13 April 2022

Dear Twitter, Meta, Google, Telegram, and Tik Tok

We stand in solidarity with the Ukrainian civil society organizations that have been working with social media platforms and have been at the forefront of the fight to ensure that all Ukrainian people have access to the free flow of life-saving information. We empathize with the difficulties that they have experienced in getting support from these companies.[1] Many of us have experienced similar difficulties, even in the midst of humanitarian crises.

The February 24, 2022 escalation of the Russian invasion of Ukraine, which started eight years ago, has been a trying time for the Ukrainian civil society organizations that continue to fight for the fundamental rights of their people while trying to ensure their own safety at the same time.[2]

The prompt response from online platforms lays bare the flawed standard operating procedure of how tech companies function globally.[3] The dominant social media platforms knew well before Russia's most recent act of aggression that Russian information warfare has been particularly dangerous in the context of this international armed conflict and brought about serious human rights implications. For years, their services have been used to spread mis- and dis-information and incitement to violence, while they have deleted documentation of human rights abuses since 2014, and suspended accounts of activists.[4] Ukrainian civil society informed them about myriad issues, such as targeted mass reporting of content critical to Russians and Putin, mostly

[1] Tanya Lokot, *Facebook's Zuckerberg Responds to Ukrainians' Complaints, But Is His Answer Enough?*, 15 May 2015, Global Voices, https://globalvoices.org/2015/05/15/facebooks-zuckerberg-responds-to-ukrainians-complaints-but-is-his-answer-enough/

[2] *Civil society calls for solidarity with Ukraine's human rights defenders in guarding against cyber threats,* 18 Feb. 2022, updated 24 Feb 2022, https://www.accessnow.org/civil-society-calls-for-solidarity-with-ukraines-human-rights-defenders-in-guarding-against-cyber-threats/ ("For the last few months we observed significant escalation in digital attacks — malware, phishing, and account takeover attempts against independent media and civic society organizations in Ukraine," said Vita Volodovska, Head of Digital Security Lab Ukraine.)

[3] Will Oremus, *Social media wasn't ready for this war. It needs a plan for the next one,* 25 Mar. 2022, Washington Post, https://www.washingtonpost.com/technology/2022/03/25/social-media-ukraine-rules-war-policy/

[4] See, Tanya Lokot, *Facebook Weighs In on the Ukraine-Russia Takedown Dispute,* 7 Oct. 2014, Global Voices, https://globalvoices.org/2014/10/07/russia-ukraine-facebook-tos-takedown/; *Open statement to the international community and intergovernmental organisations* 29 Mar. 2022, https://cedem.org.ua/en/news/open-statement/, ("The statement says "This catastrophe became possible due to laziness, cowardice, and conformism of international institutions and organisations that have failed to effectively defend the principles of democracy, the rule of law, and human rights. It also became possible because by relying confidently on the international standards and norms developed after World War II, we did not pay attention to the fact that the all-encompassing technological "disruption" of recent years could require a substantial revision and/or adaptation of at least some of these standard" and it calls on the international community to "Recognize that freedom of expression and freedom of the media should in no way extend to aggressive information operations that are funded, controlled, and/or otherwise supported by countries with unlawful intent.")

satirical in nature, or failure to moderate content using pejorative terms for Ukrainians.[5] The February 24th re-escalation highlighted all these trends, for which platforms failed to adequately prepare.[6] For years, they lacked sufficient and meaningful long-term engagement with civil society in Ukraine and started taking real action only after the EU and USA raised serious concerns about Russian aggression.

While we recognize the efforts of tech companies to uphold democracy and human rights in Ukraine, we call for long term investment in human rights, accountability, and a transparent, equal and consistent application of policies to uphold the rights of users worldwide. Once platforms began to take action in Ukraine, they took extraordinary steps that they have been unwilling to take elsewhere. From the Syrian conflict to the genocide of the Rohingya in Myanmar, other crisis situations have not received the same amount of support even when lives are at stake.

Human Rights Watch and researchers at the Shorenstein Center at Harvard University, among others, have chronicled how social media platforms have responded to the war in Ukraine, and we won't replicate the list here.[7] Not only do these responses demonstrate a willingness to act that has been absent in many other situations, but they also reflect the flawed way in which these platforms operate. For example, some platforms engage in due diligence efforts such as human rights impact assessments or other independent reviews only when they are legally required to do so, are pushed to by civil society, or garner negative media attention. And, unfortunately, some platforms only make changes to their services to protect human rights when under the same pressures. When those pressures subside, they often turn their attention away instead of learning from their mistakes.

Platforms must extend focused resourcing and policies to other situations where appropriate and address the fundamental problems with their business models everywhere in the world. Unfortunately, examples of their failure to do so, even in the midst of violent conflicts, abound.

For example, in May 2021, in anticipation of forcible evictions in the Sheikh Jarrah neighborhood in Jerusalem, Palestinians engaged in protests that Israeli security forces violently suppressed. Social media platforms removed massive amounts of content posted by

---

[5] *See fn 1, 2, 4 and* Tanya Lokot, *Ukraine Asks Facebook's Zuckerberg to Discipline Kremlin Bots*, 29 Aug. 2014, Global Voices,https://globalvoices.org/2014/08/29/ukraine-russia-facebook-bots-censorship/; Vitaly Shevchenko, *Ukrainians petition Facebook against 'Russian trolls'*, 13 May 2015, BBC, https://www.bbc.com/news/world-europe-32720965

[6] Dr. Courtney C. Radsch, *Tech Firms Caught in the Middle of Russia's War on Ukraine*, 28 Feb. 2022, Tech Policy Press, https://techpolicy.press/tech-firms-caught-in-the-middle-of-russias-war-on-ukraine/

[7]Human Rights Watch, *Russia, Ukraine, and Social Media and Messaging Apps: Questions and Answers on Platform Accountability and Human Rights Responsibilities,* 16 Mar. 2022, https://www.hrw.org/news/2022/03/16/russia-ukraine-and-social-media-and-messaging-apps; April Glaser and Jazilah Salam, *Tracking Social Media Takedowns and Content Moderation During the 2022 Russian Invasion of Ukraine,* The Media Manipulation Casebook, the Shorenstein Center at Harvard University, *Published* 1 Mar 2022*, last updated* 25 Mar, 2022, https://mediamanipulation.org/research/tracking-social-media-takedowns-and-content-moderation-during-2022-russian-invasion

Palestinians and their supporters, who were trying to document and share these human rights violations, as well as political discussions about Palestine around the world. 7amleh documented more than 500 violations against Palestinian content on these platforms between May 6 and May 19, 2021.[8] Concurrently, it observed a 15-fold increase, compared to the same period of the previous year, in hate speech, incitement, and violence against Palestinians and Arabs on social media platforms.[9] According to 7amleh, this content was not moderated or taken down despite direct harmful consequences in real life. Human rights and digital rights organizations attempted to get immediate answers to this situation as well as long-term solutions. Unfortunately, it took weeks for platforms to address what they called technical errors.[10] At the same time, livestreams of ongoing human rights abuses were shut down and content continued to be removed. Speech was removed in this context that may have been left up had it received the kind of contextual analysis platforms are claiming to do in Ukraine now.

The Undersigned Call on Social Media Platforms to provide:
1. **Real human rights due diligence:** Platforms should engage in ongoing and meaningful human rights due diligence globally, prioritizing for immediate review their operations in those countries and regions whose inhabitants are at risk of mass killings or grave human rights violations. Meaningful due diligence requires reviews by independent entities that are not financially dependent on platforms for their revenue. To ensure a diversity of perspectives and independence, the same entity should not be performing all of a company's reviews, and whatever entity a company contracts should publicly explain their processes, what other work they have done for companies, and demonstrate how they ensure their independence. In addition, companies must give the reviewing entity access to all the information necessary to do a meaningful review, and allow the entity to conduct its review free of interference from the company. These reviews should be made publicly available to the greatest extent possible, barring reasonable safety and security risks or exposure of trade secrets.

2. **Equitable investment:** Platform investments in policy, safety, and integrity must be determined by the level of risk they pose to human rights, not just by the commercial value of a particular country or whether they are located in jurisdictions with enforceable regulatory powers. Currently, platforms are devoting greater time, attention, and resources to their users in the United States and Western Europe. This happens both because of the potential for greater regulation by the United States and the European Union and because media based in the United States plays a significant role in influencing public discourse about companies, prompting greater attention to issues of public interest for the United States. As a result of this pressure, platforms sink

---

[8] *The Attacks on Palestinian Digital Rights*, 21 May 2021, 7amleh, https://7amleh.org//storage/The%20Attacks%20on%20Palestinian%20Digital%20Rights.pdf

[9] *Index of Racism and Incitement: An increase in racism and incitement against Palestinians & Arabs during the last Israeli Aggression,* 16 Jun 2021, 7amleh, https://7amleh.org/2021/06/15/index-of-racism-and-incitement-an-increase-in-racism-and-incitement-against-palestinians-and-arabs-during-the-last-israeli-aggression

[10] Maya Gebeily, *Instagram, Twitter blame glitches for deleting Palestinian posts,* 10 May 2021, Reuters, https://www.reuters.com/article/israel-palestinians-socialmedia-idUSL8N2MU624

resources into these parts of the world, and often provide more and better protections to people in the US and Europe than people elsewhere, including those living with the gravest threats to their human and democratic rights and freedoms.

3. **Meaningful engagement:** Platforms must build meaningful relationships with civil society globally that are based not on extraction of information to improve products, but also provide civil society with meaningful opportunities to shape platform tools and policies. This means bringing in civil society at the design stage and involving civil society in due diligence efforts at early stages. Platform should ensure that staff who are in contact with civil society have decision-making power and access to engineers. It also means that there should be a two-way information flow, with companies reporting back on policy and tool developments when civil society has provided input. Finally, where civil society organizations have ongoing relationships with platforms, those platforms should make an effort to respond quickly to emergency requests, even if only to provide status updates.

4. **Linguistic equity in content moderation**: Platforms must hire adequate numbers of content moderators and staff for every language in which they provide services. They must fully translate all of their policies into all the languages in which they operate. That includes not only the official policies, but also other content such as a company's blog posts and articles announcing or explaining important changes. They should ensure that they are hiring native-language speakers with sufficient knowledge of the local contexts. Additionally, they should employ experts in dialects of widely spoken languages like Arabic and Spanish, as well as experts in languages widely used for inciting violence. They should not use fully automated moderation without human involvement unless they have a high confidence in the accuracy of their automated systems, such as those for detecting spam and malware. When automated tools are used as part of a content moderation process, they must offer users the ability to appeal decisions to human moderators. Platforms must also invest in developing more accurate automated content moderation tools in languages other than English, including improving the labeling of training materials and the ability of automation to manage non-Latin alphabets. In addition, platforms should work with civil society to understand hyper local contexts such as the use of transliteration or substitute words to avoid automation.

5. **Increased transparency:** Platforms should increase transparency and accountability in their content moderation practices. The Santa Clara Principles, which were updated and elaborated in 2021, provide concrete guidance for doing so[11]. Because users cannot adhere to rules that are not made public and easily accessible, platforms should publish clear, precise, and accessible content policies that explain when they will take action with

---

[11] The Santa Clara Principles on Transparency and Accountability in Content Moderation, https://santaclaraprinciples.org/ (The Principles cover a broad range of specific information types and practices, from publishing "clear and precise rules and policies" to notifying users about where and how automation is used to ensuring "that those making moderation and appeal decisions understand the language, culture, and political and social context of the posts they are moderating."

respect to users' content or accounts, and they should allow users to see the changes to their policies made over time. As discussed above, platforms should publish human rights due diligence reviews publicly as much as possible. Platforms should also provide information about how much they are spending on content moderation and protection of human rights in different countries and the number of staff or contractors working on content moderation and protection of human rights broken down by specific markets. At a minimum, if they cannot provide raw numbers, they should provide comparative percentages.

6. **Clarity about so-called "Terrorist and Violent Extremist Content" (TVEC):** Platforms should be fully transparent regarding any content guidelines or rules related to the classification and moderation of "terrorism" and "extremism", including how they define TVEC, exceptions to those rules, and how the company determines when to make such exceptions. Platforms should push back against attempts by governments to use the TVEC label to silence dissent and independent reporting, and should be clear about how their "TVEC" policies relate to other policies such as incitement to violence. Determining whether content promotes terrorism or violent extremism in violation of a platform's content policies or legal requirements is often highly context-dependent. Platforms should use models of contextual analysis to make such determinations on a case-by-case basis. However, if platforms do use list-based designations of groups and individuals classified as "terrorist" or "extremist, even just for surfacing content for review, they should at a minimum publish their internal lists of such groups and individuals and explain how they refer to external lists such as those published by the UN and USA. They should also ensure users have a way to promptly and meaningfully appeal decisions to remove allegedly violating policies against terrorism or violent extremism, and incorporate contextual considerations into reviews of such appeals.

7. **Multi Stakeholder Debriefs:** Finally, we call on platforms to ensure that when they take extraordinary actions or are forced to engage in a "surge response" to emergencies, they take stock afterwards. What have platforms learned from how they were misused in Myanmar, how they managed content from human rights defenders in Syria and Yemen, or how they failed to address incitement to violence against minorities in India and Sri Lanka? Those lessons should be shared. There is no need to reinvent the wheel every time. In the global civil society community, we learn from each other and from history. Platforms should do the same.

As a global community of civil society actors, we do not demand a one-size-fits-all approach to responding to human rights crises. What we are asking platforms to do is to invest more time and effort in improving their operations now, not when unfolding violence gets into the media spotlight and it is often already too late to act.

Finally, when the news cameras and politicians' statements are turned away from Ukraine, we insist that platforms do not ignore the needs of the Ukrainian people.

Representatives of our organizations are happy to meet with representatives from social media platforms to discuss this statement as well as possible follow-up actions.

Signed,
Access Now
Association for Progressive Communications (APC)
Australian Muslim Advocacy Network (AMAN)
Center for Democracy & Technology (CDT)
Civil Liberties Union for Europe (Liberties)
Chayn
Derechos Digitales
Digital Action
Digital Africa Research Lab
Digital Security Lab Ukraine
Digital Rights Foundation
Doublethink Lab
Electronic Frontier Foundation (EFF)
European Sex Workers' Rights Alliance (ESWA)
Fight for the Future
Global Forum for Media Development (GFMD)
Global Project Against Hate and Extremism
Global Voices Advox
INSM Network for Digital Rights in Iraq
Jordan Open Source Association
Miaan Group
Mnemonic
New America's Open Technology Institute
Ranking Digital Rights
Social Media Exchange (SMEX)
Taraaz
The Dangerous Speech Project
WITNESS
Woodhull Freedom Foundation
Zašto ne (Bosnia and Herzegovina)
7amleh- The Arab Center for the Advancement of Social Media