptible to the same kinds of discrimination posed by artificial intelligence. Artificial Intelligence is defined by the Special Rapporteur in, *A/73/348*, as a collection of technologies that allow computers to reinforce or replace tasks done by humans.[220] The Special Rapporteur has identified the "potential for AI to embed and perpetuate bias and discrimination…in the exercise of freedom of opinion and expression."[221] Discriminatory effects and bias in artificial intelligence are produced by the data sets in the design of the intelligence.[222] Additionally, the lack of transparency around the manner artificial intelligence effects the information environment prevents people "from understanding when and according to what metric information is disseminated, restricted or targeted."[223]

On social media, the content users see and information personalized for their viewing is often dictated by artificial intelligence.[224] Massive amounts of data including "browsing histories, semantic and sentiment analyses" are entered into algorithms to curate information displayed to users.[225] Social media companies use subjective assessments to gauge how interesting content will

[219] *Online Content Regulation*, *supra* note 10 at, ¶ 11.
[220] *Artificial Intelligence, supra* note 11, at ¶ 3.
[221] *Artificial Intelligence, supra* note 11, at ¶ 37.
[222] *Artificial Intelligence, supra* note 11, at ¶ 6.
[223] *Artificial Intelligence, supra* note 11, at ¶ 31.
[224] *Artificial Intelligence, supra* note 11, at ¶ 10.
[225] *Artificial Intelligence, supra* note 11, at ¶ 10.

be to a user, thereby limiting exposure to different perspective across ideological or political lines.[226] Artificial intelligence optimized for user engagement, promotes virality and demotes independent and user-generated content,[227], thus promoting a noisy, frictionless information environment.[228]

In addition to determining what content is seen by users, artificial intelligence is also involved in content moderation and removal.[229] The Special Rapporteur has raised concerns about AI-driven content moderation, expressly recognizing the limited ability for artificial intelligence to account for linguistic and cultural context.[230] The exclusion of information by AI-driven content moderation and removal "increases the risk of manipulation of individual users through the spread of disinformation" by limiting diverse perspectives.[231]

While nothing inherent to frictive measures requires that they be carried out through artificial intelligence, there is no reason computer technology cannot be used in implementing frictive measures. If measures like communication delays and virality disruptors are implemented using artificial intelligence, they are at risk for the kinds of discrimination and bias that threaten the freedoms of opinion and expression online.[232] While the potential for bias in implementing frictive measures is worth acknowledging, the potential bias does not necessarily change their legality under the international law framework reiterated by the Special Rapporteur.

CONCLUSION:

[226] *Artificial Intelligence, supra* note 11, at ¶ 12.
[227] *Artificial Intelligence, supra* note 11, at ¶ 12.
[228] *Digital Information Fidelity and Friction*, *supra* 143 at 3; *see also The Law of Friction*, *supra* 142 at 46.
[229] *Artificial Intelligence, supra* note 11, at ¶ 13.
[230] *Artificial Intelligence, supra* note 11, at ¶ 15.
[231] *Artificial Intelligence, supra* note 11, at ¶ 18.
[232] *Artificial Intelligence, supra* note 11, at ¶ 37.

Frictionless sharing on social media degrades information quality on social media.[233] Degradation of information quality increases noise which misinforms.[234] One way for the social media companies to combat misinformation on social media is to introduce frictive measures which would result in boosting signal by slowing information flows and allowing time to raise cognitive defenses.[235] Communication delays and virality disruptors present new frictions that can be implemented in a content neutral way.[236] Content neutral frictive measures are especially valuable because they most readily comply with the international framework identified by the Special Rapporteur, based on the ICCPR and UN Guiding Principles on Business.

In the last seven special rapporteur reports discussing the freedom of opinion and expression in the ICT sector, frictive measures have not been mentioned by name. While concerns about states using their power to restrict speech through private actors like social media companies is a very legitimate concern, so too is limiting the spread of disinformation. In the Special Rapporteur's upcoming report on Disinformation and freedom of opinion and expression, the Special Rapporteur should analyze and advocate the measures social media companies have at their disposal to combat disinformation. Content neutral frictive measures comply with international law and present a viable bulwark against the proliferation against disinformation online.

[233] *The Law of Friction*, *supra* 142 at 46.
[234] *Digital Information Fidelity and Friction*, *supra* 143 at 2.
[235] *Digital Information Fidelity and Friction*, *supra* 143 at 22.
[236] *Digital Information Fidelity and Friction*, *supra* 143 at 21.