University of California, Irvine
School of Law

Summary of Multi-Stakeholder Workshop on Hate Speech and the Challenge of Local Engagement with Global Online Platforms

by UCI Law International Justice Clinic

Table of **Contents**

I. Introduction

1. The United Nations Special Rapporteur on freedom of opinion and expression co-organized and participated in a workshop held on November 27-28, 2018. The workshop was held to discuss how content moderation policies and processes should integrate local engagement and input in addressing hate speech and threats of violence online. The workshop was held in Berlin, Germany with support from the German Federal Ministry for Economic Cooperation and Development (BMZ).

2. The workshop was organized in three parts: (1) a discussion of social media and hate speech in Germany and the NetzDG law; (2) a discussion on the role of social media, particularly Facebook, in Myanmar and the crisis of the Rohingya; and (3) small group break-out discussions and brainstorming of best practices and on issues identified in the case studies. It concluded with a roundtable of recommendations and take-aways.

3. The workshop was conducted under the Chatham House Rule.

4. Over thirty participants (in addition to the Special Rapporteur, his legal advisor, and student-rapporteurs from the University of California, Irvine) attended the workshop in Berlin. The participants included representatives from the private sector, civil society organizations, academia, and international organizations.

5. This report reflects points raised during the consultations but does not necessarily reflect the views of the Special Rapporteur or all participants. The Special Rapporteur intends to integrate these discussions into his upcoming report to the UN General Assembly in October 2019.

6. This report was compiled by Mackenzie Anderson and Carey Hughes of the UCI School of Law International Justice Clinic and Amos Toh, Legal Advisor to the Special Rapporteur.

II. Session 1: Germany Case Study

7. The day began with a discussion of the German law mandating Netzwerkdurchsetzungsgesgestz ("NetzDG"). The primary stated goal of NetzDG is to strengthen implementation of Germany's hate speech laws by enacting measures to ensure that social media companies monitor it more effectively.

Underlying Problems NetzDG Looked to Solve

8. After Chancellor Merkel announced that Germany would relax its immigration rules to permit a large number of refugees to apply for refugee status, there seemed to be an increase in hostility and violence towards refugees in Germany. There was a perception that the increase of hate speech online is partially to blame for the increase of attacks against immigrant and refugee populations in Germany.

9. More generally, there has been a proliferation of various forms of hate speech on social media. While there is no universal or agreed legal definition for what constitutes "hate speech", it is common for individuals to report coming into contact with (or otherwise "flagging") hate speech on social media in some form.

10. Germany has particularly robust hate speech laws which prohibit speech that "assaults the human dignity" of various protected classes including "a national, racial, religious group or a group defined by their ethnic origins." Although German law also punishes those who "disseminates" or "otherwise makes . . . accessible" such illegal hate speech, it would appear that this provision, in and of itself, is not understood to cover social media sites.

11. In Germany, the most popular mobile app is WhatsApp. YouTube has a bigger following that traditional television. While Twitter has less penetration than Facebook and YouTube, it is popular among politicians, journalists, and activists.

12. The role of social media companies in public discourse creates a regulatory quandary. In Germany as well as many other countries, the distinction between private platforms and the public square is blurred. The sheer reach of certain platforms also amplifies hate speech that would otherwise fail to receive media or popular attention.

Implementation Issues

13. Before the creation of NetzDG, a task force was created which included civil society groups focused on child safety online, government officials, and social media companies. This group came up with many of the goals later put forward in NetzDG but on a voluntary basis by the social media companies. The degree of adherence was mixed, though it appears government officials believed that it was not achieving the goals of reducing online hate. Some participants believed this approach may have been abandoned too soon.

14. The execution of NetzDG requirements has been mixed across various platforms. Participants noted various barriers to reporting under NetzDG including: forms being available in a limited number of languages despite many languages being spoken in Germany; reporting mechanisms being difficult to locate on the platforms' respective websites; criteria for reporting being unclear to users; and questions as to how locations and nationality affect the ability to report.

15. Many participants believed that the language within the law referring to "obviously illegal content" is vague and confusing for social media companies and outside civil society monitors alike. There is great concern that this may lead to overblocking of legitimate content (i.e., content not subject to other restrictions under domestic or international law).

16. At the time of the workshop, the government had not imposed fines or sought enforcement action for violations of NetzDG.

Information Gaps

17. There is a pressing need for more research into the effects of hate speech. Although counter-speech has been proposed as a way to mitigate hate speech, there is a dearth of research on its impacts.

18. There is also a continuing lack of transparency from social media companies on its efforts to address hate speech online, such as how algorithms flag possible hate speech for human review.

Rule of Law

19. The role of enforcing a country's laws is the responsibility of the country and its law enforcement agencies. NetzDG puts the onus on the social media companies to be the monitors of such hate speech instead of the government. Many participants expressed the view that attempts to address hate speech, or other forms of "difficult" speech, should not result in putting the ultimate adjudicatory power in the hands of a private company.

20. A few participants, however, argued that NetzDG simply places the burden of compliance with existing hate speech laws on companies. It would require companies to comply with these laws in the same way that individuals are required to comply.

21. The civil legal system, it was argued, has been a more successful avenue for individuals who seek redress for harms arising from online hate speech. Law enforcement agencies are generally not resourced to a level where criminal cases can be meaningfully pursued against instances of online hate speech.

22. The actions taken in Germany regarding NetzDG and by social media companies generally have intense global effects. Other countries with repressive governments have seized on NetzDG to draft or implement similar laws to NetzDG. Given that social media platforms are critical (and sometimes the only) avenues for free expression and journalism, this trend is concerning.

Community-Based Solutions

23. One problem with social media is the lack of social media or digital literacy.

24. Much of the solutions seem to be failing to address the ultimate harm of hate speech and the underlying issues. There is a lack of focus on the experiences of victims of hate speech, and their views on recourse and redress. There is also a lack of research on the factors that drive engagement with and the spread of hate speech. For example, anecdotal research suggests that individuals who feel left out of society take to social media to express their frustrations. However, their motivations and the circumstances of their marginalization are unclear.

25. There was a general recognition of the need to broaden the dialogue moving forward. This should include more relevant stakeholders previous left out of the discussions, such as non-governmental organizations that monitor and respond to hateful speech and conduct and researchers on hate speech.

26. Expanding the role of users in content moderation should be explored. There are limits, however, based on the size of the platforms. Additionally, community-based moderation may lead to abuse or neglect of minority users, since the majority of the community would exercise significant power over moderation decisions.

***

27. During session 2, the participants reviewed Facebook's role in the incitement of ethnic cleansing against the Rohingya Muslims.

History of Oppression Against the Rohingya

28. There have been numerous ethnically-based armed conflicts in Myanmar since the 1940s. These conflicts have persisted since the transition from military rule to the government established under the 2008 constitution.

29. Historically, the Rohingya have experienced profound oppression at the hands of the government. In the 1970s and 1990s, many Rohingya were expelled from Myanmar. Government policy today restricts the Rohingya from freely traveling throughout the country. The government also denies the Rohingya citizenship, marriage authorization and birth registration.

30. In 2012, the oppression of the Rohingya escalated and turned violent on a large scale. Twelve Rohingya townships were attacked, killing and injuring hundreds of Rohingya, and leaving over 8,000 homes destroyed. Similar attacks against the Rohingya, some perpetrated by the Burmese military, occurred over the next several years.

31. These attacks were not spontaneous. Rather, these attacks were instigated by radical Buddhist monk organizations, government officials, and nationally influential leaders. Ultimately, over 140,000 people, mostly Rohingya, were displaced.

32. In 2017, the United Nations estimated that over 700,000 Rohingya refugees have fled Myanmar due to extreme violence and discrimination. Information collected by the United Nations suggests that the estimate of up to 10,000 deaths is conservative. Mass killings were perpetrated in villages and in some attacks hundreds of people were raped, beaten, and killed.

Facebook Usage and Popularity in Myanmar

33. In 2013, as part of a series of reforms, the government began to liberalize the country's telecommunications sector, reducing the costs of mobile phones and increasing mobile penetration. Pre-liberalization, mobile SIM cards were a luxury, and could cost up to $7000 USD per card. Today, the average cost of a SIM card is $1.50.

34. In 2013, Facebook launched an initiative called "Free Basics", in collaboration with various telecommunications providers. The Free Basics program enabled users to download Facebook's mobile application and provided them with access to websites and services selected by Facebook that was not charged against the user's mobile plan. In addition to providing access to Facebook, Free Basics provided users with information like the weather and local news.

35. Free Basics introduced millions to Facebook, and the Internet itself. Although Facebook no longer provides Free Basics in Myanmar, Facebook has become the de facto Internet. By 2016, ten million people in Myanmar used Facebook; by 2018, that number had jumped to an estimated 20 million out of a population of about 53 million. For many in Myanmar today, Facebook is the central (if not exclusive) medium for accessing information on the Internet. Mobile phones are pre-installed with Facebook, and shopkeepers assist users in setting up their Facebook accounts.

36. Facebook's prominence in the country has made it a useful instrument for those seeking to spread hate, utilized by military personnel and civilians to incite violence against the Rohingya.

Facebook's Response to the Crisis: Issues and Concerns

37. Until recently, Facebook had mostly been relying on civil society organizations and other third parties to report hate speech for review and removal. There were also gaps in taking down Burmese content that violated Facebook's community guidelines because its software was unable to process Burmese characters.

38. Criticism of Facebook's inaction has led the company to hire increasing numbers of local language content reviewers. At the time of the workshop, Facebook had hired 100 Burmese literate content reviewers.

39. However, there is no indication that these reviewers are familiar with the ethnic tensions in the country, or that these reviewers will be capable of keeping track of how the expression of hate speech and incitement to violence evolves on the platform (in part to evade reviewers). Indeed, these content reviewers are not working from Myanmar, but remotely in another country. Concerns were also raised that some of these reviewers were also biased against the Rohingya.

40. Facebook has removed accounts of individuals and groups that generate hate speech, including radical groups and military leaders. These removals come after repeated failed attempts by civil society organizations and government officials to raise concern about the spread of hate speech and incitement to violence on the platform.

41. Since Facebook does not have local presence, civil society organizations that reported anti-Rohingya content on Facebook early on in the crisis experienced difficulties contacting anyone who had the power to make expedient decisions to remove hateful content.

42. For a long time, Facebook did not have Burmese-language reporting instructions or moderators, making it difficult to report content that violates its rules. Facebook has begun to create Myanmar-specific products so that users would better be able to flag inappropriate content. However, it is unclear whether Facebook is still capable of managing its 20 million Burmese users, many of whom come from different ethnic backgrounds and speak languages other than Burmese. Many users in Myanmar open a Facebook account and select Burmese as their language but participate on the platform in another language.

43. Many Burmese users, according to some participants, remain unaware of Facebook's community standards and that a reporting system exists to report violence-inciting

content. Many users are also reticent about reporting hateful content, as they are unclear whether their complaints would be addressed. For some, it is not in their culture to report or lodge official complaints.

<center>***</center>

IV. Session 3: Breakout Sessions and Plenary Discussion

44. On the second day of the workshop, participants broke up into small groups, and each group was given a different discussion question. These questions were: (1) How should platforms expand, diversify, or refine their policymaking processes and workforce to integrate local input and engagement, particularly from historically marginalized populations? (2) Can an effective rights-oriented approach be achieved through changes in content policy and practice or is broader structural change necessary? (3) To what extent are the ongoing challenges with regulating hate speech attributable to the dominance and size of a particular platform, and how should this relationship be addressed? (4) What is the role of external accountability and oversight mechanisms, and how should these be established?

45. The paragraphs below capture major themes that were discussed during the individual brainstorming sessions and the concluding plenary discussion.

Transparency

46. Greater transparency is required to determine how companies should develop and implement content-related standards, and whether broader structural change is required. Platforms should, for example, provide greater insight into how they deploy algorithms and other AI technologies in content moderation and curation.

47. Platforms should explain their decision making and clarify when and how local communities' input will be taken into account. This is particularly important during situations of crises. Situations can change rapidly, and commitment to communication and transparency in real time is key.

48. Transparency alone, however, is not enough. Transparency must be viewed as part of a range of responses that include effective local and public engagement and an openness to rethinking and reconceptualizing problems as they arise. Additionally, companies have some responsibility to contextualize information they make available to the public and actively solicit feedback on transparency-related processes.

Internal Practices

49. Platforms should consider how human rights norms apply at every level of the company, including for developers and engineers, and explain how international

human rights standards are considered and implemented throughout their operations (for example, how updates to community standards are guided by human rights law). States should advocate for companies to create such changes, and the process should include relevant stake-holders.

50. As a starting point, platforms should conduct human rights impact assessments to better understand landscape of the countries and regions they are in or seeking to enter. From there they must develop internal processes which regularize local engagement.

51. Diversity and representation are crucial to successful policy development. This includes the hiring of employees from a wide variety of cultural and linguistic backgrounds. Meaningful diversity should be implemented at all levels of the company, and not just in leadership. Employees from the Global South should be encouraged to speak *in* and *for* the company.

Engagement with External Stakeholders

52. Platforms should actively engage in public outreach and regularly consult local civil society groups and advocates. To build relationships with civil society, companies must be clear about the advice they are seeking, how external input guides their decision-making and next steps. Companies should also follow up on decisions that were made as a result of the input given. To avoid extractive relationships, companies should offer to cover travel and related expenses or provide the option of monetary compensation for time spent.

53. Platform should be careful not to have "gatekeepers." Instead, they should be sure to engaged with a variety of groups and advocates and continue to switch out and diversify their pool of experts of stakeholders.

54. Companies should make a concerted effort to evaluate obstacles in identifying and engaging with marginalized groups, including those not represented within mainstream civil society. An individual is not necessarily an effective representative just because they come from a particular country or has a marginalized background.

External Accountability

55. A global external body may be an effective way to assist platforms in integrating human rights standards into their operations. Such a body, could create pressure on governments to refrain from or cease human rights abuses where relevant local laws are inconsistent with international human rights standards.

56. In regions and countries where Facebook is effectively the Internet, the stakes for participating or not are much higher. If the user leaves the forum or are suspended, they effectively leave public discourse. There should be appeals mechanisms, and ideally these should be designed with input from relevant civil society groups. These mechanisms should be widely accessible to the everyday user.

57. Platforms currently rely heavily on flagging by users to identify violation of community standards or laws. The frequency and types of complaints flagged depend on users' preferences, biases and values. Accordingly, this may lead to inconsistent

enforcement of the platform's rules, in ways that further marginalize minority groups and at-risk communities.

Concluding Observations

58. Participants generally agreed that international human rights considerations must be integrated into platform policies and processes from the outset, but differed on how these considerations should be taken into account.

59. Human rights law not only provides a source of legitimacy for companies, but also meets a variety of business needs. It helps them conduct an assessment of risks and benefits associated with product design, market entry and policy development.

60. However, it is still unclear how, when and to what extent major internet platforms integrate human rights considerations into their decision-making. It is unclear how and when the company's economic interests would conflict with the implementation of human rights norms. At the same time, companies risk capture and backlash when they rely on community standards developed in Silicon Valley, which inevitably reflect Western cultural values and biases.

61. Solutions to emerging human rights challenges online are greatly influenced by the size of the organization, and much of the discussion has focused on how to solve the problems faced by the larger platforms. Smaller companies will struggle to implement solutions that are focused only on addressing the challenges posed by tech giants.

62. While the discussion focused largely on the platforms, the state's role should not be minimized. States have obligations to ensure that companies respect human rights. Conversely, companies can and should rely on human rights law to resist pressure to facilitate state-based abuses.