

7-2020

In a World of “Fake News,” What’s a Social Media Platform to do?

Evelyn Mary Aswad

University of Oklahoma College of Law

Follow this and additional works at: <https://dc.law.utah.edu/ulr>



Part of the [Communications Law Commons](#), and the [First Amendment Commons](#)

Recommended Citation

Aswad, Evelyn Mary (2020) "In a World of “Fake News,” What’s a Social Media Platform to do?," *Utah Law Review*. Vol. 2020 : No. 4 , Article 4.

Available at: <https://dc.law.utah.edu/ulr/vol2020/iss4/4>

This Article is brought to you for free and open access by Utah Law Digital Commons. It has been accepted for inclusion in Utah Law Review by an authorized editor of Utah Law Digital Commons. For more information, please contact valeri.craigle@law.utah.edu.

IN A WORLD OF “FAKE NEWS,” WHAT’S A SOCIAL MEDIA PLATFORM TO DO?

Evelyn Mary Aswad*

Abstract

While the circulation of disinformation and misinformation online can pose a variety of risks to societies around the world, it should also be of concern that overreacting to such false information can undermine human rights, including freedom of expression. The business operations of global social media platforms frequently intersect with this latter concern because of a spike in the adoption of national laws that ban “fake news” as well as their own platform policies to tackle false information. This Essay assesses the corporate responsibility standards afforded by the United Nations’ Guiding Principles on Business & Human Rights as well as the International Covenant on Civil and Political Rights and explains several key ways in which the guidance that these instruments provide is relevant to social media companies in tackling false information on their platforms.

INTRODUCTION

*Strong protections for democratic freedoms are necessary to ensure that the internet does not become a Trojan horse for tyranny and oppression.*¹

In its 2019 report on internet freedom, Freedom House—a non-profit human rights organization—focused a spotlight on how governments and non-state actors throughout the world intentionally manipulate and undermine the connectivity that social media provides to infringe human rights and engage in election interference.² The report highlighted that such interference generally occurs in the form of three

* © 2020 Evelyn Mary Aswad. The author is the Herman G. Kaiser Chair in International Law and the Director of the Center for International Business & Human Rights at the University of Oklahoma’s College of Law (@EvelynAswad). Previously, she was the director of the human rights law office at the U.S. Department of State (2010–2013) and served as an attorney-adviser in that human rights law office from 2004–2009. The views are solely those of the author. The author thanks Bailey Betz for his research assistance.

¹ ADRIAN SHAHBAZ & ALLIE FUNK, FREEDOM HOUSE, FREEDOM ON THE NET 2019: THE CRISIS OF SOCIAL MEDIA 2 (2019), https://freedomhouse.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf [<https://perma.cc/W4YJ-N5UE>].

² *Id.* at 6–20. This report examines the situation in “65 countries around the globe, covering 87 percent of the world’s internet users. . . . [The countries] represent diverse geographical regions and regime types.” *Id.* at 4.

types of measures. First, state and private actors have engaged in *informational* measures, such as creating false information, developing effective dissemination networks for such information (including automated accounts), and infiltrating real social media accounts.³ Second, governments have engaged in *technical* measures to restrict online communications and connectivity, including restricting access to particular websites or the internet altogether.⁴ Third, governmental authorities have also deployed *legal* measures to punish online falsehoods, such as laws banning the defamation of rulers.⁵

With respect to *legal* measures, a recent troubling trend shows governments citing to the rise of disinformation and misinformation⁶ to justify new laws that prohibit online falsehoods and often require social media platforms to assist in implementing such measures.⁷ Within the one-year period from June 2017 through May 2018, “at least 17 countries approved or proposed laws that would restrict

³ *Id.* at 6–9. For example, in India, the Prime Minister sent millions of people “misleading and inflammatory content” through an app “marketed to all Indians as a way to keep up with official government news.” *Id.* at 8. In the Philippines, “political operatives spread information through closed groups on public platforms, where there is less content moderation. . . .” *Id.* They also paid online social media personalities who are “micro-influencers” to endorse particular politicians, which gave the impression of authentic endorsements and helped political operatives avoid campaign spending limits. *Id.*

⁴ *Id.* at 9. In Egypt, Cambodia, and Zimbabwe, the governments blocked access to websites that were critical of government officials. *Id.* In Bangladesh, the government “repeatedly restricted mobile internet service throughout the country prior to and on election day.” *Id.*

⁵ *Id.* at 10. Turkey, India, and Malawi arrested and/or brought criminal charges against those who engaged in online speech that defamed their rulers. *Id.* Bangladesh adopted a new law prohibiting propaganda while Thai officials issued vague rules about online discourse involving elections. *Id.*

⁶ This Essay uses the term “disinformation” as defined by Claire Wardle and Hossein Derakhshan: “information that is false and deliberately created to harm a person, social group, organization or country.” Samuel Spies, *Defining “Disinformation,”* MEDIAWELL (Oct. 22, 2019), <https://mediawell.ssrc.org/literature-reviews/defining-disinformation/versions/1-0/> [<https://perma.cc/7DNJ-HZXM>]. This Essay also uses Wardle and Derakhshan’s definition of “misinformation,” i.e., “information that is false, but not created with the intention of causing harm.” *Id.*

⁷ Allie Funk, *Citing ‘Fake News,’ Singapore Could Be Next to Quash Freedom of Expression*, JUST SECURITY (Apr. 8, 2019), <https://www.justsecurity.org/63522/citing-fake-news-singapore-could-be-next-to-quash-free-expression/> [<https://perma.cc/9QTJ-Y8R5>] (noting a regional trend in Asia to ban falsehoods, including in China, Cambodia, Malaysia, the Philippines, and Vietnam); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Freedom of Expression and Elections in the Digital Age*, 8–9, U.N. Research Paper 1/2019 (June 2019) [hereinafter *Elections in the Digital Age*], https://www.ohchr.org/Documents/Issues/Opinion/Elections_ReportDigitalAge.pdf [<https://perma.cc/M6EG-2CHY>] (expressing concern “with recent legislative and regulatory initiatives to restrict ‘fake news’ and disinformation” throughout the world, including in Italy, Malaysia, and France).

online media in the name of fighting ‘fake news’ and online manipulation.”⁸ These laws have taken hold throughout the world, including in Western democracies such as Canada.⁹ Human rights watchdogs have consistently criticized these laws as violations of international freedom of expression protections.¹⁰

Should governments and platforms seek to ban online falsehoods? Does exercising corporate responsibility mean social media platforms should assist or resist the implementation of laws that outlaw the dissemination of online falsehoods? Regardless of national laws on this topic, what does corporate responsibility mean in the context of global platforms’ own policies on disinformation and misinformation?

Part I of this Essay sets forth the international standards for corporate responsibility and freedom of expression. Part II examines how platforms should apply these standards amid a proliferation of laws banning falsehoods, with a particular focus on a high-profile law recently adopted in Singapore. This Part also reflects on what these standards mean for platform policies, particularly with regard to political advertisements. Part III concludes by contemplating the usefulness of the global corporate responsibility framework in helping companies react in a principled manner to “fake news” laws as well as guiding the development of their own platform policies on false information.

⁸ Funk, *supra* note 7. Nigeria is one of the countries to most recently consider adoption of a “fake news” ban. Danielle Paquette, *Nigeria’s ‘Fake News’ Bill Could Jail People for Lying on Social Media. Critics Call it Censorship.*, WASH. POST (Nov. 25, 2019, 8:46 AM), https://www.washingtonpost.com/world/africa/nigerias-fake-news-bill-could-jail-people-for-lying-on-social-media-critics-call-it-censorship/2019/11/25/ccf33c54-0f81-11ea-a533-90a7becf7713_story.html [<https://perma.cc/5CVE-UKB2>] (reporting that Nigeria’s draft bill would criminalize speech that could “threaten national security, sway elections or ‘diminish public confidence’ in the government” and Internet access could be disrupted for violators).

⁹ See Michael Karanicolas, *Canada’s Fake News Laws Face a Charter Challenge. That’s a Good Thing*, OTTAWA CITIZEN (Oct. 14, 2019), <https://ottawacitizen.com/opinion/columnists/karanicolas-canadas-fake-news-laws-face-a-charter-challenge-thats-a-good-thing> [<https://perma.cc/6BDK-T2LD>] (discussing a variety of overbroad provisions in Canada’s new law banning falsehoods during campaigns, which “gives the government a potential weapon to wield against its critics” and “can exert a chilling effect against legitimate speech, particularly when a prison term is attached to the offence.”).

¹⁰ See, e.g., *Elections in the Digital Age*, *supra* note 7; Funk, *supra* note 7; Jordan: *Fake News Amendments Need Revision*, HUMAN RIGHTS WATCH (Feb. 21, 2019, 12:00 AM), <https://www.hrw.org/news/2019/02/21/jordan-fake-news-amendments-need-revision> [<https://perma.cc/2SW5-AGMR>]; *Philippines: Reject Sweeping ‘Fake News’ Bill*, HUMAN RIGHTS WATCH (July 25, 2019, 8:00 PM), <https://www.hrw.org/news/2019/07/25/philippines-reject-sweeping-fake-news-bill> [<https://perma.cc/2D5W-VZQG>]; *Singapore: Reject Sweeping ‘Fake News’ Bill*, HUMAN RIGHTS WATCH (Apr. 3, 2019, 9:00 AM), <https://www.hrw.org/news/2019/04/03/singapore-reject-sweeping-fake-news-bill> [<https://perma.cc/6ZAE-UAN2>].

I. BACKGROUND ON THE UN'S HUMAN RIGHTS FRAMEWORK
ON CORPORATE RESPONSIBILITY

A. *The UN Guiding Principles on Business & Human Rights*

The U.S. government has repeatedly encouraged American companies to implement the corporate responsibility standards set forth in the United Nations (UN) Guiding Principles on Business & Human Rights (UNGPs) and to treat them as a floor rather than a ceiling in their operations.¹¹ The UNGPs were first adopted by the UN Human Rights Council in 2011 and reflect international expectations for how companies should act when their operations intersect with human rights issues.¹² The UNGPs state that companies have a “responsibility to respect human rights,” which means they should (1) avoid “infringing on the human rights of others” and (2) address “adverse human rights impacts with which they are involved.”¹³

Under this global corporate responsibility framework, companies are expected to proactively engage in a variety of measures, including adopting and mainstreaming human rights policies that implement the UNGPs, hiring appropriate staff to implement those policies, and engaging with external stakeholders to learn about potential human rights impacts.¹⁴ Moreover, companies are supposed to conduct due diligence to determine the human rights risks of their operations and to devise strategies to avoid undermining rights¹⁵ as well as communicate publicly on their plans to address potential harms.¹⁶ Where national law is inconsistent with international human rights law standards, companies are expected to actively seek paths to avoid infringing on human rights, but they are not supposed to violate local law.¹⁷ As is frequently noted, the UNGPs mandate that companies “know and show,”

¹¹ See, e.g., U.S. DEP'T OF STATE, RESPONSIBLE BUSINESS CONDUCT: FIRST NATIONAL ACTION PLAN FOR THE UNITED STATES OF AMERICA 17 (2016), <https://2009-2017.state.gov/documents/organization/265918.pdf> [<https://perma.cc/AWL5-8EVB>]; U.S. DEP'T OF STATE BUREAU OF DEMOCRACY, HUMAN RIGHTS, AND LABOR, U.S. GOVERNMENT APPROACH ON BUSINESS AND HUMAN RIGHTS 4 (2013), https://photos.state.gov/libraries/korea/49271/july_2013/dwoa_USG-Approach-on-Business-and-Human-Rights-updated-June2013.pdf [<https://perma.cc/9QEX-X4YT>].

¹² See Human Rights Council Res. 17/4, U.N. Doc. A/HRC/RES/17/4 (July 6, 2011); John Ruggie (Special Representative of the Secretary-General), *Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework*, U.N. Doc. A/HRC/17/31 (Mar. 21, 2011) [hereinafter UNGPs].

¹³ UNGPs, *supra* note 12, at princ. 11.

¹⁴ *Id.* at princs. 16, 18, 19.

¹⁵ *Id.* at princs. 17, 19–20.

¹⁶ *Id.* at princ. 21.

¹⁷ *Id.* at princ. 23(a), (b), cmt. Companies are expected to respect international human rights standards even if the countries they operate in have not undertaken international law obligations regarding those standards. *Id.* at princ. 11 cmt.

which means *know* their potential human rights impacts and *show* what they are doing about it.¹⁸

When social media companies tackle disinformation and misinformation—either pursuant to “fake news” laws or under their own “platform law”¹⁹—one of the most salient human rights that will be impacted is freedom of expression.²⁰ For social media companies to implement the UNGPs, they need to understand the scope of the right to freedom of expression under international human rights law²¹ in order to assess potential infringements on expression. Such infringements in the context of combatting “fake news” will often happen either by (1) helping a government to implement a law that is inconsistent with international protections for expression or (2) adopting business models or platform law that adversely affect the enjoyment of freedom of expression. The next section therefore focuses on the international standard for freedom of expression as that is key for determining whether a company is infringing on freedom of expression under either of these two scenarios.

B. The International Standard for Freedom of Expression

Article 19 of the International Covenant on Civil and Political Rights (ICCPR), which is one of the foundational treaties in the United Nations’ human rights system,

¹⁸ *Id.* at princ. 15 cmt.

¹⁹ The phrase “platform law” was used by the United Nations’ top free expression expert to refer to the rules that social media companies (which he called “enigmatic regulators”) deploy to regulate human activity on their platforms. David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶¶ 1, 63, U.N. Doc. A/HRC/38/35 (Apr. 6, 2018) [hereinafter *UN SR 2018 Report*].

²⁰ See Evelyn Mary Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26, 39–40 (2018). Additional important human rights, such as privacy, can be adversely affected in this context as well, but the focus of this Essay regards freedom of expression.

²¹ The UNGPs define “human rights” according to UN instruments rather than *regional* ones. UNGPs, *supra* note 12, at princ. 12, cmt. (citing to the UN’s international bill of human rights and related UN instruments as defining the content of international human rights standards). The UN’s *international* protections for human rights can be broader than *regional* human rights approaches. See David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶¶ 26–27 U.N. Doc. A/74/486 (Oct. 9, 2019) [hereinafter *UN SR 2019 Report*] (noting that the European Court of Human Rights’ approach to freedom of expression has often fallen short of international protections for speech and that regional norms cannot be invoked to justify departures from international standards). To implement the UNGPs, companies therefore should be focused on *international* human rights standards.

provides the international standard for protecting free expression.²² In particular, Article 19(2) establishes a broad right to seek and receive information of all kinds, regardless of frontiers and through any media.²³ Article 19(3) permits governments to limit speech when they can prove that a three-prong test is met.²⁴ To be valid, speech restrictions must be: (1) “provided by law” and (2) “necessary” to (3) achieve an enumerated public interest objective (i.e., protection of the rights or reputations of others, national security, public order, or public health or morals).²⁵

1. The “Provided by Law” Condition

The independent experts that comprise the UN human rights machinery’s oversight mechanisms have recommended interpretations for ICCPR Article 19, which help apply the tripartite test. For example, the Human Rights Committee, which is the body of independent experts charged with overseeing ICCPR implementation and elected by the treaty’s State Parties, has interpreted “provided by law” to mean, *inter alia*, that a law “must be formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly”²⁶ The UN Special Rapporteur on freedom of opinion and expression, the UN’s top expert focused on the state of free speech in all UN member states, has endorsed this

²² International Covenant on Civil and Political Rights, Dec. 16, 1966, S. Exec. Doc. E, 95-2 (1978), 999 U.N.T.S. 171 [hereinafter ICCPR].

²³ *Id.* at art. 19 ¶ 2.

²⁴ *Id.* at art. 19 ¶ 3.

²⁵ *Id.* It should be noted that UN instruments also contain mandatory bans on hateful speech, but such bans also must pass ICCPR Article 19(3)’s tripartite test to be valid. *See UN SR 2019 Report, supra* note 21, at ¶¶ 13–16. In addition, speech restrictions should be consistent with the ICCPR’s other protections, including the prohibition on discrimination. *See ICCPR, supra* note 22, at arts. 2, 26 (prohibiting discrimination when State Parties implement the treaty’s obligations and requiring State Parties to guarantee equal protection of the law without discrimination based on race, color, sex, language, religion, political or other opinion, national or social origin, property, birth, or other status). The ICCPR does, however, permit State Parties to derogate from their freedom of expression obligations during certain public emergencies if particular conditions are met. *Id.* at art. 4(1) (“In time of public emergency which threatens the life of the nation and the existence of which is officially proclaimed, the States Parties to the present Covenant may take measures derogating from their obligations under the present Covenant to the extent strictly required by the exigencies of the situation, provided that such measures are not inconsistent with their other obligations under international law and do not involve discrimination solely on the ground of race, colour, sex, language, religion or social origin.”).

²⁶ U.N. Human Rights Comm., *General Comment No. 34*, ¶ 35, U.N. Doc. CCPR/C/GC/34 (Sept. 12, 2011) [hereinafter GC 34]. The Committee has further explained that “[a] law may not confer unfettered discretion for the restriction of freedom of expression on those charged with its execution.” *Id.*

interpretation²⁷ and provided significant guidance on terminology that is inappropriately vague.²⁸

The Special Rapporteur has further elaborated on the issue of vagueness in the context of “fake news” laws. For example, in a joint statement with regional free expression experts, the Special Rapporteur has opined that “[g]eneral prohibitions on the dissemination of information based on vague and ambiguous ideas, including ‘false news’ or ‘non-objective information,’ are incompatible with international standards . . . and should be abolished.”²⁹ The Special Rapporteur has also reacted to specific “fake news” laws adopted or considered in various countries, which has resulted in further clarification of phraseology that is improperly vague. For instance, the Special Rapporteur criticized an Italian law for banning “unfounded,” “biased,” “false,” and “fake” information, noting such words do not give sufficient notice of what is off-limits to individuals or government officials implementing the law, which can empower censorship and chill speech.³⁰

²⁷ See, e.g., David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 12, U.N. Doc. A/HRC/71/373 (Sept. 6, 2016) [hereinafter *UN SR 2016 Report*] (restrictions must “be formulated with sufficient precision to enable both the individual and those charged with its execution to regulate conduct accordingly and be made accessible to the public. It cannot confer discretion for the restriction of freedom of expression on those charged with its execution.”); *UN SR 2018 Report*, *supra* note 19, at ¶ 7 (enunciating that restrictions on speech must “limit government discretion in a manner that distinguishes between lawful and unlawful expression with ‘sufficient precision’”). The Special Rapporteur has also made clear that “[r]ules should be subject to public comment and regular legislative or administrative processes. Procedural safeguards, especially those guaranteed by independent courts or tribunals, should protect rights.” *UN SR 2019 Report*, *supra* note 21, at ¶ 6.

²⁸ For example, the Special Rapporteur has criticized as unduly vague draft Chinese legislation that would ban engaging “in activities harming national security” or that “upset social order”; Kenyan legislation that criminalized “obscene, gory or offensive material which is likely to cause fear and alarm to the general public”; and European counter-terrorism approaches that ban the “glorification of terrorism.” *UN SR 2016 Report*, *supra* note 27, at ¶¶ 13–14. The Special Rapporteur has also highlighted inappropriately vague provisions in the platform law of social media companies. See *UN SR 2018 Report*, *supra* note 19, at ¶ 26 (“Company policies on hate, harassment and abuse also do not clearly indicate what constitutes an offence. Twitter’s prohibition of ‘behavior that incites fear about a protected group’ and Facebook’s distinction between ‘direct attacks’ on protected characteristics and merely ‘distasteful or offensive content’ are subjective and unstable bases for content moderation.”).

²⁹ ORGANIZATION FOR SECURITY AND CO-OPERATION IN EUROPE, JOINT DECLARATION ON FREEDOM OF EXPRESSION AND “FAKE NEWS” ¶ 2(a), Mar. 3, 2017, FOM.GAL/3/17, <https://www.osce.org/fom/302796> [<https://perma.cc/9NL2-M2VU>] [hereinafter *Fake News Declaration*]. This Joint Declaration also specifies that the “human right to impart information and ideas is not limited to ‘correct’ statements. . . .” *Id.* at preambular para. 7.

³⁰ *Elections in the Digital Age*, *supra* note 7, at 9.

2. *The “Necessary” Condition*

The second prong of ICCPR Article 19(3)’s tripartite test requires that any restriction be “necessary” to achieve an enumerated public interest objective.³¹ The word choice of “necessary” is significant. To be acceptable, a speech restriction may not be merely related to or useful in achieving a public interest goal. As the UN’s human rights machinery has stated, to prove a restriction is “necessary,” governments must show, *inter alia*, it is the “least intrusive means” of achieving the public interest goal.³²

To prove a speech restriction poses the least possible burden on speech, a three-part inquiry should be undertaken.³³ First, is it possible to achieve the public interest goal without restricting speech? This question forces governments to engage in good governance measures before resorting to speech bans. In the context of combatting “fake news,” it is important to assess if governments, for example, have promoted digital and media literacy campaigns to make their populations as resistant as possible to disinformation and misinformation. It is also important to assess whether government actors themselves have been spreading such false information.³⁴ A governmental failure to engage in good governance does not empower governments to censor speech. Rather, it highlights that such governments have not sought the least intrusive way of tackling disinformation or misinformation.

Second, if a government can show that good governance measures are insufficient to achieve the goal, has it selected, from the range of potential options, the restriction on speech that least burdens expression? An issue that often comes up in the context of bans on false information is whether the government imposes civil or criminal sanctions and their level of severity. The UN’s human rights machinery has repeatedly warned against the use of criminal sanctions for defamation and other bans on false information.³⁵ The UN Special Rapporteur has concluded that this

³¹ ICCPR, *supra* note 22, at art. 19 ¶ 3.

³² GC 34, *supra* note 26, at ¶ 34 (a restriction “must be the least intrusive instrument amongst those which might achieve [the] protective function”); *UN SR 2018 Report*, *supra* note 19, at ¶ 7 (“States must demonstrate that the restriction imposes the least burden on the exercise of the right and actually protects, or is likely to protect, the legitimate State interest at issue. States may not merely assert necessity but must demonstrate it. . .”).

³³ I proposed this three-part test in 2018, and it was cited to favorably by the UN Special Rapporteur. See *UN SR 2019 Report*, *supra* note 21, at ¶ 52 (citing Evelyn M. Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26 (2018)).

³⁴ As noted by the UN Special Rapporteur and regional freedom of expression experts, “[s]tate actors should . . . take care to ensure that they disseminate reliable and trustworthy information, including about matters of public interest, such as the economy, public health, security and the environment.” Fake News Declaration, *supra* note 29, at ¶ 2(d).

³⁵ *Id.* at ¶ 2(b) (“Criminal defamation laws are unduly restrictive and should be abolished. Civil law rules on liability for false and defamatory statements are legitimate only if defendants are given a full opportunity and fail to prove the truth of those statements and also benefit from other defences, such as fair comment.”); GC 34, *supra* note 26, at ¶ 47

“presumption against the criminalization of expression also applies to restrictions on ‘fake news’ and disinformation.”³⁶

Lastly, governments must ask whether the speech restrictions they choose actually help—or are at least likely to help—achieve the public interest objective. If speech restrictions are not effective in achieving the goal or are counter-productive, they cannot be defended as “necessary” to achieve the objective. For example, attempts to ban speech can often raise the profile of certain speech, which defeats the purpose of the ban.³⁷ In this regard, it should be noted that at all stages in this three-part inquiry, it is important that policy choices be grounded in evidence-based decision-making.³⁸

3. The “Legitimacy” Condition

The third prong of ICCPR Article 19(3)’s tripartite test requires that the reason for the speech restriction be one of the enumerated public interest objectives listed in the treaty: (1) “for respect of the rights or reputations of others,” or (2) “the protection of national security or of public order . . . or of public health or morals.”³⁹ The UN’s Human Rights Committee has made clear that governments cannot invoke these objectives as pretexts to, for instance, protect the ruling party.⁴⁰ In the context of laws banning “fake news,” potentially legitimate governmental interests could include the rights of others (e.g., the rights to receive information and take part in elections)⁴¹ as well as protection of public order (e.g., if disinformation threatens to incite imminent violence). That said, the UN Special Rapporteur has found that governments often invoke the problems of disinformation and misinformation as pretexts to improperly control speech,⁴² which is inconsistent with the third prong of ICCPR Article 19(3)’s tripartite test.

(“States parties should consider the decriminalization of defamation and, in any case, the application of the criminal law should only be countenanced in the most serious of cases and imprisonment is never an appropriate penalty.”).

³⁶ *Elections in the Digital Age*, *supra* note 7, at 10.

³⁷ See Justin Parkinson, *The Perils of the Streisand Effect*, BBC NEWS MAG. (July 31, 2014), <https://www.bbc.com/news/magazine-28562156> [<https://perma.cc/J5X8-G966>] (highlighting attempts by celebrities to remove speech about themselves and demonstrating how those attempts resulted in significant additional exposure for the information they were seeking to suppress).

³⁸ See *Elections in the Digital Age*, *supra* note 7, at 11.

³⁹ ICCPR, *supra* note 22, at art. 19 ¶ 3.

⁴⁰ GC 34, *supra* note 26, at ¶ 30 (highlighting that it is inappropriate for governments to limit speech on national security grounds when such claims are merely pretexts).

⁴¹ *Elections in the Digital Age*, *supra* note 7, at 9 (“[D]isinformation and propaganda may mislead populations and interfere with the public’s right to know, particularly during elections.”).

⁴² *Id.* (“[G]overnments are also capitalizing on the phenomenon of disinformation to propose and enact laws and regulations that interfere with the freedom of expression by restricting legitimate speech, especially during elections.”).

II. UN CORPORATE RESPONSIBILITY STANDARDS & “FAKE NEWS” ONLINE

With the rise of “fake news” laws that prohibit falsehoods, social media companies will need to answer two key questions to ensure they act consistently with the UNGPs in each country that has adopted such laws. First, is a particular “fake news” law consistent with international human rights law standards? Second, if such a law does not meet international human rights standards, what should the company do to fulfill its responsibility to respect these standards? This Part begins by unpacking each of these two questions and then turns to how UN corporate responsibility standards can provide guidance with respect to a social media company’s own platform law.

A. Singapore’s Recent Online Falsehoods Ban

In October 2019, one of the world’s most recent laws banning “fake news” went into effect in Singapore: The Protection from Online Falsehoods & Manipulation Act (POFMA).⁴³ Singapore’s government justified the law’s adoption by referring, among other things, to the need to protect its society from “hostile parties” who could “turn different groups against one another and cause disorder.”⁴⁴ One governmental official in particular represented to the Parliament that this law would be necessary to deal with the type of false information that had been circulated online in 2016 prior to the U.S. elections and the UK’s Brexit vote.⁴⁵

An overview of POFMA’s key provisions reveals its potential breadth and impacts. The Act makes it illegal for a person to share—knowingly or having reason to believe—a “false statement of fact” when that information is, *inter alia*, likely “prejudicial” to Singapore’s security, public health, safety or tranquility, “friendly relations . . . with other countries,” or likely to incite feelings of “ill-will.”⁴⁶ A statement is false “if it is false or misleading, whether wholly or in part, and whether

⁴³ James Griffiths, *Singapore ‘Fake News’ Law Comes into Force, Offenders Face Fines and Prison Time*, CNN (Oct. 2, 2019, 1:05 AM), <https://www.cnn.com/2019/10/02/asia/singapore-fake-news-internet-censorship-intl-hnk/index.html> [https://perma.cc/WUH6-KR4N].

⁴⁴ *Id.*

⁴⁵ Adam Taylor, *First Target of Singapore’s ‘Fake News’ Law Is Facebook Post that Alleged Failed State Investment in Restaurant*, WASH. POST (Nov. 27, 2019, 11:04 AM), <https://www.washingtonpost.com/world/2019/11/25/first-target-singapores-fake-news-law-is-facebook-post-that-alleged-failed-state-investment-salt-bae/> [https://perma.cc/UPC4-6QGS].

⁴⁶ Protection from Online Falsehoods and Manipulation Act 2019 (No. 18 of 2019), June 3, 2019, §§ 7(1)(a)–(b)(i), (ii), (iii), (v), <https://sso.agc.gov.sg/Act/POFMA2019?ValidDate=20191002&TransactionDate=20191007> [https://perma.cc/2DUW-A53K] [hereinafter POFMA]. This prohibition applies to a person in or outside of Singapore. *Id.*

on its own or in the context in which it appears.”⁴⁷ Violators face criminal penalties of up to S\$50,000 (about US\$37,000) and/or five years in prison.⁴⁸

Without the approval of a judge, any Singapore government minister can order a person to remove a “false” post or to add a correction to such post if that serves the public interest.⁴⁹ Individuals face criminal fines and/or twelve months in prison for not implementing these orders.⁵⁰ If a person does not comply with such an order, then the government can order an internet access service to block access to the offending statement or order an internet intermediary to correct the offending material.⁵¹ Non-compliant companies can face hefty criminal fines.⁵² Individuals and companies can appeal such orders, but the grounds for appeal are limited and the order remains in effect during the appeal.⁵³

Singapore’s first POFMA enforcement action was against one of the regime’s political opponents (Brad Bowyer).⁵⁴ Mr. Bowyer had posted statements critical of the investments of two state-owned companies, which he implied the government had influenced in making those financial decisions.⁵⁵ Invoking POFMA authority, the Minister of Finance contacted Mr. Bowyer and instructed him to correct his Facebook post.⁵⁶ Mr. Bowyer did so by stating that his post contained falsehoods and telling readers to visit the government’s website for the correct facts.⁵⁷ Although Singapore did not require Mr. Bowyer to remove his post or impose a fine or prison time,⁵⁸ it was evident that the law was enforced for the purpose of silencing a regime

⁴⁷ *Id.* at § 2(2)(b). Also, a statement of “fact” is a “statement which a reasonable person . . . would consider to be a representation of fact.” *Id.* at § 2(2)(a).

⁴⁸ *Id.* at § 7(2). Additional penalties apply if inauthentic accounts or bots were used to share falsehoods. *Id.* at § 7(3).

⁴⁹ *Id.* at §§ 10–13. Correction or removal orders can be imposed even if the individual did not know or have reason to believe a statement was false. *Id.* at §§ 11(4), 12(4).

⁵⁰ *Id.* at § 15(1).

⁵¹ *Id.* at §§ 16, 21. These measures appear geared toward preventing end-users in Singapore from viewing the “fake news” rather than those outside the country.

⁵² *Id.* at §§ 16, 27, 28.

⁵³ *Id.* at § 29. The grounds for appeal of correction or disabling orders generally involve arguing (1) the offensive material was not communicated in Singapore; (2) that the statement was not factual in nature or was true; and (3) that it is technically impossible to comply with the order. *Id.* at § 29(5). Before appealing, a request for reconsideration must be made to the relevant minister who issued the order. *Id.* at § 29(2).

⁵⁴ Taylor, *supra* note 45.

⁵⁵ *Id.*

⁵⁶ *Id.*

⁵⁷ *Id.*

⁵⁸ Tim Cushing, *Singapore Government Tests Out Its Fake News Law Against an Opposition Party Leader*, TECH CRUNCH (Nov. 27, 2019, 10:42 AM), <https://www.techdirt.com/articles/20191125/17321343457/singapore-government-tests-out-fake-news-law-against-opposition-party-leader.shtml> [<https://perma.cc/MX64-HAJY>] (concluding that the government’s “initial salvo may have been delivered with a light touch,

critic, rather than for the reasons originally cited in lobbying for the law's adoption, i.e., protecting society from online foreign interference and promoting sectarian harmony.⁵⁹

A few days later, Singapore issued its first POFMA order to a platform.⁶⁰ The government initially ordered the editor of an Australian website to issue a correction about a Facebook post that alleged an unlawful arrest had occurred in Singapore, but the editor refused to implement the order, as he resides outside of Singapore.⁶¹ The government then issued an order to Facebook to post a correction on the offending post.⁶² Facebook complied with the order by adding a label to the original post that said "Facebook is legally required to tell you that the Singapore government says this post has false information,"⁶³ which could only be seen in Singapore.⁶⁴ Facebook also included a link to a page that states the company "doesn't endorse the truthfulness of either the posts on its site or government corrections."⁶⁵ Again, it was evident that Singapore's "fake news" law was enforced when the government was criticized, which serves as a lesson about why giving already powerful authorities even more power to become arbiters of truth can be so dangerous.⁶⁶

but when things heat up around elections and the discussion of controversial legislation, expect the government's blows to land with a little more force.").

⁵⁹ *Id.*; Taylor, *supra* note 45.

⁶⁰ Niharika Mandhana & Phred Dvorak, *Ordered by Singapore, Facebook Issues a Correction*, WALL STREET J.: MARKETWATCH (Nov. 30, 2019, 7:15 AM), <https://www.marketwatch.com/story/ordered-by-singapore-facebook-posts-a-correction-2019-11-30> [<https://perma.cc/F77L-9Y4Y>].

⁶¹ *Id.* The editor says the Singapore government never contacted him. Cameron Wilson, *This Man's Post Was the First to Be "Corrected" by Facebook under Singapore's Fake News Law*, BUZZFEED NEWS (Dec. 2, 2019, 9:31 PM), <https://www.buzzfeed.com/cameron-wilson/singapore-fake-news-law-facebook-correction> [<https://perma.cc/RUK5-KQXQ>]. The editor explained he was contacted by a friend who saw the correction order and then the editor posted that he would not comply with foreign governmental orders. *Id.*

⁶² Mandhana & Dvorak, *supra* note 60.

⁶³ *Id.*

⁶⁴ Mike Butcher, *Facebook Bowed to a Singapore Government Order to Brand a News Post as False*, TECHCRUNCH (Nov. 30, 2019, 1:05 AM), <https://techcrunch.com/2019/11/30/facebook-bowed-to-a-singapore-government-order-to-brand-a-news-post-as-false/> [<https://perma.cc/8FML-Z6LU>].

⁶⁵ Mandhana & Dvorak, *supra* note 60.

⁶⁶ See Casey Newton, *Singapore's Fake News Law Should Be a Warning to American Lawmakers*, THE VERGE (Dec. 3, 2019, 6:00 AM), <https://www.theverge.com/interface/2019/12/3/20991422/singapore-fake-news-law-censorship-politics-usa> [<https://perma.cc/N652-S2JE>] ("But the lesson of Singapore is that the fake-news law you want probably won't be used in the way that you want. In fact, it may be used in ways that you don't want at all!"). Singapore's next two POFMA orders were also against "an opposition party or politician, or a government critic." Kirsten Han, *Want to Criticize Singapore? Expect a 'Correction Notice,'* N.Y. TIMES (Jan. 21, 2020), <https://www.nytimes.com/2020/01/21/opinion/fake-news-law-singapore.html> [<https://perma.cc/BWD2-M6KB>].

B. Application of UN Standards

1. The POFMA and ICCPR Article 19

Although Singapore is not a party to the ICCPR,⁶⁷ the UNGPs expect companies to assess whether national laws are consistent with international standards regardless of whether a government has taken on particular human rights obligations.⁶⁸ The POFMA fails to meet the ICCPR Article 19(3) standards for speech restrictions for a variety of reasons.⁶⁹ For example, its language banning “false statements of fact” and the definition of such falsehoods⁷⁰ are classic examples of inappropriately vague terminology that the UN Special Rapporteur has condemned as giving too much discretion to governing officials and a lack of notice to the public.⁷¹

In terms of the “necessity” condition, the law fails the “least intrusive means” test as well. Singapore has not met its burden of showing why good governance and related measures, including independent fact-checking, media and digital literacy efforts, and the issuance of governmental notices on its own official website, are insufficient options in tackling false information. In addition, the law contemplates criminal punishments for noncompliance, which the UN’s machinery has noted will rarely constitute the least intrusive means to achieve public interest goals in the context of combatting false information.⁷² The third part of Article 19(3)’s tripartite

⁶⁷ *International Covenant on Civil and Political Rights*, U.N. TREATY COLLECTION, https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-4&chapter=4&clang=_en [<https://perma.cc/3SXD-K5PQ>] (last visited Feb. 7, 2020). It is also important to recall that the UNGPs mandate respecting *international* human rights standards rather than *regional* standards. See *supra* note 21. Singapore is a member of the Association of Southeast Asian Nations, which has a regional human rights declaration. Ass’n of Southeast Asian States [ASEAN] Human Rights Declaration (Nov. 19, 2012), <http://asean.org/asean-human-rights-declaration/> [<https://perma.cc/NED2-4MT4>]. However, this declaration provides fewer protections than UN standards. See Press Release, U.S. Dep’t of State, ASEAN Declaration on Human Rights Press Statement (Nov. 20, 2012), <https://2009-2017.state.gov/r/pa/prs/ps/2012/11/200915.htm> [<https://perma.cc/FJ8C-RPYS>] (criticizing the ASEAN Declaration for limiting human rights in a variety of ways that are inconsistent with UN standards).

⁶⁸ UNGPs, *supra* note 12, at princ. 11 cmt.

⁶⁹ Though the focus of this Essay is on freedom of expression, it should be noted as well that the POFMA also poses a variety of privacy concerns. See Jennifer Daskal, *This ‘Fake News’ Law Threatens Free Speech. But it Does Not Stop There.*, N.Y. TIMES (May 30, 2019), <https://www.nytimes.com/2019/05/30/opinion/hate-speech-law-singapore.html> [<https://perma.cc/AJ4N-4ETF>] (explaining how the POFMA allows the government to “mandate that service providers track the viewing habits of their users in ways that dangerously threaten their privacy.”).

⁷⁰ See *supra* notes 46–47 and accompanying text.

⁷¹ See *supra* notes 27–30 and accompanying text.

⁷² See *supra* notes 35–36 and accompanying text.

test (the legitimacy condition) also does not appear to be met as the initial POFMA enforcement measures were targeted at regime critics rather than public interest goals, which indicates the publicly stated goals were mere pretexts for protecting the government from criticism.⁷³

2. *The POFMA and the UNGPs*

Given that POFMA is not consistent with the international standard for freedom of expression, what do UN corporate responsibility standards expect of companies that face enforcement orders issued under this law? To begin with, platforms should engage with governments when such bills are being drafted and use their leverage to urge them not to adopt laws that are inconsistent with international human rights standards.⁷⁴ Once a law such as POFMA is adopted, a platform should have in place procedures to deal with enforcement orders at the appropriate level in the company and assure steps are taken to avoid infringing human rights to the extent possible.⁷⁵

When a platform receives a POFMA enforcement order, it should consider whether to appeal, even though an appeal would not stay the enforcement order during the appellate process.⁷⁶ At the very least, a social media company should consider seeking a request for reconsideration from the minister who issued the order.⁷⁷ If the minister denies the reconsideration request, the company should assess whether to appeal the order for review by a court.⁷⁸ While POFMA limits the standards for a successful appeal, lodging an appeal can send a distinct signal to the government about a corporation's profound concerns about the law and shed an additional spotlight on the government's implementation of the law, both of which could impact the government's future implementation choices.

To minimize negative human rights impacts, companies should narrowly interpret POFMA enforcement orders. For example, when a company places a correction notice on a post, it should not give oxygen to the government's

⁷³ See *supra* notes 39–42 and accompanying text.

⁷⁴ In the case of POFMA, it appears that Google and Facebook voiced opposition to the draft bill. James Griffiths, *Singapore Just Used Its Fake News Law. Critics Say It's Just What They Feared*, CNN BUS. (Nov. 30, 2019, 9:47 AM), <https://www.cnn.com/2019/11/29/media/singapore-fake-news-facebook-intl-hnk/index.html> [https://perma.cc/7XJ5-DMFM]. However, it is unclear how strenuously such objections were made. In addition, Google, Facebook, and Twitter noted concerns with the law after it was passed. Michelle Toh, *Google Says Singapore Risks Hurting Innovation with Fake News Law*, CNN BUS. (May 9, 2019, 10:46 AM), <https://www.cnn.com/2019/05/09/tech/singapore-fake-news-law-tech/index.html> [https://perma.cc/6HN5-6V7B].

⁷⁵ See UNGPs, *supra* note 12, at princ. 19.

⁷⁶ See *supra* note 53 and accompanying text.

⁷⁷ *Id.* From the reporting about the first POFMA order issued to a platform, it does not appear that Facebook requested the government minister to reconsider the order. See Griffiths, *supra* note 74; see also Toh, *supra* note 74.

⁷⁸ See *supra* note 53 and accompanying text.

assessment of the truth and do the minimum to comply with the law.⁷⁹ Companies should also give notice to affected users about the label that will be placed on their posts.⁸⁰ Finally, a platform should address any adverse impacts to human rights from its selected course of action.⁸¹ In sum, although implementing the UNGPs does not exempt social media companies from implementing national laws such as POFMA, it does require them to “know and show,” which helps shape corporate reactions and minimize human rights infringements to the extent possible.

C. Falsehoods, Platform Law, Business Models, & Corporate Responsibility

Although the focus of this Essay regards the recent uptick in governmental “fake news” laws and how responsible corporations should react to them, it is also worth reflecting briefly on a related issue: social media companies’ own policies and business models and how they impact human rights when tackling disinformation. The UNGPs apply not only to corporate reactions to state regulation but also to a company’s own business decisions that impact human rights.⁸² According to the UNGPs, therefore, a social media company must also assess if, in tackling disinformation, its own business model and platform law infringe on human rights, including freedom of expression.

One particularly hot issue involving platform law, business models, and the spread of online falsehoods regards how social media companies treat political ads, including those placed by politicians as well as issue-specific ads. Recently, major U.S. social media companies have issued differing policy approaches for political ads.⁸³ In September 2019, Facebook confirmed it would continue to implement its existing policy that exempts politicians’ ads (and their organic content) from its third-party fact-checking program.⁸⁴ There was a significant backlash that this policy

⁷⁹ In response to Singapore’s correction notice order, Facebook’s addition to the user’s post avoided giving credence to the government’s assertion of truth and even provided an additional link with more information about POFMA as well as the company’s stance in not taking a position on the truthfulness of either the users’ posts or the government’s assessments of user speech. *See supra* notes 63–65 and accompanying text.

⁸⁰ In the case of Facebook’s implementation of the POFMA correction order, the affected user stated that the company did not give him notice that a label would be put on his post. Wilson, *supra* note 61.

⁸¹ *See* UNGPs, *supra* note 12, at princ. 22.

⁸² *Id.* at princ. 11, cmt. (stating that the business responsibility to respect human rights “exists over and above compliance with national laws and regulations protecting human rights”).

⁸³ Gennie Gebhart, *Three Political Ad Policies and No Good Answers*, ELECTRONIC FRONTIER FOUND. (Nov. 2, 2019), <https://www.eff.org/deeplinks/2019/11/three-political-ad-policies-and-no-good-answers> [<https://perma.cc/C575-ZXMG>] (analyzing the differences among the political advertisement policies of Facebook, Twitter, and Google).

⁸⁴ Nick Clegg, *Facebook, Elections and Political Speech*, FACEBOOK NEWSROOM (Sept. 24, 2019), <https://about.fb.com/news/2019/09/elections-and-political-speech/> [<https://perma.cc/3UWG-4WV8>].

would enable politicians to promote online falsehoods.⁸⁵ Facebook's chief defense of its policy was that people should be able to see and scrutinize politicians' ads.⁸⁶ However, Facebook's employees stated that such public scrutiny is undermined because the company provides politicians and campaigns with microtargeting for their ads, which means users see different ads based on information derived from, *inter alia*, behavioral tracking tools.⁸⁷

Amid reactions to Facebook's approach, Twitter's CEO tweeted that his company would soon ban political advertising, including candidate and issue ads.⁸⁸ Another Twitter official clarified that the company's existing policies defined issue ads as those covering topics of national legislative importance, such as climate change, immigration, and healthcare.⁸⁹ Twitter's CEO stated this ban was not about free expression because it involved paid advertising rather than organic content.⁹⁰ Shortly thereafter, the UN Special Rapporteur replied to his tweet, noting political advertisement policies do indeed impact freedom of expression.⁹¹ Commentators spoke out about the potential adverse free expression impacts of the impending ban.⁹² Ultimately, Twitter decided to ban candidate and campaign ads but generally

⁸⁵ See, e.g., Jeanine Santucci, *Elizabeth Warren Targets Facebook Fact-Checking Policy with False Ad Saying Zuckerberg Endorsed Trump*, USA TODAY (Oct. 12, 2019, 5:17 PM), <https://www.usatoday.com/story/news/politics/elections/2019/10/12/elizabeth-warren-targets-trump-zuckerberg-false-facebook-ad/3960138002/> [<https://perma.cc/AX5S-Q47R>] (explaining how a candidate for U.S. President purchased a Facebook ad containing a demonstrably false statement in order to highlight concerns that Facebook's political ads policy facilitates the circulation of disinformation around elections).

⁸⁶ Clegg, *supra* note 84.

⁸⁷ *Read the Letter Facebook Employees Sent to Mark Zuckerberg About Political Ads*, N.Y. TIMES (Oct. 28, 2019), <https://www.nytimes.com/2019/10/28/technology/facebook-mark-zuckerberg-letter.html?action=click&module=RelatedLinks&pgtype=Article> [<https://perma.cc/GCS8-4RE9>] (calling for an end to microtargeting political ads because it undermines the public scrutiny needed to assess political speech). In early 2020, Facebook announced some updates to its political ads policy. Rob Leathern, *Expanded Transparency and More Controls for Political Ads*, FACEBOOK NEWSROOM (Jan. 9, 2020), <https://about.fb.com/news/2020/01/political-ads/> [<https://perma.cc/Q9Y8-NYXJ>]. Essentially, the company plans to make its public archives of political ads more transparent and accessible. It will also allow users greater control over the ads they see. *Id.*

⁸⁸ Jack Dorsey (@jack), TWITTER (Oct. 30, 2019, 2:05 PM), <https://twitter.com/jack/status/1189634360472829952?s=20> [<https://perma.cc/3DJE-AK4G>].

⁸⁹ Vijaya Gadde (@vijaya), TWITTER (Oct. 30, 2019, 4:04 PM), <https://twitter.com/vijaya/status/1189664481263046656?s=20> [<https://perma.cc/6T5P-4LGE>].

⁹⁰ Jack Dorsey (@jack), TWITTER (Oct. 30, 2019, 2:05 PM), <https://twitter.com/jack/status/1189634377057067008?s=20> [<https://perma.cc/V4Z6-ZNGD>].

⁹¹ David Kaye (@davidakaye), TWITTER (Oct. 31, 2019, 1:02 AM), <https://twitter.com/davidakaye/status/1189769465816158210?s=20> [<https://perma.cc/T7JZ-H6PN>].

⁹² See, e.g., Will Oremus, *Twitter's Ban on Political Ads Will Hurt Activists, Labor Groups, and Organizers*, MEDIUM (Oct. 31, 2019), <https://onezero.medium.com/twitters-ban-on-political-ads-will-hurt-activists-labor-groups-and-organizers-c339908b841d> [<https://perma.cc/841D-3399>].

allows issue ads, although limiting aspects of microtargeting for such ads.⁹³ At around the same time, Google issued a policy update that did not ban election ads, but it did limit certain microtargeting tools for such ads.⁹⁴

Under the UNGPs, the three companies should have engaged in a global human rights due diligence analysis that involved input from internal and external stakeholders to assess how potential policies for political advertisements (combined with their underlying business models that encompass microtargeting) could impact human rights globally, including freedom of expression.⁹⁵ Given the core significance of political advertising and disinformation to election integrity and human rights, the companies should also have been as transparent as possible with the results of any human rights impact assessments that were conducted about their political ads policies.⁹⁶ Unfortunately, it seems that such human rights due diligence may not have occurred prior to the adoption of these policies. Would Twitter’s CEO have tweeted that banning political ads does not impact freedom of expression if a global human rights impact analysis had been conducted? Moreover, none of the companies released human rights impact assessments before adopting their political ads policies.

To meet the UNGPs corporate responsibility standards, companies need to assess whether their own business models and platform law infringe human rights, including freedom of expression.⁹⁷ In the context of political ads policies, it is worth briefly examining Twitter’s decision to ban candidate ads.⁹⁸ Specifically, was an

[//perma.cc/94J7-3UXH](https://perma.cc/94J7-3UXH)] (noting downsides of Twitter’s impending ban, including favoring commercial speech by extractives companies and preventing issue ads about climate change); Isaac Stanley-Becker, *Ban Political Ads on Facebook? Upstart Anti-Trump Candidates Object*, WASH. POST (Nov. 10, 2019, 10:00 AM), https://www.washingtonpost.com/politics/ban-political-ads-on-facebook-upstart-anti-trump-candidates-object/2019/11/09/12be4f1a-fffd-11e9-8501-2a7123a38c58_story.html [<https://perma.cc/984S-JMMF>] (highlighting that a ban on political ads on social media would favor incumbents).

⁹³ Gebhart, *supra* note 83.

⁹⁴ *Id.*

⁹⁵ See UNGPs, *supra* note 12, at princs. 17–20 (describing how and when to conduct human rights due diligence).

⁹⁶ See *id.* at princ. 21 (discussing the need to communicate with those outside the company about human rights impacts).

⁹⁷ It should be noted that Amnesty International recently issued a report about how Google and Facebook’s business model of extracting and monetizing the personal data of users has adverse impacts on a variety of human rights, including privacy and expression. AMNESTY INT’L, SURVEILLANCE GIANTS: HOW THE BUSINESS MODEL OF GOOGLE AND FACEBOOK THREATENS HUMAN RIGHTS 7, 21 (2019), <https://www.amnesty.org/en/documents/pol30/1404/2019/en/> [<https://perma.cc/8JW4-NCBA>].

⁹⁸ See generally Lauren Feiner, *Twitter Bans Political Ads After Facebook Refused to Do So*, CNBC (Oct. 30, 2019), <https://www.cnn.com/2019/10/30/twitter-bans-political-ads-after-facebook-refused-to-do-so.html> [<https://perma.cc/LM2S-M355>] (noting that Twitter decided to ban ads from political campaigns). There are additional human rights

outright ban on such ads the least intrusive means of tackling disinformation, or were there other measures Twitter could have taken that would have burdened political speech less while achieving its goal?⁹⁹

A former Chair of the U.S. Federal Election Commission has argued eloquently against bans by social media companies on political ads.¹⁰⁰ Instead, she favors an end to the microtargeting that accompanies such ads.¹⁰¹ According to former Chair Weintraub, microtargeting amplifies disinformation harms by targeting susceptible groups with political ads that the rest of society does not see.¹⁰² Ending such microtargeting would, she argues, deter disinformation because “malicious advertisers, foreign and domestic, would be less likely to say to an entire state what they have been willing to say to a small audience targeted for its susceptibility.”¹⁰³ She assesses that ending microtargeting would also alleviate the problem of divisive political ads because advertisers “would have to appeal to a wider audience.”¹⁰⁴ Professor Siva Vaidhyanathan has also advocated for an end to microtargeting of political ads in order to bring wider public scrutiny to advertising as a means of countering potential disinformation harms.¹⁰⁵

Before resorting to speech bans (such as prohibiting candidate ads) as a means of tackling disinformation, Twitter should have engaged in an analysis of whether it has other options at its disposal to achieve legitimate public interest goals involving election integrity.¹⁰⁶ Serious arguments have been made that indicate limiting

issues with regard to the various political ads policies discussed in this section, but they are beyond the scope of this Essay.

⁹⁹ To respect international freedom of expression standards, Twitter needs to show, among other things, that its restrictions on speech are the least intrusive means of achieving a public interest goal. *See supra* notes 31–36 and accompanying text.

¹⁰⁰ Ellen L. Weintraub, *Don't Abolish Political Ads on Social Media. Stop Microtargeting.*, WASH. POST (Nov. 1, 2019, 6:51 PM), <https://www.washingtonpost.com/opinions/2019/11/01/dont-abolish-political-ads-social-media-stop-microtargeting/> [<https://perma.cc/XT3D-TY4S>].

¹⁰¹ *Id.*

¹⁰² *Id.*

¹⁰³ *Id.*

¹⁰⁴ *Id.* Weintraub proposes that targeting political ads only be allowed when it is generally no more specific “than one political level below the election at which the ad is directed.” *Id.* In other words, an ad for governor could run across a particular state or in certain districts, but not target individuals more precisely. *Id.* For presidential elections, she would be comfortable with ads “targeted down two levels, to the state and then to the county or congressional district level.” *Id.*

¹⁰⁵ Siva Vaidhyanathan, *The Real Reason Facebook Won't Fact Check Political Ads*, N.Y. TIMES (Nov. 2, 2019), <https://www.nytimes.com/2019/11/02/opinion/facebook-zuckerberg-political-ads.html> [<https://perma.cc/6CAC-84V2>]. He also argues that platforms cannot be trusted with global fact-checking of all candidate and campaign claims throughout the world. *Id.*

¹⁰⁶ The Electronic Frontier Foundation (EFF) has also challenged platforms to “examine the parts of their infrastructure that are acting as a megaphone for dangerous

microtargeting may sufficiently address potential disinformation harms and could obviate the need to ban political ads.¹⁰⁷ Under the UNGPs, Twitter should grapple with these questions and provide the public with an assessment of potential negative human rights impacts from its ban on candidate ads as well as whether ending microtargeting for such ads would be sufficient to counter disinformation harms.

III. CONCLUDING OBSERVATIONS

The corporate responsibility of social media companies to respect human rights when combatting disinformation and misinformation is triggered not only when companies help governments to implement laws that are inconsistent with international human rights protections, but also when they adopt platform laws or business models that adversely affect the enjoyment of human rights. While there can be no denying that the viral spread of falsehoods over global social media platforms is problematic, societies must remain vigilant in guarding against solutions that undermine human rights, including freedom of expression.

The rallying cry for governments to “do something” about false information is being used to justify a proliferation of “fake news” laws that improperly empower governments to control discourse, including by enlisting social media companies to help implement these speech codes. As exemplified in Singapore’s POFMA, such laws can be inconsistent with international freedom of expression standards because they are vague, do not reflect the least intrusive means for tackling falsehoods, and are applied against regime critics rather than for public interest goals.¹⁰⁸ The breadth and implementation of such laws should give us pause in seeking to empower those already in power—whether in government or corporations—to censor falsehoods.

Even in the absence of state regulation on falsehoods, platforms can undermine their users’ enjoyment of freedom of expression and other rights when their business models and policies are not developed with human rights in mind. With rising concerns about the relationship between social media business models, platform law, and disinformation harms, it is important that companies assess their policies, including their approaches to microtargeting, in light of the UNGPs.

Indeed, as complex and challenging as the issue of combatting disinformation online throughout the world can be, it is important to recall that there is a global, principled framework for corporate responsibility that provides helpful guidance for how to proceed. The UNGPs provide companies with substantive standards they should use to judge human rights impacts of measures to combat disinformation. For example, while national freedom of expression laws vary significantly throughout

content and address that root cause of the problem rather than censoring users.” See Jillian C. York, David Greene & Gennie Gebhart, *Censorship Can’t Be the Only Answer to Disinformation Online*, ELECTRONIC FRONTIER FOUND. (May 1, 2019), <https://www EFF.ORG/deeplinks/2019/05/censorship-cant-be-only-answer-disinformation-online> [<https://perma.cc/3TQE-29SR>] (noting that platform curation of content can amplify “incendiary content” and take users “into a rabbit hole of disinformation”).

¹⁰⁷ See *supra* notes 100–105 and accompanying text.

¹⁰⁸ See *supra* notes 69–73 and accompanying text.

the world, the UNGPs anchor assessments of such laws to the international human rights standard for freedom of expression. This standard gives companies a substantive benchmark with which to judge not only the legitimacy of governmental attempts to regulate falsehoods but also to assess whether their own corporate policies infringe on freedom of expression. The UNGPs also provide a host of helpful action steps, including human rights due diligence, that companies should take to ensure they are doing what is possible to avoid infringing on human rights when combatting disinformation.

Given the extraordinary power and reach of global social media companies in the worldwide information ecosystem,¹⁰⁹ those concerned with the impacts of disinformation on democracy should encourage such platforms to adhere to the UNGPs and the substantive standards in international human rights law to guide their responses to state regulation as well as the development of company policies. Otherwise, we risk heading not only toward a post-truth world, but also a post-human rights world.

¹⁰⁹ Facebook, for example, had 2.45 billion people using its platform by the third quarter of 2019. Josh Constone, *Facebook Shares Rise on Strong Q3, Users Up 2% to 2.45B*, TECHCRUNCH (Oct. 30, 2019), <https://techcrunch.com/2019/10/30/facebook-earnings-q3-2019/> [<https://perma.cc/Z54Q-JNAP>]. Google's YouTube had over 2 billion users in early 2020, which it notes is "almost one-third of the internet." *YouTube For Press*, YOUTUBE ABOUT, <https://www.youtube.com/about/press/> [<https://perma.cc/NQ3S-E7H6>] (last visited Jan. 24, 2020).