

Research Report by the Mandate of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, with the support of the International Justice Clinic at the University of California, Irvine School of Law

July 2020
Freedom of Expression and Oversight of Online Content Moderation
Supported by Knight Foundation

Table of Contents

	<i>Page</i>
I. Introduction.....	1
II. Overview of company responsibilities under human rights standards.....	2
III. Recent developments in content moderation oversight.....	5
IV. Case study: Facebook’s Oversight Board.....	7
A. Factual background.....	8
B. Preliminary Analysis.....	13
V. Conclusion.....	25

I. Introduction

1. In 2018, the mandate released a thematic report examining State regulation and company moderation of online content (the “2018 Report”). The 2018 Report provided an overview of the ways in which both public and private actors interfere with the individual’s right to freedom of opinion and expression. It emphasized, *inter alia*, radically better transparency by both States and companies and promoted accountability and remedy to protect the use of online platforms as forums for free expression, access to information and engagement in public life.¹ The Report underscored that companies, especially those in social media with significant power over user-generated content, must adopt and apply human rights standards at all stages of their operations.² The Report encouraged companies to pursue independent mechanisms of appeal and remedy, including cross-industry social media councils (“SMCs”).³
2. During the two years since the Report was submitted to the Human Rights Council, public debate and significant research have pushed forward an agenda for content moderation oversight (as a form of appeal and remedy) with human rights as a guiding force. While the 2018 Report noted early work on SMCs, the international non-governmental organization ARTICLE 19, in collaboration with experts at Stanford University, the University of California, Irvine and elsewhere, has further articulated how cross-industry regulation may improve the application of human rights standards by social media companies.⁴ Meanwhile, Facebook, a platform with nearly 2.6 billion active users worldwide, has developed a self-regulatory Oversight Board to address a selection of what might be considered “hard” content moderation decisions.
3. The mandate, with the support of the International Justice Clinic at the University of California, Irvine School of Law (the “Clinic”)⁵, has remained engaged with the developing space of content moderation oversight. For example, thematic reports on the human rights implications of online hate speech and artificial intelligence technologies in the information environment have called for effective remedies for human rights violations.⁶ Also, the

¹ A/HRC/38/35, para. 64.

² *See id.*, para. 70.

³ *Id.*, para. 72. The instant report will not reiterate the contents of the 2018 Report but should be read in light of its findings and recommendations.

⁴ *See* ARTICLE 19, “The Social Media Councils: Consultation Paper,” June 2019.

⁵ The International Justice Clinic at the University of California, Irvine School of Law works with international activists, non-governmental organizations and scholars across the globe to develop and implement strategies for the protection of human rights, particularly the right to freedom of opinion and expression. With the support of Knight Foundation, the Clinic has established the Human Rights and Social Media Initiative, which monitors and provides analysis on issues at the intersection of international human rights law and digital spaces. For more information, please refer to the Clinic’s website at: www.law.uci.edu/academics/real-life-learning/clinics/international-justice.html.

⁶ *See* A/74/486, para. 53 (“In short, the process of remediation must begin with an effective way for individuals to report potential violations of hate speech policies and must ensure protections against abuse of the reporting system as a form of hate speech. It should include a transparent and accessible process for appealing platform decisions, with companies providing a reasoned response that should also [be] publicly accessible.”); A/73/348, para. 60 (“Adverse impacts of AI systems on human rights must be remediable and remedied by the companies responsible.”).

mandate and the Clinic have participated in extensive discussions with civil society organizations and social media companies concerning SMC proposals.⁷ Further, they have monitored and assessed the beginnings of Facebook’s Oversight Board.⁸ With the support of a grant from Knight Foundation, the Clinic will continue to engage with this space and, in 2021, will release a report evaluating the status of content moderation oversight, by companies and governments; the report will include a detailed assessment of the next phase of Facebook’s Oversight Board.

4. Building on the 2018 Report, this report evaluates recent developments in content moderation oversight, with a particular focus on company initiatives and the emerging Oversight Board. Part II provides an overview of the human rights standards applicable to company oversight mechanisms. Part III reviews current proposals for social media councils and other regulatory models. Part IV sets forth an interim assessment of the Oversight Board at this early stage of development. The assessment looks at the Board’s structure, scope and functions and offers a preliminary analysis of significant human rights issues raised by the Board, as well as recommendations to address such issues. Part V concludes the report with a brief reflection on the future of content moderation oversight.

II. Overview of company responsibilities under human rights standards

5. The 2018 Report detailed the international human rights legal framework applicable to State and company regulation of online content.⁹ In light of its focus on company initiatives, particularly the Oversight Board, this report sets forth the human rights standards applicable to company oversight of content moderation.
6. While content moderation and its oversight implicate an array of substantive international human rights, such as rights to religious freedom and belief, assembly and association, privacy, public participation, among others,¹⁰ this report will focus primarily on the right to freedom of opinion and expression under Article 19 of the International Covenant on Civil and Political Rights (the “ICCPR”).¹¹ Under Article 19 of the ICCPR, everyone has the right to hold opinions without interference (Article 19(1)) and the right to exercise their freedom of expression (Article 19(2)). Article 19(2) defines freedom of expression as the “freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of [their] choice.”¹² Article 19 rights apply with equal force online and offline.¹³

⁷ See, e.g., GDPi, ARTICLE 19 and David Kaye, “Social Media Councils: From Concept to Reality,” February 2019.

⁸ In fact, the mandate sent a communication to Facebook in 2019 analyzing the first iteration of the Board’s charter. See Communication No. OL OTH 24/2019.

⁹ See A/HRC/38/35, paras. 5-11.

¹⁰ *Id.*, para. 5. See also Business for Social Responsibility (BSR), “Human Rights Review: Facebook Oversight Board,” December 2019, at 17 (listing a range of human rights relevant to content moderation and its oversight).

¹¹ International Covenant on Civil and Political Rights, 999 U.N.T.S. 171 [hereinafter “ICCPR”], art. 19. Article 19 codifies, as a binding obligation, the same rights found in Article 19 of the Universal Declaration of Human Rights.

¹² *Id.*

¹³ A/HRC/38/35, para. 1.

7. While the right to freedom of opinion is absolute, not subject to any restrictions, Article 19(3) recognizes that freedom of expression may be subject to a limited set of narrow restrictions. These restrictions must be provided by law and necessary to protect a legitimate interest, namely “[f]or respect of the rights or reputations of others. . . [or] the protection of national security or of public order (ordre public), or of public health or morals.”¹⁴ Any restriction must be implemented in a non-discriminatory manner.¹⁵ In short, limitations on freedom of expression must meet the well-established conditions of legality, necessity and proportionality, legitimacy and non-discrimination.¹⁶ These global standards are also reflected in European, inter-American and African legal instruments and jurisprudence.¹⁷ Article 20 of the ICCPR requires States to prohibit national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence, a standard that the UN Human Rights Committee, the monitoring body for the ICCPR, has clarified must be read together with the standards found in Article 19(3).¹⁸
8. The UN Guiding Principles on Business and Human Rights (the “Guiding Principles”)¹⁹ indicate how companies should respect rights — including Article 19 of the ICCPR — through policy, due diligence, implementation and remedy. Endorsed by the UN Human Rights Council in 2011, the Guiding Principles identify “global standards of expected conduct” by private companies and governments.²⁰ While many of the Principles are relevant to oversight, Principle 31 is particularly instructive and provides an overarching human rights framework for non-judicial grievance mechanisms. This Principle states:

In order to ensure their effectiveness, non-judicial grievance mechanisms, both State-based and non-State-based, should be:

(a) Legitimate: enabling trust from the stakeholder groups for whose use they are intended, and being accountable for the fair conduct of grievance processes;

(b) Accessible: being known to all stakeholder groups for whose use they are intended, and providing adequate assistance for those who may face particular barriers to access;

(c) Predictable: providing a clear and known procedure with an indicative time frame for each stage, and clarity on the types of process and outcome available and means of monitoring implementation;

¹⁴ ICCPR, art. 19.

¹⁵ *Id.*, art. 2; UN Human Rights Committee, General Comment No. 34 (2011) on Article 19: Freedoms of Opinion and Expression, para. 26..

¹⁶ ICCPR, art. 2; UN Human Rights Committee, *supra* note 15, para. 26; A/HRC/38/35, para. 7.

¹⁷ *See* African Charter on Human and Peoples’ Rights, art. 9; American Convention on Human Rights, art. 13; Convention for the Protection of Human Rights and Fundamental Freedoms, art. 10.

¹⁸ UN Human Rights Committee, *supra* note 15, paras. 50-52. *See also* A/74/486.

¹⁹ United Nations Guiding Principles on Business and Human Rights, March 21, 2011 [hereinafter “Guiding Principles”].

²⁰ *Id.*, Principle 11, Commentary.

(d) Equitable: seeking to ensure that aggrieved parties have reasonable access to sources of information, advice and expertise necessary to engage in a grievance process on fair, informed and respectful terms;

(e) Transparent: keeping parties to a grievance informed about its progress, and providing sufficient information about the mechanism's performance to build confidence in its effectiveness and meet any public interest at stake;

(f) Rights-compatible: ensuring that outcomes and remedies accord with internationally recognized human rights;

(g) A source of continuous learning: drawing on relevant measures to identify lessons for improving the mechanism and preventing future grievances and harms;

(h) Based on engagement and dialogue: consulting the stakeholder groups for whose use they are intended on their design and performance, and focusing on dialogue as the means to address and resolve grievances.²¹

These criteria “provide a benchmark for designing, revising or assessing” oversight mechanisms, in an effort to avoid “compounding a sense of grievance amongst affected stakeholders by heightening their sense of disempowerment and disrespect by the process.”²² Ultimately, as the Principles emphasize, “[a] grievance mechanism can only serve its purpose if the people it is intended to serve know about it, trust it and are able to use it.”²³

9. The 2018 Report, in urging companies and governments to adhere to the Guiding Principles, warned that “[o]paque forces are shaping the ability of individuals worldwide to exercise their freedom of expression,” calling for online platforms to demonstrate “radical transparency, meaningful accountability and a commitment to remedy.”²⁴ With respect to the latter, the Report underscored that “[c]ompanies should institute robust remediation programmes.”²⁵ To this end, the Report endorsed SMCs as one model for providing content moderation oversight and recommended that “[a]ll segments of the ICT sector that moderate content or act as gatekeepers should make the development of industry-wide accountability mechanisms (such as a social media council) a top priority.”²⁶

²¹ *Id.*, Principle 31. These criteria should be viewed and applied through a gender-sensitive lens. As the Gender Dimensions of the Guiding Principles on Business and Human Rights provide, “All effectiveness criteria for non-judicial grievance mechanisms should be interpreted in a gender-responsive manner, considering especially the intersectional nature of discrimination faced by women in accessing such mechanisms and in enforcing effective remedies against business enterprises.” A/HRC/41/43, at 26.

²² *Id.*, Principle 31, Commentary.

²³ *Id.*

²⁴ A/HRC/38/35, para. 64.

²⁵ *Id.*, para. 59 (explaining that “[i]f the failure to remediate persists, legislative and judicial intervention may be required.”).

²⁶ *Id.*, paras. 59, 72 (“There has been some convergence among several companies in their content rules, giving rise to the possibility of inter-company cooperation to provide remedies through a social media council, other ombudsman programmes or third-party adjudication.”).

III. Recent developments in content moderation oversight

10. Since the 2018 Report's endorsement, social media councils have continued to emerge as one of several models, described below, in the developing landscape of content moderation oversight. The SMC model envisions "a transparent, multi-stakeholder, voluntary compliance mechanism that would oversee content moderation practices of social media platforms on the basis of international standards on human rights."²⁷ Discussions among civil society organizations and social media companies have defined the general features of the model:²⁸ independence from government, commercial and special interests; establishment by an inclusive process with public consultation; democratic and transparent member selection and decision-making; broad representation reflecting societal diversity; a robust complaint mechanism applying human rights standards; and extensive transparency in service of public accountability.²⁹ Although the SMC concept and general features enjoy support among many stakeholders, there are different views on what the exact structure, scope and functions of such a mechanism (or set of mechanisms) should be.³⁰
11. A key point of divergence among stakeholders is the appropriate geographic scope: whether SMCs should be national, regional, global or some combination of the three levels.³¹ Each model has advantages and disadvantages. National SMCs would be relatively entrenched in the local context and, therefore, their decision-making would benefit from familiarity with the linguistic, social, cultural, economic and political complexities of their respective countries.³² However, as a recent report commissioned by the European Parliament points out, "there are risks of interference or even appropriation by States as well as difficulties in determining the jurisdiction of a given country."³³ By contrast, a global SMC could help foster uniformity in decision-making and create a consistent body of "case law,"³⁴ which could clarify how various content moderation approaches comport with international human rights standards.³⁵ Yet a global SMC could

²⁷ ARTICLE 19, "The Facebook Oversight Board: A significant step for Facebook and a small step for freedom of expression," 21 May 2020.

²⁸ In particular, a February 2019 international working meeting, convened by Stanford University's Global Digital Policy Incubator, ARTICLE 19 and the mandate, helped to progress and refine the SMC concept and culminated in a detailed report. See generally GDPi, et al., *supra* note 7.

²⁹ *Id.*; ARTICLE 19, *supra* note 4, at 10.

³⁰ See Pierre François Docquir, "The Social Media Council: Bringing Human Rights Standards to Content Moderation on Social Media," *Centre for International Governance Innovation*, 28 October 2019.

³¹ GDPi, et al., *supra* note 7, at 16-20.

³² See Docquir, *supra* note 30.

³³ European Parliament, "Online Platforms' Moderation of Illegal Content Online," *European Parliament Think Tank*, June 2020, at 61 n.129.

³⁴ The 2018 Report underscored the benefits of a public compilation of content moderation "case law," stating: "Ideally, companies should develop a kind of case law that would enable users, civil society and States to understand how the companies interpret and implement their standards. While such a 'case law' system would not involve the kind of reporting the public expects from courts and administrative bodies, a detailed repository of cases and examples would clarify the rules much as case reporting does. A social media council empowered to evaluate complaints across the ICT sector could be a credible and independent mechanism to develop such transparency." A/HRC/38/35, para. 63.

³⁵ See European Parliament, "Online Platforms' Moderation of Illegal Content Online," *European Parliament Think Tank*, June 2020, at 61 n.129; Docquir, *supra* note 30.

have difficulty accessing local contexts in its decision-making, as well as ensuring diverse representation.³⁶ Meanwhile, as civil society organizations have explained, regional SMCs could “go a long way in alleviating concerns about state capture, would still be grounded in contexts that are closer to local, and would have an easier time facilitating the participation of local experts”; however, they “would still need to deal with concerns about representation and diversity.”³⁷ These three models need not be mutually exclusive and could work together in a complementary system.³⁸

12. Currently, a group of stakeholders led by ARTICLE 19 aims to set up the first SMC in Ireland.³⁹ Once established, the SMC would have a national scope and include civil society organizations, media regulators and representatives from Twitter, Google and Facebook.⁴⁰ ARTICLE 19 is in the process of organizing a workshop with stakeholders to finalize the details of the SMC and draft its governing documents.⁴¹
13. It is worth noting that discussions concerning non-governmental oversight are taking place in an environment with significant potential for governmental regulation. Most prominently, the European Union has launched public consultations regarding a proposed Digital Services Act.⁴² The French Constitutional Council recently found unlawful a new French hate speech law that would have imposed broad requirements on social media companies.⁴³ Brazil is considering legislation targeting “fake news” that would impose stringent requirements on social media companies operating there.⁴⁴ And the U.S. is exploring various options for social media regulation.⁴⁵ These are just a few examples reflecting a possible global trend toward State regulation. Meanwhile, creative approaches are regularly suggested. For instance, one proposed model for content moderation oversight would involve specialized, online “e-courts.”⁴⁶ Under this model, States would set up an e-courts network with the capacity to render expedient decisions by specially trained

³⁶ European Parliament, “Online Platforms’ Moderation of Illegal Content Online,” *European Parliament Think Tank*, June 2020, at 61-62.

³⁷ GDPi, et al., *supra* note 7, at 19.

³⁸ See Docquir, *supra* note 30.

³⁹ Aaron Rogan, “Plans for world’s first ‘social media council’ in Ireland,” *Business Post*, 12 April 2020.

⁴⁰ *Id.* Recently, ARTICLE 19 sent an email to stakeholders, explaining why Ireland was a good fit for the project: “There are many elements that lead to the choice of Ireland as our target country in the EU: firstly, Ireland has an important tradition of media self-regulation, as evidenced by the existence of a vibrant press council; secondly, the attention that the Irish government has given to increasing online safety through the discussion of a bill and the establishment of a Commissioner on the topic; and thirdly, Ireland hosts the European headquarters of social media platforms such as Facebook, Twitter and YouTube, which could give proximity of the SMC to these platforms.” *Id.*

⁴¹ *Id.*

⁴² European Commission Press Release IP/20/962, “Commission launches consultation to seek views on Digital Services Act package,” 2 June 2020.

⁴³ ARTICLE 19, “France: Constitutional Council declares French hate speech ‘Avia’ law unconstitutional,” 18 June 2020.

⁴⁴ Electronic Frontier Foundation, “New Hasty Attempt to Tackle Fake News in Brazil Heavily Strikes Privacy and Free Expression,” 7 June 2020.

⁴⁵ See, e.g., Tony Romm, “Justice Department recommends new legislation holding Facebook, Google and Twitter liable for some online content,” *The Washington Post*, 17 June 2020.

⁴⁶ See Annenberg Public Policy Center, “Freedom and Accountability: A Transatlantic Framework for Moderating Speech Online,” *The University of Pennsylvania*, June 2020. This model is based on the idea that “[i]n a democracy, moderation decisions that implicate law or human rights require judicial redress.” *Id.*, at 8.

magistrates and render a body of public jurisprudence to guide content moderation decisions.⁴⁷ This model seeks to offer “legitimacy through due process and independence, and protect democracy by bringing decisions about the legality of content into public view.”⁴⁸

14. Companies are also proposing self-regulatory oversight mechanisms. For example, TikTok recently announced that it is forming a “committee of outside experts to advise on and review the company’s content moderation policies covering a wide range of topics, including child safety, hate speech, misinformation, bullying, and other potential issues.”⁴⁹ This “Content Advisory Council” is to provide “unvarnished views” to TikTok regarding its content moderation decisions.⁵⁰ Also, TikTok recently announced the creation of two “Transparency Centers” to allow outside experts the opportunity to directly observe and give feedback on how the company moderates content on a daily basis.⁵¹ On May 15, 2020, the mandate sent a communication to TikTok, urging the company to incorporate human rights, particularly the right to freedom of expression, into its content moderation policies and practices.⁵²
15. Another company-specific mechanism, Facebook’s Oversight Board, may be the most salient development in the space of content moderation oversight. The following case study focuses on the Board but should be understood against the backdrop of a broader landscape, including the foregoing initiatives.

IV. Case Study: Facebook’s Oversight Board

16. As of the time of writing, Facebook is in the process of constituting, training and launching its Oversight Board. The stated purpose of the Board is “to promote free expression by making principled, independent decisions regarding content on Facebook and Instagram and by issuing recommendations on the relevant Facebook company content policy.”⁵³ It should be noted at the outset that, due to the limited mandate and capacity of the Board, discussed below, the Board will not address all (or even most) of the social, political and

⁴⁷ *Id.*, at 28.

⁴⁸ *Id.*, at 27; *See also* OBSERVACOM, “Contributions for the democratic regulation of big platforms to ensure freedom of expression online,” at 19 (“Individual cases where there is a violation of user rights and that are not satisfactorily resolved within the internal scopes and mechanisms for dispute resolution should be resolved by judicial bodies, Public Defenders or similar independent and specialized public bodies.”).

⁴⁹ Vanessa Pappas, “An update on TikTok’s efforts in the US,” *TikTok Newsroom*, 20 February 2020; Elizabeth Culliford, “TikTok Reveals First Members of New U.S. Content Moderation Committee,” *U.S. News*, 18 March 2020.

⁵⁰ Vanessa Pappas, “Introducing the TikTok Content Advisory Council,” *TikTok Newsroom*, 18 March 2020; Culliford, *supra* note 49. Likewise, the video live streaming service Twitch recently announced the creation of a “Safety Advisory Council.” This body will advise on a number of topics, including: “[d]rafting new policies and policy updates; [d]eveloping products and features to improve safety and moderation; [p]romoting healthy streaming and work-life balance habits; [p]rotecting the interests of marginalized groups; and [i]dentifying emerging trends that could impact the Twitch experience” Twitch, “Introducing the Twitch Safety Advisory Council,” *Twitch Blog*, 14 May 2020.

⁵¹ Vanessa Pappas, “TikTok to launch Transparency Center for moderation and data practices,” *TikTok Newsroom*, 11 March 2020; Roland Cloutier, “TikTok’s security and data privacy roadmap,” *TikTok Newsroom*, 9 June 2020.

⁵² Communication No. OL OTH 37/2020.

⁵³ Oversight Board, <https://www.oversightboard.com/>.

public harms that many allege to be the company’s responsibility⁵⁴ — leaving broad swaths of rightsholders without access to a remedy and potentially “compounding a sense of grievance.”⁵⁵

17. This section sets forth the factual contours of the Board, describing its membership and oversight processes.⁵⁶ The section then analyzes the human rights implications of the Board at this fledgling stage.

A. Factual background

18. In 2018, Facebook founder and CEO Mark Zuckerberg stated, “I do not believe individual companies can or should be handling so many of these issues of free expression and public safety on their own.”⁵⁷ Accordingly, Facebook set out to create an independent body to oversee the company’s content moderation decisions.⁵⁸ The Board will provide users and Facebook, itself, access to an external and limited review process beyond the internal review system already in place.⁵⁹

19. Facebook has committed \$130 million to fund the Oversight Board for its first six years.⁶⁰ This amount will cover compensation for the Board’s members and administrative staff, among other expenses.⁶¹ The funding will be administered by the Oversight Board Trust, responsible for “ensuring that the board is operating according to its purpose and its governing documents.”⁶² Facebook currently appoints all Oversight Board Trustees.⁶³ The structure of the Trust allows for the possibility of other social media companies joining the Board and appointing their own Trustees.⁶⁴

20. In 2019, Facebook released a “draft charter” for the Board and then conducted a months-long global consultation, soliciting feedback on the document from a range of

⁵⁴ For example, Dipayan Ghosh has argued for the Board to provide oversight of the company’s consumer data practices, strategic acquisitions and data governance to protect against anticompetitive practice. Dipayan Ghosh, “Facebook’s Oversight Board Is Not Enough,” *Harvard Business Review*, 16 October 2019. However, at least for now, the Board’s limited capacity seems to preclude such a vast expansion of scope.

⁵⁵ See Guiding Principles, Principle 31, Commentary.

⁵⁶ In conducting factual research for this paper, the mandate and the Clinic benefitted from interviews with Facebook employees, as well as other social media companies and civil society organizations.

⁵⁷ Mark Zuckerberg, “A Blueprint for Content Governance and Enforcement,” *Facebook Newsroom*, 15 November 2018.

⁵⁸ See Ben Gilbert, “Facebook is spending \$130 million to create a ‘Supreme Court’ that can overrule Mark Zuckerberg – here’s everything we know about it,” *Business Insider*, 16 February 2020.

⁵⁹ Mark Zuckerberg, “Facebook’s commitment to the oversight board,” *Facebook Newsroom*, September 2019.

⁶⁰ Brent Harris, “An Update on Building a Global Oversight Board,” *Facebook Newsroom*, 12 December 2019.

⁶¹ *Id.*

⁶² Oversight Board Bylaws, January 2020, Introduction [hereinafter “Bylaws”], art. 4 § 1.2 (noting, “The trust will not have a role in reviewing cases or interfere with the board’s exercise of its independent judgment on substantive content issues.”).

⁶³ Oversight Board Charter, September 2019 [hereinafter “Charter”], art. 5 § 2.

⁶⁴ See Issie Lapowsky, “Facebook tells us how its new board will oversee Mark Zuckerberg,” *Protocol*, 6 May 2020; Issie Lapowsky, “How Facebook’s oversight board could rewrite the rules of the entire internet,” *Protocol*, 6 May 2020. See generally Oversight Board Trust, “Trust Agreement,” 16 October 2019. The Clinic will be addressing this possibility, as well as the interests of other private actors, in further reporting.

stakeholders.⁶⁵ Based on this feedback, Facebook subsequently released a revised charter, as well as a set of bylaws.⁶⁶ These governing documents set forth the parameters for, *inter alia*, Board membership and oversight processes.

Board membership

21. On May 6, 2020, Facebook announced the first twenty members of the Board, including four co-chairs.⁶⁷ In constituting the initial Board, Facebook selected the four co-chairs, and then the company and the co-chairs, collectively, recruited candidates for the remaining sixteen vacancies.⁶⁸ Going forward, the Board, without input from Facebook, will assume responsibility for recruitment.⁶⁹ Once recruited, all candidates must be formally approved by the Trustees to become members.⁷⁰ There will eventually be approximately forty members on the Board;⁷¹ however, the bylaws state that “the total number of members, at any given time, may be calibrated to accommodate case volume and scope.”⁷² Facebook is expected to announce the remaining Board members in the coming months.⁷³
22. Members serve three-year terms with the possibility of renewal for up to nine years.⁷⁴ With Board approval, the Trustees may remove a Board member prior to the expiration of their term.⁷⁵ They may do so only if that member is deemed to have violated the Board’s code of conduct, which Facebook prepared.⁷⁶ According to the bylaws, “[m]embers will not be removed due to content decisions they have made.”⁷⁷ Requests for removal may come from the Board, the public or Trustees themselves.⁷⁸
23. Board membership must strike a geographical balance.⁷⁹ “In particular, this means that board membership should encompass the following regions: United States and Canada; Latin America and the Caribbean; Europe; Sub-Saharan Africa; Middle East and North Africa; Central and South Asia; and Asia Pacific and Oceania.”⁸⁰ Of the first twenty Board members, five are from the United States, two are from Latin America, four are from

⁶⁵ Brent Harris, “Global Feedback and Input on the Facebook Oversight Board for Content Decisions,” *Facebook Newsroom*, 27 June 2019.

⁶⁶ Brent Harris, “Preparing the Way Forward for Facebook’s Oversight Board,” *Facebook Newsroom*, 28 January 2020. By their own terms, the bylaws will not be operational until their adoption by Facebook, the Trustees and the Board. Bylaws, Introduction.

⁶⁷ Nick Clegg, “Welcoming the Oversight Board,” *Facebook Newsroom*, 6 May 2020.

⁶⁸ Bylaws, art. 1 § 1.2.2.

⁶⁹ *Id.*, art. 1 § 1.4.2.

⁷⁰ *Id.*, art. 4 § 2.1.1.

⁷¹ *Id.*, art. 1 § 1.4. See Steven Levy, “Why Mark Zuckerberg’s Oversight Board May Kill His Political Ad Policy,” *Wired*, 28 January 2020.

⁷² Bylaws, art. 1 § 1.4.

⁷³ Steven Overly and Alexandra S. Levine, “Facebook announces first 20 picks for global oversight board,” *Politico*, 6 May 2020.

⁷⁴ Bylaws, art. 1 § 1.4.2.

⁷⁵ *Id.*, art. 4 § 2.1.2, art. 1 § 1.2.2.

⁷⁶ *Id.*, art. 4 § 2.1.2

⁷⁷ *Id.*

⁷⁸ *Id.*

⁷⁹ *Id.*, art. 1 § 1.4.1.

⁸⁰ *Id.*

Europe, two are from sub-Saharan Africa, two are from the Middle East and North Africa, two are from Central and South Asia and three are from Asia Pacific.⁸¹ Two of the four co-chairs are from the U.S.⁸²

24. In terms of qualifications, the charter provides that members must “have demonstrated experience at deliberating thoughtfully and as an open-minded contributor on a team; be skilled at making and explaining decisions based on a set of policies or standards; and have familiarity with matters relating to digital content and governance, including free expression, civic discourse, safety, privacy and technology.”⁸³ The governing documents do not require Board members to have a specific background in international human rights. The members are to participate in training sessions on Facebook’s policies, values and enforcement practices, as well as international human rights standards.⁸⁴

Oversight processes

25. The Board may decide individual cases referred by users;⁸⁵ meanwhile, it can decide individual cases, broad policy questions and urgent matters referred by Facebook.⁸⁶ Currently, the Board does not have competence to review referrals by civil society organizations or other stakeholders (other than as user referrals). Nor does the Board have competence to take up matters on its own accord, in the absence of a referral from a user or Facebook.⁸⁷
26. In order to refer individual cases to the Board, users must first exhaust Facebook’s internal appeals process,⁸⁸ which allows users to seek and obtain the company’s review of its own content moderation decisions.⁸⁹ After Facebook notifies a user that their internal appeal has been unsuccessful, they have fifteen (15) days to prepare and submit their referral to the Board.⁹⁰ The Board may only decide user referrals involving certain types of user-generated content (including posts, photos, videos and comments) that was removed by Facebook.⁹¹ At some indeterminate time “[i]n the future,” users will be able to refer cases involving other types of content and moderation actions other than removal.⁹² Until then, notable examples of cases excluded from review are those involving political

⁸¹ Clegg, *supra* note 67; Oversight Board, “Announcing the First Members of the Oversight Board,” *Oversight Board*, 6 May 2020.

⁸² *Id.*

⁸³ Charter, art. 1 § 2.

⁸⁴ Bylaws, art. 1 § 1.4.4.

⁸⁵ *Id.*, art. 3 § 1.1.

⁸⁶ *Id.*, art. 2 § 2.1

⁸⁷ Recent statements from at least one Board member, however, indicate that the Board may create a channel to do so. *See, e.g.*, IANS, “Facebook Oversight Board may take up cases by itself: Sudhir Krishnaswamy,” *The New Indian Express*, 10 May 2020.

⁸⁸ Bylaws, art. 3 § 1.1; Brent Harris, “Establishing Structure and Governance for an Independent Oversight Board,” *Facebook Newsroom*, 17 September 2019.

⁸⁹ *See* Monika Bickert, “Publishing Our Internal Enforcement Guidelines and Expanding Our Appeals Process,” *Facebook Newsroom*, 24 April 2018.

⁹⁰ Bylaws, art. 3 § 1.1.

⁹¹ *Id.*

⁹² *Id.*, art. 3 § 1.1.2.

advertisements, events and groups, as well as content allowed to remain on the platform and subjected to algorithmic down-ranking — such as misinformation.⁹³ In general, moderation actions taken pursuant to legal obligations, including localized blocking of content due to illegality under State laws (known as “geo-blocking”), will not be reviewable by the Board.⁹⁴ The Board must decide cases referred by users within ninety (90) days.⁹⁵ The governing documents do not indicate that there will be any interim relief provided to users while a case is pending before the Board.⁹⁶

27. Facebook can refer matters to the Board either on a regular basis or under emergency circumstances, with the latter triggering an expedited review.⁹⁷ Facebook’s regular (that is, non-emergency) case submissions must be “significant and difficult”: in this context, “significant” means that “the content in question involves real-world impact and issues that are severe, large-scale, and/or important for public discourse,” while “difficult” means “the content raises questions about current policies or their enforcement, with strong arguments on both sides for either removing or leaving up the content under review.”⁹⁸ Unlike users, Facebook may refer cases involving a broad range of content types.⁹⁹ Additionally, Facebook can request policy advisory statements from the Board on virtually any content policy matter for which it seeks guidance (and involving most types of content and moderation actions).¹⁰⁰ Moreover, Facebook can trigger an emergency review process in cases of “exceptional circumstances, including when content could result in real world consequences.”¹⁰¹ The Board must decide Facebook’s regular case submissions and requests for policy advisory statements within ninety (90) days and emergency cases within thirty (30) days.¹⁰²

28. When the Board receives a referral from a user or Facebook, it will decide whether to take up the matter pursuant to criteria to be selected by the Board.¹⁰³ “In its selection, the board

⁹³ See *id.* Recently, Facebook has removed misinformation related to COVID-19; presumably, these removals may be reviewable by the Board. See Kang-Xing Jin, “Keeping People Safe and Informed About the Coronavirus,” *Facebook Newsroom*, 2 July 2020.

⁹⁴ Bylaws, art. 2 §§ 1.2.1-1.2.2. According to the bylaws, decisions made pursuant to legal obligations are those involving the following situations: “Where the underlying content has already been blocked, following the receipt of a valid report of illegality, and not removed for a Community Standards violation; [w]here the underlying content is criminally unlawful in a jurisdiction with a connection to the content (such as the jurisdiction of the posting party and/or the reporting party) and where a board decision to allow the content on the platform could lead to criminal liability for Facebook, Facebook employees, the administration, or the board’s members; or [w]here the underlying content is unlawful in a jurisdiction with a connection to the content (such as the jurisdiction of the posting party and/or the reporting party) and where a board decision to allow the content on the platform could lead to adverse governmental action against Facebook, Facebook employees, the administration, or the board’s members.” *Id.*, art. 2 § 1.2.2.

⁹⁵ *Id.*, art. 1 § 3.1.

⁹⁶ See Gabriella Casanova Carlos Lopez and Sam Zarifi, “Some questions regarding Facebook’s oversight board and remediation of human rights impacts (Part I),” *Opinio Juris*, 3 March 2020.

⁹⁷ Bylaws, art. 2 § 2.1.

⁹⁸ *Id.*, art. 2 § 2.1.1.

⁹⁹ *Id.*, art. 2 § 2.1.

¹⁰⁰ *Id.*, art. 2 § 2.1.3.

¹⁰¹ *Id.*, art. 2 § 2.1.2.

¹⁰² *Id.*, art. 1 § 3.1, art. 2 § 2.1.2.

¹⁰³ *Id.*, art. 1 § 1.2.1.

will seek to consider cases that have the greatest potential to guide future decisions and policies.”¹⁰⁴ One member has predicted that the Board will only have the capacity to decide approximately 100-200 cases each year.¹⁰⁵ After a case is selected for review, it will be referred to a panel of five members, one of whom will be from the geographical region “which the content primarily affects.”¹⁰⁶ During case deliberations, this panel may, at its discretion, request information and analysis from civil society organizations, as well as a “global pool of outside subject-matter experts” selected by the Board.¹⁰⁷ The panel may also request information from Facebook, but Facebook is not required to comply with such requests.¹⁰⁸ The panel will then draft a written decision, setting forth its reasoning and, if desired, policy guidance for Facebook.¹⁰⁹ The only form of relief that the Board is authorized to order is the reinstatement of content wrongfully removed.¹¹⁰ The Board as a whole will review the panel’s decision and either approve the decision, provide feedback for the panel’s consideration or refer the case to a new panel for another review.¹¹¹ Case deliberations will be confidential;¹¹² however, the Board will send communications to users whose cases are reviewed upon selection for review and finalization of the decision.¹¹³

29. Once finalized, each decision (containing the panel’s reasoning and policy guidance, if any) will be posted on the Board’s public website.¹¹⁴ Prior to posting, the Board will delete identifying details of any users, at their request.¹¹⁵ The Board will also post its case selection criteria, as well as data on the volume and types of cases referred and selected for review.¹¹⁶ Additionally, the Board will issue annual reports containing summary data on cases reviewed, a breakdown of case submissions by region, information on Facebook’s implementation of decisions and “[a]n analysis of how the board’s decisions have considered or tracked the international human rights implicated by a case.”¹¹⁷
30. This annual report requirement is one of only a few references to international human rights in the Board’s governing documents. Although the Board will receive training in international human rights (and the co-chairs have stated that they intend to apply human rights “principles”¹¹⁸), the Board is not required to apply international human rights

¹⁰⁴ Charter, art. 2 § 1.

¹⁰⁵ Omer Kabir, “Facebook Won’t Fudge Content Oversight, Says Israeli Legal Expert,” *CTech*, 7 May 2020.

¹⁰⁶ Bylaws, art. 1 § 3.1.3.

¹⁰⁷ *Id.*, art. 1 § 3.1.4.

¹⁰⁸ *Id.*, art. 1 § 3.1.3, art. 2 § 2.2.2.

¹⁰⁹ *Id.*, art. 1 § 3.1.7. The decision may also contain statements of any concurring or dissenting viewpoints. *Id.*

¹¹⁰ *See id.*, art. 1 § 3.1.7; Charter, art. 3 § 4.

¹¹¹ Bylaws, art. 1 § 3.1.8.

¹¹² *Id.*, Code of Conduct, at 41.

¹¹³ *Id.*, art. 3 § 1.2.

¹¹⁴ *Id.*, art. 1 § 3.2; Charter, art. 3 § 6. The Board’s decisions will initially be posted in English. Bylaws, art. 1 § 4.3. Within twenty-one (21) days of release, the administration will translate each decision into the Board’s eighteen (18) official languages. *Id.*, art. 1 § 3.2.

¹¹⁵ Oversight Board, “Frequently Asked Questions,” *Oversight Board*, March 2020. Notably, the Board’s review of cases will not be anonymous; the Board will receive relevant information about users referring cases, such as their names, locations, age and gender. *Id.*

¹¹⁶ Bylaws, art. 1 § 1.2.1.

¹¹⁷ *Id.*, art. 1 § 4.1.

¹¹⁸ Catalina Botero-Marino, Jamal Greene, Michael W. McConnell and Helle Thorning-Schmidt, “We Are a New Board Overseeing Facebook. Here’s What We’ll Decide,” *The New York Times*, 6 May 2020 (“And we are all

standards in its decision-making. Rather, as the charter provides, “[t]he board will review content enforcement decisions and determine whether they were consistent with Facebook’s content policies and values” and “[w]hen reviewing [content enforcement] decisions, the board will pay particular attention to the impact of removing content in light of human rights norms protecting free expression.”¹¹⁹

31. Decisions on individual cases, whether referred by users or the company, will be binding and must be implemented by Facebook within seven (7) days.¹²⁰ Further, the company must search for “identical content with parallel context” on its platforms, and if it “determines that it has the technical and operational capacity to take action on that content as well, it will do so promptly.”¹²¹ By contrast, decisions on policy matters will not be binding on Facebook; instead, Facebook will consider such decisions and issue a public response within thirty (30) days explaining its plans, if any, for implementation.¹²²
32. The charter and bylaws leave room for the Board to amend these documents, but amending most of the provisions requires approval by Facebook, the Trustees or both Facebook and the Trustees¹²³ Those few provisions that can be unilaterally amended by the Board relate to internal operations and administration.¹²⁴ Facebook, on the other hand, can unilaterally amend bylaws provisions regarding the types of content the Board can review and appeal-submission procedures.¹²⁵ Amendments to the bylaws are not permissible if they contravene the Board’s charter.¹²⁶
33. The Board has indicated that it will start reviewing cases by the end of 2020.¹²⁷

B. Preliminary Analysis

34. In designing the Oversight Board, Facebook has taken steps that appear consistent with its responsibilities under the Guiding Principles. Such steps include conducting a global consultation with stakeholders, creating the Trust as a layer of separation between the

committed to freedom of expression within the framework of international norms of human rights. We will make decisions based on those principles and on the effects on Facebook users and society, without regard to the economic, political or reputational interests of the company.”) The appointment of Thomas Hughes, formerly of ARTICLE 19, as Director of Oversight Board Administration is another positive indication that the Board may pursue a human rights approach. See Sam Shead, “Human rights expert to keep Zuckerberg in check,” *BBC News*, 28 January 2020.

¹¹⁹ Charter, art. 2 § 2.

¹²⁰ Bylaws, art. 2 § 2.3. Facebook is not required to implement such decisions if doing so would violate the law, by the company’s determination. *Id.*

¹²¹ *Id.*; Charter, art. 4.

¹²² Bylaws, art. 2 § 2.3.2. See also Gabriella Casanova Carlos Lopez and Sam Zarifi, “Some questions regarding Facebook’s oversight board and remediation of human rights impacts (Part I),” *Opinio Juris*, 3 March 2020.

¹²³ Charter, art. 6 § 1; Bylaws, art. 5 § 1.

¹²⁴ Bylaws, art. 5 § 1.

¹²⁵ *Id.*

¹²⁶ *Id.*

¹²⁷ Oversight Board, “An update on the Oversight Board’s progress,” *Oversight Board*, 3 June 2020; Sam Shead, “Facebook Oversight Board says it won’t get started until late fall,” *CNBC*, 8 July 2020.

company and the Board and requiring the Board to publicize case-related information and a compendium of its decisions.

35. There are, however, four key areas in which the Board seems to be out of alignment with the Principles: (1) application of international human rights standards, (2) remedial scope, (3) transparency and stakeholder engagement and (4) independence and global legitimacy. These topics and the recommendations below are by no means exhaustive. Also, they are subject to change as the Board begins operations and as the Clinic continues its monitoring efforts.

The Board's application of international human rights standards

36. In light of the Guiding Principles, the governing documents should provide for the Board to apply international human rights standards in its decision-making. As noted above, Principle 31 provides that non-judicial grievance mechanisms should be “[r]ights-compatible: ensuring that outcomes and remedies accord with internationally recognized human rights.”¹²⁸ Principle 12 clarifies that the responsibility of companies to respect human rights refers to all internationally recognized human rights, including the ICCPR and International Covenant on Economic and Social Rights (the “ICESCR”). “Because business enterprises can have an impact on virtually the entire spectrum of internationally recognized human rights,” Principle 12 notes, “their responsibility to respect applies to all such rights.”¹²⁹ Principle 16 states that companies should express their commitments to human rights through formal, public statements of policy.¹³⁰
37. Early indications from Facebook and Board members suggest some measure of human rights orientation. The Board is to receive training in international human rights, and the co-chairs have signaled a commitment to human rights principles. The director of the Board’s administration and several of the members have extensive human rights expertise, despite the lack of membership criteria requiring a human rights background. The charter and bylaws make sporadic reference to human rights. In particular, the charter provides that when reviewing content enforcement decisions, the Board is to “pay particular attention” to the impact of removing content “in light of” human rights norms for free expression.¹³¹
38. This charter provision, however, calls for the Board to “pay particular attention” to human rights norms — not apply them — and does not require the Board to consider any human rights other than freedom of expression.¹³² Simply paying attention to human rights will not, on its own, “ensur[e] that outcomes and remedies accord with internationally recognized human rights,” as compelled by Principle 31.¹³³ Application of human rights standards is the ideal means to satisfy this criterion. Also, under Principle 12, the governing

¹²⁸ Guiding Principles, Principle 31(f).

¹²⁹ *Id.*, Principle 12, Commentary.

¹³⁰ *Id.*, Principle 16(a)-(e).

¹³¹ Charter, art. 2 § 2.

¹³² Additionally, it is unclear whether this provision applies to the Board’s review of requests for policy advisory statements, as these may not always constitute “content enforcement decisions” (an undefined term).

¹³³ Guiding Principles, Principle 31(f).

documents should include the full spectrum of human rights. By only identifying freedom of expression, they imply that this right is more important than, or exclusive of, other rights relevant to content moderation.¹³⁴ International law makes clear that human rights are indivisible, interdependent and interrelated; they must be applied “in a fair and equal manner, on the same footing, and with the same emphasis.”¹³⁵

39. Instead of requiring the application of human rights standards, the governing documents establish that the Board must directly apply Facebook’s content policies, including its Community Standards,¹³⁶ and five core values.¹³⁷ There are aspects of Facebook’s content policies and its core values (voice, authenticity, safety, privacy and dignity) that may have analogues in human rights law.¹³⁸ However, civil society organizations, including ARTICLE 19, have identified important ways in which they diverge from international human rights standards.¹³⁹ The 2018 Report found that company standards based on “vague assertions of community interests” have “created unstable, unpredictable and unsafe environments for users and intensified government scrutiny.”¹⁴⁰ Indeed, these policies and values are Facebook’s unilateral effort to define for itself its public responsibilities; they did not come from an elected or judicial body and lack democratic legitimacy.¹⁴¹ Facebook is free to change them at any time (and does change them regularly) to suit company interests.¹⁴²
40. As a practical matter, it is unclear how the Board will give weight to Facebook’s content policies and each of the values in a given case, as well as how it will render consistent decisions across cases and policy decisions.¹⁴³ Commentators have speculated that the Board will likely develop a balancing test among values; for example, “voice” might be pitted against “safety” and “dignity” in cases involving hate speech.¹⁴⁴ Any such balancing

¹³⁴ BSR has created a non-exhaustive list of such rights, which include rights related to physical harm and bodily integrity risk (e.g., the right to life); civil liberties risk (e.g., the freedom of assembly and association); and risk to basic needs (e.g., the right to education). BSR, *supra* note 10, at 17.

¹³⁵ A/CONF.157/23, Vienna Declaration and Programme of Action, para. 5. As BSR has explained, “a human rights-based approach implies that the Oversight Board should be aware of the human rights impacts at stake in each case and should not limit itself to considerations of freedom of expression.” BSR, *supra* note 10, at 17.

¹³⁶ The content policy for Facebook’s platform is entitled “Community Standards,” while Instagram’s content policy is entitled “Community Guidelines.”

¹³⁷ Charter, art. 2 § 2. *See also* Bylaws art. 1 § 3. (“The board will review and decide on content in accordance with Facebook’s content policies and values.”)

¹³⁸ *See* Monika Bickert, “Updating the Values That Inform Our Community Standards,” *Facebook Newsroom*, 12 September 2019.

¹³⁹ *See, e.g.*, ARTICLE 19, “Facebook Community Standards,” June 2018.

¹⁴⁰ A/HRC/38/35, para. 41.

¹⁴¹ Chinmayi Arun, “The Facebook Oversight Board: An Experiment in Self-Regulation,” *Just Security*, 6 May 2020. Chinmayi Arun contends that this issue might have been avoided or mitigated by importing existing democratically legitimate goals from another source, such as international law. She concludes, “Given the global application of these rules, international human rights law is the only source of norms that might be acceptable to the majority of Facebook’s users.” *Id.*

¹⁴² *See* Bickert, *supra* note 138. A list of Facebook’s updates to its Community Standards within the last two months can be found at <https://www.facebook.com/communitystandards/recentupdates/>.

¹⁴³ *See generally* Evelyn Douek, “What Kind of Oversight Board Have You Given Us?” *The University of Chicago Law Review*, 11 May 2020.

¹⁴⁴ *See, e.g., id.*

test, however, would differ from the approach required under international human rights standards, specifically ICCPR Article 19.¹⁴⁵ Article 19(3)'s requirements of legality, necessity and proportionality and legitimacy, as well as Article 2's prohibition of discrimination, provide guidance on how Facebook should approach content decisions in a manner that respects freedom of expression and other rights. As explained in the 2018 Report, a human rights approach offers a well-established set of tools and a common vocabulary for addressing difficult content moderation issues.¹⁴⁶ These touchstones are especially important given the relatively short timeframe for assessing and deciding cases (ninety (90) days for regular case submissions and requests for policy advisory statements and thirty (30) days for emergency cases).

41. The Board should not be obligated to apply Facebook's content policies and values even as they contravene international human rights. Applying human rights standards as an overarching framework would ensure that the Board's decision-making is "[r]ights-compatible"¹⁴⁷ and enable Facebook to cultivate "an inclusive environment that accommodates the varied needs and interests of [] users while establishing predictable and consistent baseline standards of behavior."¹⁴⁸ In light of Principle 16, a written human rights policy should be adopted, stating that the Board will apply international human rights standards in its decision-making. At a minimum, these standards should include the rights set forth in the ICCPR and ICESCR.

Broadening the remedial scope of the Board

42. The Guiding Principles provide a compelling basis for expanding the Board's jurisdiction and authority. As noted above, Principle 31 requires oversight bodies to be compatible with human rights,¹⁴⁹ such rights include the right to a remedy for each and every human rights violation.¹⁵⁰ Similarly, Principle 29 establishes that company oversight mechanisms should identify and address "any legitimate concerns," noting that otherwise concerns "may over time escalate into more major disputes and human rights abuses."¹⁵¹ Acknowledging that it may not always be possible to address all adverse human rights impacts simultaneously, Principle 24 provides that "[w]here it is necessary to prioritize actions to address actual and potential adverse human rights impacts, business enterprises should first seek to prevent and mitigate those that are most severe or where delayed response would make them irremediable."¹⁵² Also, Principle 20 stresses the importance of tracking the

¹⁴⁵ The 2018 Report called for companies to "incorporate directly into their terms of service and 'community standards' relevant principles of human rights law that ensure content-related actions will be guided by the same standards of legality, necessity and legitimacy that bind State regulation of expression." A/HRC/38/35, para. 45. See generally David Kaye, "Against Balancing," *Medium*, 4 May 2020.

¹⁴⁶ A/HRC/38/35, para. 42-43. Furthermore, a consistently human rights-based approach to content decisions will enable Facebook to "stand on firmer ground when [it] seek[s] to hold States accountable to the same standards," and resist State attempts to exploit its content policies to censor content. *Id.*, para. 42.

¹⁴⁷ Guiding Principles, Principle 31(f).

¹⁴⁸ A/HRC/38/35, para. 43.

¹⁴⁹ Guiding Principles, Principle 31(f).

¹⁵⁰ See, e.g., ICCPR, art. 2(3) ("Each State Party to the present Covenant undertakes: (a) To ensure that any person whose rights or freedoms as herein recognized are violated shall have an effective remedy . . .")

¹⁵¹ Guiding Principles, Principle 29, Commentary (emphasis supplied).

¹⁵² *Id.*, Principle 24.

effectiveness of company responses to human rights impacts, explaining that oversight mechanisms can provide important feedback on the effectiveness of such responses.¹⁵³

43. Facebook has indicated that it is taking an incremental approach to the Board’s remedial scope: initially only allowing user referrals involving user-generated posts that the company removed from the platform, while gesturing toward expanding this subject matter scope “[i]n the future.” Although an incremental approach may be reasonable in the context of this experiment, it means that, for now, the Board’s scope is severely limited relative to the universe of cases in need of oversight. Thus, the Board leaves myriad rightsholders without a remedy and is not set up to decide “any legitimate concerns,” notwithstanding Principles 31 and 29.
44. Several of the content types and moderation actions excluded from the Board’s review have been at the forefront of human rights controversies. Political advertisements, events and groups are among the excluded content types. Each of these content types have been recently linked to human rights abuses.¹⁵⁴ Content subject to moderation actions other than removal has also been the subject of recent controversy. Content that Facebook has left up has caused serious harm to rightsholders; for example, incitement and hate speech that Facebook did not remove contributed to the genocidal attacks on the Rohingya community in Myanmar.¹⁵⁵ Indeed, as the mandate has detailed, “[t]he consequences of ungoverned online hate can be tragic.”¹⁵⁶
45. Another moderation action excluded from review is “down-ranking,” defined as the use of algorithms to limit the visibility of specific content.¹⁵⁷ The impact of down-ranking on freedom of expression is similar to that of removal;¹⁵⁸ thus, the omission of down-ranked content from review provides a way for Facebook to circumvent Board review while taking action akin to removal.¹⁵⁹ The exclusion of cases involving down-ranked content from the Board’s review is particularly problematic given that Facebook generally moderates misinformation using this action.¹⁶⁰

¹⁵³ *Id.*, Principle 20.

¹⁵⁴ See, e.g., Mike Isaac, “Why Everyone Is Angry at Facebook Over Its Political Ads Policy,” *The New York Times*, 22 November 2019; Sam Adler-Bell, “Facebook Is Removing Protest Pages. That’s a Terrible Precedent,” *Medium*, 24 April 2020; Kevin Roose, “Facebook’s Private Groups Offer Refuge to Fringe Figures,” *The New York Times*, 3 September 2018.

¹⁵⁵ The independent international fact-finding mission on Myanmar, established by the UN Human Rights Council, implicated Facebook in the Myanmar atrocities, stating: “The role of social media is significant. Facebook has been a useful instrument for those seeking to spread hate, in a context where, for most users, Facebook is the Internet. Although improved in recent months, the response of Facebook has been slow and ineffective.” A/HRC/39/64, para. 74. See also Evelyn Douek, “Why Were Members of Congress Asking Mark Zuckerberg About Myanmar? A Primer,” *Lawfare*, 26 April 2018.

¹⁵⁶ A/74/486, para. 41.

¹⁵⁷ Facebook’s use of algorithms to promote and amplify content is also not reviewable by the Board. See Bylaws, art. 2 § 1.2.

¹⁵⁸ See Abby Ohlheiser, “Facebook wants to limit the reach of bogus medical ‘cures’ by treating them like spam,” *The Washington Post*, 2 July 2019.

¹⁵⁹ Douek, *supra* note 143.

¹⁶⁰ See Facebook Community Guidelines, § 21 (“For these reasons, we don’t remove false news from Facebook but instead, significantly reduce its distribution by showing it lower in the News Feed.”).

46. “Geo-blocking” is a third moderation action excluded from review. This action typically involves restricting the visibility of content to users located in certain geographical areas, based on State assertions that the content violates local laws in those areas.¹⁶¹ According to Facebook, content subject to such restrictions is not reviewable because the company cannot grant to the Board greater power than the company itself has.¹⁶² However, as the 2018 Report explains, the requirements of local law and other State pressures to restrict content may compel company action that is inconsistent with human rights standards.¹⁶³ The Board’s oversight regarding the company’s approach to State restrictions could help ensure that Facebook is minimizing any adverse human rights impacts as much as possible.¹⁶⁴ As noted in the 2018 Report, “[a]rrangements to coordinate content actions with State input exacerbate concerns that companies perform public functions without the oversight of courts and other accountability mechanisms.”¹⁶⁵
47. The bylaws provide that, in some indeterminate future, users will be able to refer cases involving a broader array of content types and moderation actions.¹⁶⁶ The Board’s scope should be expanded as soon as possible and, at a minimum, should include the content types and moderation actions described above. Several Board members have made statements indicating that they intend to prioritize the expansion of the Board’s scope; however, the Board’s purview remains limited until these statements become institutional reforms.¹⁶⁷
48. In the meantime, Facebook may refer matters to the Board that involve many of these content types and moderation actions, although only matters involving individual cases are

¹⁶¹ See Kate Klonick, “The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression,” 129 *Yale Law Journal* 2418, 2430 n.32 (2020).

¹⁶² Laura Hecht-Felella and Faiza Patel, “Facebook Bylaws for Takedown Oversight Board: Questions of Independence,” *Just Security*, 5 March 2020.

¹⁶³ “The commitment to legal compliance can be complicated when relevant State law is vague, subject to varying interpretations or inconsistent with human rights law. For instance, laws against ‘extremism’ which leave the key term undefined provide discretion to government authorities to pressure companies to remove content on questionable grounds. Similarly, companies are often under pressure to comply with State laws that criminalize content that is said to be, for instance, blasphemous, critical of the State, defamatory of public officials or false.” A/HRC/38/35, para. 23. Prevention and mitigation strategies range from adopting interpretations of local laws that “resolve any legal ambiguity in favour of respect for freedom of expression [and other human rights],” challenging overbroad requests in court, to creative approaches to transparency that “disclose all relevant and publishable information” concerning government requests. A/HRC/35/22, para. 71.

¹⁶⁴ Pursuant to Principles 19 and 23, if Facebook is unable to prevent adverse human rights impacts in connection with State requests, it has a responsibility to minimize these impacts to the “greatest extent possible.” Guiding Principles, Principle 19, Commentary; Guiding Principles, Principle 23, Commentary. See Gabriella Casanova Carlos Lopez and Sam Zarifi, “Some questions regarding Facebook’s oversight board and remediation of human rights impacts (Part II),” *Opinio Juris*, 3 March 2020.

¹⁶⁵ A/HRC/38/35, para. 20. “The Board is also not allowed to review certain decisions of staff to remove content from the platform when this may cause unfavourable government action against Facebook. Setting aside the vagueness and imprecision of language such as ‘unfavourable action’ . . . Many human rights defenders in the field in countries with poor human rights records may wonder whether Facebook is a reliable ally for human rights in the world or will it blink at the first sign of government bullying.” Gabriella Casanova Carlos Lopez and Sam Zarifi, “Some questions regarding Facebook’s oversight board and remediation of human rights impacts (Part II),” *Opinio Juris*, 3 March 2020.

¹⁶⁶ Bylaws, art. 3 § 1.1.2.

¹⁶⁷ See, e.g., Botero-Marino, et al., *supra* note 118.

binding on the company. This channel provides an opportunity for Facebook to seek guidance on a range of timely human rights issues.¹⁶⁸ For example, Facebook will have (and should seize) the opportunity to seek oversight regarding its use of artificial intelligence and related technologies in moderating content, which, as the mandate has explained, can “invisibly supplant, manipulate or interfere with the ability of individuals to form and hold their opinions or access and express ideas in the information environment.”¹⁶⁹

49. In effect, the matters referred by Facebook could largely determine the true breadth of the Board’s scope. However, reliance on Facebook to seek oversight renders the Board “less ‘Supreme Court’¹⁷⁰ and more ‘optional consultant’” regarding critical human rights issues.¹⁷¹ The human rights of billions of rightsholders may be dependent upon what questions the company sees fit to ask. Thus, Facebook should engage with civil society organizations and other stakeholders in deciding what matters, particularly requests for policy advisory opinions, to refer to the Board. The importance of this issue is compounded by the Board’s limited capacity, such that it is expected to decide only a small fraction of the cases referred to it. Dependency on Facebook may be mitigated if the Board gains the authority to take up matters on its own accord, a possibility that at least one member has suggested.¹⁷²
50. The Board’s case selection criteria will also be critical in determining the matters decided by the Board.¹⁷³ These criteria should be grounded in human rights standards. The charter provides that “[i]n its selection, the board will seek to consider cases that have the greatest potential to guide future decisions and policies.”¹⁷⁴ However, pursuant to Principle 24, the Board should formulate criteria prioritizing the severity and irremediability of human rights impacts. Severity, in this context, “can be judged by characteristics such as scope (i.e., the number of people affected by the harm) [and] scale (i.e., the seriousness of the harm for the victim),”¹⁷⁵ while irremediability looks at whether delaying a response will lead to

¹⁶⁸ See A/HRC/38/35, paras. 12–40 (identifying salient human rights issues for content moderation).

¹⁶⁹ A/73/348, para. 58. Furthermore, a recent report by the UN Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance has detailed the myriad ways in which algorithms recapitulate and exacerbate discrimination. A/HRC/44/57. “As ‘classification technologies that differentiate, rank, and categorize’, artificial intelligence systems are at their core ‘systems of discrimination’. Machine-learning algorithms reproduce bias embedded in large-scale data sets capable of mimicking and reproducing implicit biases of humans, even in the absence of explicit algorithmic rules that stereotype.” *Id.*, para. 7.

¹⁷⁰ The Oversight Board was initially styled as a “Supreme Court.” Ezra Klein, “Mark Zuckerberg on Facebook’s hardest year, and what comes next,” *Vox*, 2 April 2018.

¹⁷¹ Douek, *supra* note 143.

¹⁷² See “Facebook Oversight Board may take up cases by itself: Sudhir Krishnaswamy,” *The New Indian Express*, 10 May 2020. Board member Sudhir Krishnaswamy was quoted as stating: “There are three possible channels through which cases can come to the Oversight Board. The Facebook content moderation team may itself refer a case to the board or a user may bring up a case and potentially even the Oversight Board itself might feel that a particular case should come to it . . . The Facebook referral is already recognised and so is the user referral. The third channel is not yet open. So at any event, the board will write up a document making very clear how cases will come up.” *Id.*

¹⁷³ See Bylaws, art. 1, § 1.2.1.

¹⁷⁴ Charter, art. 2 § 1.

¹⁷⁵ BSR, *supra* note 10, at 15 (interpreting the Guiding Principles). These criteria may somewhat overlap with the “significant and difficult” criteria for Facebook referrals, but they differ in important respects. *Id.*

irreparable harm. Implementing such criteria will require the Board to record and track the potential human rights impacts of accepted cases.¹⁷⁶

51. In addition to expanding the subject matter scope of the Board, Facebook should expand the universe of remedies available to rightsholders. Simply reinstating content that has been removed by Facebook may not be sufficient to remedy an abuse, especially after more than ninety (90) days. For example, as explained in the 2018 Report, content removals during a public protest could have significant impact on political rights, and reinstatement may not be an adequate remedy.¹⁷⁷ By the time such content could be reinstated, the protest may have ended and the reinstated content would likely be buried in a user feed. Also, content removals may result in a range of harms that reinstatement may not address, including reputational, physical, moral and financial injuries.
52. Accordingly, the Board should have the power to order remedial measures beyond reinstatement. International law sets forth five basic categories of remedial options: restitution, compensation, rehabilitation, satisfaction and guarantees of non-repetition.¹⁷⁸ The Board should be authorized to order specific measures within these categories, including, for example, monetary damages, revisions to content moderator protocols, psychological counseling, formal apologies and alterations of product features.¹⁷⁹ The Board should also have the power to order interim measures, in appropriate circumstances, to preserve the rights of those affected by Facebook's actions during the pendency of the Board's deliberations.
53. Moreover, pursuant to Principle 20, the Board should have the authority to retain oversight of the implementation of its decisions. Although the Board is to provide information regarding Facebook's adherence to its decisions in its annual reports, it is not expressly authorized to monitor and investigate Facebook's implementation. The Board should be authorized to do so and to engage with civil society organizations and other stakeholders in its monitoring and investigatory efforts.¹⁸⁰

Increasing transparency and stakeholder engagement

54. In shaping the Board, Facebook has demonstrated a commitment to transparency and stakeholder engagement; however, the Guiding Principles indicate significant room for improvement. Principle 31 places particular emphasis on transparency, which is listed as one of the effectiveness criteria and calls for "keeping parties to a grievance informed about its progress, and providing sufficient information about the mechanism's performance to build confidence in its effectiveness and meet any public interest at stake."¹⁸¹ Transparency is integral to several other effectiveness criteria, which establish that oversight mechanisms should be "[a]ccessible: being known to all stakeholder groups for whose use they are

¹⁷⁶ BSR, *supra* note 10, at 18.

¹⁷⁷ See A/HRC/38/35, para. 38.

¹⁷⁸ A/RES/60/147, paras. 18-23.

¹⁷⁹ See BSR, *supra* note 10, at 31-32, 56-57.

¹⁸⁰ Guiding Principles, Principle 20(b).

¹⁸¹ *Id.*, Principle 31(e).

intended”; “[p]redictable: providing a clear and known procedure with an indicative time frame for each stage, and clarity on the types of process and outcome available and means of monitoring implementation”; and “[e]quitable: seeking to ensure that aggrieved parties have reasonable access to sources of information.”¹⁸² With respect to stakeholder engagement, Principle 31 calls for oversight mechanisms to be “[a] source of continuous learning: drawing on relevant measures to identify lessons for improving the mechanism and preventing future grievances and harms” and “[b]ased on engagement and dialogue: consulting the stakeholder groups for whose use they are intended on their design and performance, and focusing on dialogue as the means to address and resolve grievances.”¹⁸³

55. Facebook has included numerous Board features in service of transparency, but opacities remain. Principle 31 makes clear that transparency should underpin all of the Board’s activities, including its user communications and decision-making. Accordingly, the Board should provide users with regular updates on the status of their cases, not only upon selection for review and finalization of the decision.¹⁸⁴ In the context of decisional transparency, the Board should post on its public website all of the information and analyses received by the Board, including from Facebook and third parties, during the course of its deliberations.¹⁸⁵ Also, the Board should be required to provide public explanations of its decisions to accept a case or deny review.¹⁸⁶ These transparency measures will help “build confidence in [the Board’s] effectiveness and meet any public interest at stake.”¹⁸⁷
56. Further, pursuant to Principle 31, the Board should undertake an educational campaign to publicize the Board and its features to communities across the globe. This campaign should provide detailed information regarding the procedure for users to refer cases to the Board. In particular, the Board should ensure that vulnerable and marginalized groups of rightsholders, including those without access to Facebook, are informed of the Board’s functionalities and activities.¹⁸⁸ Such efforts will help establish the Board as an accessible and predictable mechanism.¹⁸⁹
57. Improved transparency will enable meaningful engagement with a range of stakeholders; however, the Board also should include institutional channels for stakeholder involvement.

¹⁸² *Id.*, Principle 31(b-d). Principle 21 also emphasizes transparency, providing that “[i]n order to account for how they address their human rights impacts,” companies should “communicate this externally, particularly when concerns are raised by or on behalf of affected stakeholders.” *Id.*, Principle 21.

¹⁸³ *Id.*, Principle 31(g-h). Principle 18 also stresses the importance of stakeholder engagement, stating that assessing human rights impacts should “[i]nvolve meaningful consultation with potentially affected groups and other relevant stakeholders.” *Id.*, Principle 18.

¹⁸⁴ *See* Bylaws, art. 3 § 1.2.

¹⁸⁵ Also, the Board should provide users whose cases are selected for review with an opportunity to submit supplementary statements responding to information submitted by Facebook or third parties to the Board. Further, independent support (for example, user advocates) should be provided to help guide users through the review process and prepare supplementary statements. *See* BSR, *supra* note 10, at 54.

¹⁸⁶ *See* Gabriella Casanova Carlos Lopez and Sam Zarifi, “Some questions regarding Facebook’s oversight board and remediation of human rights impacts (Part I),” *Opinio Juris*, 3 March 2020; Bylaws, art. 1 § 1.2.1.

¹⁸⁷ *See* Guiding Principles, Principle 31(e).

¹⁸⁸ *See* BSR, *supra* note 10, at 53.

¹⁸⁹ *See* Guiding Principles, Principle 31(b-c).

Such channels would facilitate the Board’s and Facebook’s continuous learning and enhance the dialogic nature of the oversight process, in accordance with Principle 31.¹⁹⁰ As one option, the Board could create a channel for civil society organizations or other stakeholders, individually or collectively, to refer matters to the Board.¹⁹¹ This channel could help ensure that the most salient human rights issues are presented to the Board, instead of letting Facebook’s referrals dictate the docket. Another option is to allow for civil society organizations and other stakeholders to submit information and analysis to the Board on pending matters, particularly policy questions raised by Facebook (akin to the *amicus curiae* briefing process in many judicial settings).¹⁹² Currently, the bylaws empower the Board to request information and analysis from civil society organizations, as well as a “global pool of outside subject-matter experts” selected by the Board.¹⁹³ However, this process is discretionary: The Board can choose whether to make such requests, which organizations and experts to engage and whether to consider the input received.

58. Notably, stakeholder engagement is invaluable for understanding the context of content at issue before the Board.¹⁹⁴ Under ICCPR Article 19, restrictions on speech require a legitimate reason that is proportionate, necessary and non-discriminatory. This analysis entails gauging the geographic, temporal, linguistic and cultural context of the restricted content.¹⁹⁵ In assessing the context of restricted content, Board panels will look to the member who is from the “region which the content primarily affects.”¹⁹⁶ However, the delineated regions are extremely broad, such that it is unlikely this member will be able to offer meaningful insights.¹⁹⁷ For example, an Argentinian member could be tasked with providing context for content from a user in Trinidad and Tobago. Argentina and Trinidad and Tobago both fall within the “Latin America and the Caribbean” regional grouping but are culturally distinct States. This approach could result in inaccurate contextual understandings. To address this issue, the Board should engage with local stakeholders affected by or knowledgeable about the context of the content under review.

59. In all of its stakeholder engagement, Facebook and the Board should prioritize engaging with communities and individuals historically at risk of censorship, discrimination and other potential harms of content moderation. The Trust should allocate funds to cover

¹⁹⁰ Guiding Principles, Principle 31(g-h).

¹⁹¹ See ARTICLE 19, “Facebook oversight board: Recommendations for human rights-focused oversight,” 27 March 2019.

¹⁹² Facebook’s original draft of the charter for the Board did include as a suggested approach that “Facebook users and pertinent stakeholders may also submit arguments and material to the panel.” Draft Charter: An Oversight Board for Content Decisions, *Facebook Newsroom*, 28 January 2019, at 4. However, subsequent versions of the charter did not include this feature.

¹⁹³ Bylaws art. 1 § 3.1.4.

¹⁹⁴ Stakeholder input may also help identify “identical content with parallel context,” which Facebook is required to remove following certain Board decisions. See Bylaws, art. 2 § 2.3.

¹⁹⁵ See Communication No. OL OTH 24/2019.

¹⁹⁶ See Bylaws, art. 1 § 3.1.3.

¹⁹⁷ The regions are: (1) The United States and Canada, (2) Latin America and the Caribbean, (3) Europe, (4) Sub-Saharan Africa, (5) the Middle East and North Africa, (6) Central and South Asia and (7) Asia Pacific and Oceania. Bylaws, § Definitions.

expenses associated with stakeholder participation in the Board’s deliberations, such as travel costs and interpreter fees.¹⁹⁸

The independence and global legitimacy of the Board

60. Facebook has taken steps to ensure the Board’s independence and legitimacy, yet several concerning aspects remain. Principle 31 makes clear that oversight mechanisms should be “[l]egitimate: enabling trust from the stakeholder groups for whose use they are intended, and being accountable for the fair conduct of grievance processes.”¹⁹⁹ In enabling stakeholder trust, “[a]ccountability for ensuring that the parties to a grievance process cannot interfere with its fair conduct is typically one important factor.”²⁰⁰ Indeed, the independence and legitimacy of an oversight mechanism are intertwined. Moreover, the Basic Principles on the Independence of the Judiciary (“Basic Principles”), endorsed by the UN General Assembly, may provide some guidance given the company’s stated commitment to ensuring the independence of the Oversight Board. These Basic Principles stipulate that “[a]ny method of judicial selection shall safeguard against judicial appointments for improper motives” and removal of judges should be based only on “reasons of incapacity or behavior that renders them unfit to discharge their duties,”²⁰¹ which has been interpreted to mean “serious grounds of misconduct or incompetence.”²⁰²
61. Over the years, multiple UN Special Rapporteurs have recognized that the most effective method for securing judicial (or quasi-judicial) independence during the appointments process is through a “well-entrenched independent mechanism . . . responsible for the appointment, promotion, transfer and dismissal of judges.”²⁰³ Also, the previous mandate holder noted that “any restriction [on freedom of expression] imposed must be applied by a body that is independent of political, commercial or other unwarranted influences.”²⁰⁴
62. Contrary to this guidance, Board member selection and removal are not conducted by an independent mechanism.²⁰⁵ In selecting members, Facebook and co-chairs selected by Facebook have been tasked with recruiting the first set of potential Board candidates; going forward, the Board will assume responsibility for recruitment.²⁰⁶ Once recruited, each candidate must be formally approved by the Trust in order to become members.²⁰⁷ The Trust does provide a critical layer of separation between Facebook and the Board. However, Facebook may retain a significant amount of influence over the Board due to its selection of the co-chairs and initial slate of members, as well as the Trustees.²⁰⁸ Meanwhile, removal of a Board member requires approval by the Board and the Trustees,

¹⁹⁸ See Communication No. OL OTH 24/2019.

¹⁹⁹ Guiding Principles, Principle 31(a).

²⁰⁰ *Id.*, Principle 31, Commentary.

²⁰¹ Basic Principles on the Independence of the Judiciary, G.A. Res. 40/32 and 40/146.

²⁰² A/HRC/11/41, para. 59.

²⁰³ E/CN.4/1995/39 at 25. See also A/HRC/11/41, para. 61; A/HRC/35/31, para. 10.

²⁰⁴ A/67/357, para. 42.

²⁰⁵ *Id.*

²⁰⁶ Bylaws, art. 1 § 1.2.2.

²⁰⁷ *Id.*, art. 4 § 2.1.

²⁰⁸ *Id.*, art. 1 § 1.1.2; Charter, art. 5 § 2.

although requests for removal may come from the Board, the public or the Trustees.²⁰⁹ Given Facebook’s role in selecting members and the Trustees, the removal process may also be susceptible to company influence.

63. If Facebook is serious about its emphasis on the independence of the Board, it should consider further distancing itself from member selection and removal, such as by delegating that role to an independent body with no ties to the company. An independent body, constituted with a range of civil society organizations and other stakeholders, could be empowered to select and remove Board members. Facebook users and the general public could have opportunities to provide input into this body’s selection of members.
64. Moreover, member removals should be based on established human rights standards. The governing documents provide that Board members “will not be removed due to content decisions they have made,” but only “if the trustees have determined that that member has violated the code of conduct,” which has been prepared by Facebook.²¹⁰ Many code of conduct provisions are vague and ambiguous, allowing for broad discretion and subjective biases. Pursuant to the Basic Principles, the governing documents should specify that only “serious grounds of misconduct or incompetence” may constitute adequate grounds for removal.²¹¹
65. Additionally, Facebook’s ability to amend the bylaws may undermine the Board’s independence. For example, Facebook can unilaterally amend the types of content the Board can review and case submission procedures.²¹² While it is true that amendments are not allowed to contravene the Board’s charter,²¹³ it is unclear who will decide whether or not an amendment is in accordance with the charter and observers of the Board speculate that it will be Facebook, itself.²¹⁴ Amendment powers give Facebook the continuing ability to impact and interfere with the Board’s practical functionality at any time.²¹⁵ Facebook’s powers to amend the Board’s bylaws should be sharply curtailed.
66. The Board should have broad authority to conduct an autonomous review, which is integral to independence. The bylaws provide that the Board can request information from Facebook, but they do not require Facebook to comply with such requests.²¹⁶ Independent review of content decisions will require access to information concerning the company’s policies, processes, deliberations and actions leading to those decisions.²¹⁷ To assess the

²⁰⁹ Bylaws, art. 1 § 1.2.2, art. 4 § 2.1.2.

²¹⁰ *Id.*, art. 4 § 2.1.2.

²¹¹ A/HRC/11/41, para. 59; *see also* UN Human Rights Committee, General Comment No. 32 (2007) on Article 14: Right to equality before courts and tribunals and to a fair trial, para. 20.

²¹² Bylaws, art. 5 § 1.

²¹³ *Id.*

²¹⁴ *See* Evelyn Douek, “Facebook’s Oversight Board Bylaws: For Once, Moving Slowly,” *LawFare*, 28 January 2020.

²¹⁵ Laura Hecht-Felella and Faiza Patel, “Facebook Bylaws for Takedown Oversight Board: Questions of Independence,” *Just Security*, 5 March 2020.

²¹⁶ Bylaws, art. 1 § 3.1.3, art. 2 § 2.2.2.

²¹⁷ For additional analysis regarding the Board’s independence, based on a three-part framework of (1) jurisdictional independence, (2) intellectual independence and (3) financial independence, please refer to Kate Klonick, “The

proportionality of the removal of certain content, for example, the Board may require information pertaining to previous content actions taken against the user who posted the content, the availability and feasibility of other moderation actions and factors that may amplify the content at issue.²¹⁸

67. Finally, in light of Principle 31, the Board’s legitimacy will largely be determined by public trust. Such trust will depend on how well the Board represents Facebook users and other rightsholders. As of the time of writing, five members and two of the four co-chairs are from the United States, even though fewer than ten percent of Facebook users are from the U.S.²¹⁹ Meanwhile, only one member is from the State with the largest Facebook userbase, India.²²⁰ This imbalance should be addressed in selecting the remaining Board members. Public trust will also depend on how Facebook and the Board proceed with respect to the foregoing issues (namely, application of human rights standards, remedial scope, transparency, stakeholder engagement and independence).²²¹

IV. Conclusion

68. The environment for the regulation of online content is presently dynamic, with governments, companies and civil society struggling to identify the models that best implement their interests. It may even be that regulation and oversight of content moderation itself is a limited approach and that broader approaches taking on company business models and competition policy would have a deeper impact on content oversight. For now, however, social media companies have a major impact on a range of rights that all people enjoy. As such, they should be conducting the kind of human rights due diligence, oversight and remediation discussed in this report and in other reports of the mandate. Some of the ideas identified in this report may reinforce the promotion and protection of human rights, while some may feed into the global conversation about the appropriate contours of government regulation. The mandate of the Special Rapporteur and the Clinic will continue to monitor and report on developments in this rapidly evolving landscape and their impact on human rights.

Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression,” 129 *Yale Law Journal* 2418, 2481-87 (2020).

²¹⁸ See Communication No. OL OTH 24/2019.

²¹⁹ Steven Levy, “Facebook Names the 20 People Who Can Overrule Mark Zuckerberg,” *Wired*, 6 May 2020.

²²⁰ Leading countries based on Facebook audience size as of April 2020, *Statista*, April 2020 <https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/>.

²²¹ “[T]he more positive the public’s attitude towards an institution’s right to govern, the greater its popular legitimacy. However, this legitimacy is fragile.” Centro de Estudios en Libertad de Expresión y Acceso a la Información, “Considering Facebook Oversight Board: turning on expectations,” May 2019, at 4.