

Mandates of the Working Group on the issue of human rights and transnational corporations and other business enterprises; the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; the Special Rapporteur on the rights to freedom of peaceful assembly and of association; the Special Rapporteur on the independence of judges and lawyers; the Special Rapporteur on minority issues; the Special Rapporteur on the situation of human rights in the Palestinian territory occupied since 1967; the Special Rapporteur on the right to privacy and the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism

Ref.: AL OTH 20/2024
(Please use this reference in your reply)

18 April 2024

Dear Mr. Zuckerberg,

We have the honour to address you in our capacities as Working Group on the issue of human rights and transnational corporations and other business enterprises; Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; Special Rapporteur on the rights to freedom of peaceful assembly and of association; Special Rapporteur on the independence of judges and lawyers; Special Rapporteur on minority issues; Special Rapporteur on the situation of human rights in the Palestinian territory occupied since 1967; Special Rapporteur on the right to privacy and Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, pursuant to Human Rights Council resolutions 53/3, 52/9, 50/17, 53/12, 52/5, 1993/2A, 46/16 and 49/10.

We are independent human rights experts appointed and mandated by the United Nations Human Rights Council to report and advise on human rights issues from a thematic or country-specific perspective. We are part of the special procedures system of the United Nations, which has 60 thematic and country mandates on a broad range of human rights issues. We are sending this letter under the communications procedure of the Special Procedures of the United Nations Human Rights Council to seek clarification on information we have received. Special Procedures mechanisms can intervene directly with Governments and other stakeholders (including companies) on allegations of abuses of human rights that come within their mandates by means of letters, which include urgent appeals, allegation letters, and other communications. The intervention may relate to a human rights violation that has already occurred, is ongoing, or which has a high risk of occurring. The process involves sending a letter to the concerned actors identifying the facts of the allegation, applicable international human rights norms and standards, the concerns and questions of the mandate-holder(s), and a request for follow-up action. Communications may deal with individual cases, general patterns and trends of human rights violations, cases affecting a particular group or community, or the content of draft or existing legislation, policy or practice considered not to be fully compatible with international human rights standards.

In this connection, we would like to bring to your attention information we have received concerning **a number of alleged human rights violations in Israel and the occupied Palestinian territory since 7 October 2023, including increased criminalization, surveillance, harassment and attacks against individuals who are**

Meta

peacefully expressing dissent or advocating for Palestinian rights on social media. The proliferation of hate speech, incitement to violence, disinformation, and bias and systematic censorship of Palestinian and pro-Palestinian human rights voices and content are allegedly also on the rise on social media platforms including Meta, X, Google (headquartered in the United States of America), and Telegram (headquartered in the United Arab Emirates).

We would like to remind your company that our mandates expressed deep concern about the restriction of a significant number of accounts and content published by users of Facebook, Instagram and WhatsApp, (now Meta) during the September 2021 conflict between Israel and Palestine on 29 December 2022 (OTH 126/2022) and 8 September 2021 (OTH 212/2021). We regret that in those instances, we did not receive a response from your company.

According to the information received:

Disproportionate censorship of Palestinian voices and content on Palestine

The alleged disproportionate censorship, moderation and restriction of pro-Palestinian human rights voices, including of journalists and activists reporting from the ground, content creators and others on Meta's platforms appears to be rooted in the company's apparent discriminatory content moderation practices and policies.

Meta designated the 7 October events as a Terrorist Attack under its Dangerous Organisations and Individuals (DOI policy). This means that Meta would remove all content linked to the event, unless the user clearly shares it in the context of news reporting or condemnation of the event. Meta has stated that it has taken several measures in response to an increase in violent and graphic content being posted to its platforms, including lowering the confidence threshold for the automatic systems classifiers to identify and remove content, blocking hashtags, and restricting live streaming.

Since 7 October 2023, content supporting Palestinian rights has faced disproportionate censorship, including deletion, restriction, takedowns and the deliberate suppression of hashtags. Between 7 and 10 of October 2023, Meta reportedly removed or marked as disturbing more than 795,000 posts. The Palestinian Observatory for Digital Rights Violations has documented 2,846 violations from 7 October until 15 December, 826 of which occurred on Meta's online platforms Instagram and Facebook. For example, Meta immediately censored the Arabic hashtag #iron_swords (طوفان_الاقصى) on 7 October but did not censor the parallel hashtag in Hebrew (חרבות_ברזל).

Palestinian journalists, activists, and human rights defenders have faced significant and disproportionate censorship and have been routinely suspended on Meta's platforms. Palestinian and international news outlets, including but not limited to Quds News Network, Ajjyal Radio Network, BreakThroughNews, 24FM, Mondoweiss, Palestinian Refugees Portal, as well as journalist accounts such as Faten Elwan, Motaz Azaiza, Saleh Al-Jafarawi, and Ahmed Shihab-Eldin have all experienced and/or continue to experience

content takedowns and account restrictions on Instagram and Facebook. Journalists have also experienced account suspensions and content takedowns under Meta's "community guidelines" and Dangerous Individuals and Organizations policy for reporting news in Arabic language and/or from occupied Gaza. Users have also experienced shadow banning on a wide scale for posting content on Palestine. This includes limiting views on Instagram stories particularly if users share news content from other pages.

The Palestinian flag emoji has been flagged as "potentially offensive" by Instagram, resulting in it being hidden. Meta and Instagram have reportedly prevented the display of images of Palestinian victims in hospitals because they were deemed to be "nude" pictures. Additionally, reports reveal that WhatsApp's artificial intelligence image generator recently created emojis of gun-wielding children when prompted with "Palestinian".

According to the information received, Meta rolled out a notification discouraging users from following or sharing content from popular Palestinian accounts with a "Are you sure" prompt for sharing news updates from occupied Gaza, under the allegation that these accounts "repeatedly posted false information" or violated Meta's community guidelines. It is also alleged that Meta has restricted pro-Palestinian human rights accounts from posting comments on Instagram, including comments being automatically hidden when talking about Palestine or using the Palestinian flag emoji.

On 15 October, a Meta spokesperson blamed the reduced reach of pro-Palestine human rights posts on a bug that "affected accounts equally around the globe and had nothing to do with the subject matter of the content".

Reportedly, Instagram auto-translated user bios that included "Palestinian" and an Arabic phrase meaning "praise be to God" to "Palestinian terrorists are fighting for their freedom". Meta publicly apologized and said in a statement to several newspapers that the problem that caused the "inappropriate Arabic translations" has been fixed.

Recently, Meta's internal documents acknowledged that Meta had lowered the level of certainty its algorithmic censorship system needed to remove Arabic posts from 80 percent confidence that the post broke the rules, to 25 percent. In other words, the accuracy rate of posts being removed by Meta significantly decreased for Arabic posts with regard to whether these posts actually contained policy violations. Academics have repeatedly criticized the lack of efficient content moderation in Arabic, stating that Arabic content is subject to "inconsistent moderation", meaning that some content will be over-moderated, while other content will be left untouched despite violating the platforms' standards. These inconsistencies have limited users' ability to engage in meaningful political debates in the region. Further, Meta's internal documents indicated that the Hebrew hostile speech classifiers did not detect speech violating policies effectively, because they did not have sufficient data for the system to function accurately. In fact, in 2022, an independent due diligence report commissioned by Meta to assess the impact of its content moderation actions and policies in Israel and the occupied Palestinian territory pointed to

the absence of these classifiers for the under-enforcement of moderation of hate speech content in Hebrew. It also found that Meta had negatively affected the human rights of Palestinian users in areas such as freedom of expression, freedom of assembly, political participation and non-discrimination.

In the second week of November, Meta removed a video posted on Instagram showing what appears to be the aftermath of a strike on or near Al-Shifa Hospital in occupied Gaza City, during Israel's ground offensive in the north of the occupied Gaza Strip, for violating its Violent and Graphic Content Community Standard. On 19 December 2023, Meta's Oversight Board overturned this decision citing "the importance of Meta respecting the right to freedom of expression and other human rights of all those affected by these events, and their ability to communicate in this crisis."

Dissemination of hate speech, incitement to violence and the dehumanization of Palestinians

The proliferation of hate speech and incitement to violence in the context of the 7 October attacks and its aftermath on various platforms presents a grave concern. While hate speech against Israelis, Jews and antisemitism are also strongly on the rise, social media platforms appear to have a biased and discriminatory approach when it comes to protecting Palestinians against hate speech. We have received reports that content relating to hate speech, incitement to violence, dehumanization and calls for genocide against Palestinians remains uncensored and prominent on the platform. Of the 2,800+ violations recorded through the Palestinian Observatory for Digital Rights Violations between 7 October and 15 December, more than half (2,012 cases) were classified as harmful content. This content includes hate speech, incitement to violence, online harassment and smear campaigns. Most of those cases occurred on Facebook, X (formerly Twitter), and Telegram.

Specifically related to hate speech in Israel against Palestinians, 7amleh's "Violence Indicator" —a linguistic model that uses artificial intelligence technology that monitors the spread of hate speech and violence in Hebrew against Palestinians and their advocates on social media platforms— has counted over 2,381,444 pieces of hate speech and violent content posted on X, Facebook, and Telegram.

There have been reports of Meta's platforms being used by Israelis, including senior officials, to explicitly call for violence against Palestinians. Meta is allowing verified state accounts that belong to the Israeli government — including politicians, the Israeli army and its spokespeople — to disseminate content that justifies military intervention, including attacks on hospitals and ambulances, filmed confessions of Palestinian detainees, and almost daily "evacuation" orders for Palestinian civilians. It has been reported that Meta approved ads that explicitly called for a "holocaust for the Palestinians" and wiping out "Gazan women and children and the elderly".

Until September 2023, Meta did not have classifiers to automatically detect and remove hate speech in Hebrew. A recent internal memo revealed that Meta lacks

sufficient training data to use the newly built Hebrew classifier on Instagram comments.

Dissemination of disinformation

There have been widespread instances of disinformation and misinformation on various online platforms, in addition to AI-generated deepfakes that increase fears about AI-powered models fostering misleading information. This significantly harms freedom of expression and access to information. Since 7 October, false information has been weaponized to propagate hatred and incite violence against Palestinians. It is also used to manipulate the opinions of users globally. This disinformation appears to be employed to rationalize collective punishment against all Palestinians and is often accompanied by incitement and calls for violence. Meta indicated they have “the largest third-party fact-checking network of any platform”, but given the number of false accusations that spread across their platforms throughout October, their efforts still seem to be insufficient.

While we do not wish to prejudge the accuracy of these allegations, we express our deep concern that Meta’s treatment of pro-Palestinian human rights content and voices on its platforms represents a serious threat to freedom of expression and access to information, freedom of peaceful assembly and public participation about issues of global concern, stifling political participation and creating a chilling effect. This censorship not only affects activists and human rights defenders, but also impedes journalists’ ability to report effectively and hampers the right of the public to be informed.

The six months of conflict since 7 October 2023, now plausibly amounting to a real and imminent risk of genocide according to the International Court of Justice,¹ have placed a spotlight on the role and obligations of technology companies, particularly social media platforms, in situations of armed conflicts. We remind Meta of the importance of free public discourse during conflicts, particularly given the real and imminent risk of genocide in this instance. It is critical for the public to know about incidents and for human rights groups to document and report on human rights violations. Silencing pro-Palestinian human rights voices across digital platforms and social media limits the visibility and accessibility of Palestinian voices, making it difficult to build public understanding of their situation. International platforms such as Meta therefore have an important duty and role to play in protecting freedom of expression across the globe, and it is therefore essential that your platform allows for a level playing field for all to have their voices heard and their content fairly disseminated.

We are also deeply concerned that many of the voices of journalists, human rights defenders, civil society activists and organisations and many others concerned or directly affected by the current situation in occupied Gaza have been suppressed by Meta’s policies and practices. This is particularly concerning as many prominent Palestinian journalists, citizen journalists and human rights defenders in the occupied Palestinian territory have no means other than social media to report on human rights

¹ <https://www.icj-cij.org/case/192>

violations and share the most up to date and detailed information on events as they unfold.

When compared to its moderation of Hebrew-language content or its response to the military occupation by Israel, Meta's actions reveal potential discrimination in the way it treats Palestinian users and content about Palestine. It is deeply worrying that Meta significantly relies on automated content moderation tools that seem to be insufficiently trained in Arabic and its dialects. We are concerned that the allegations outlined above suggest a deep-rooted bias in Meta's datasets and machine-learning systems. We wish to highlight that any form of content moderation restricting freedom of expression must be applied in a non-discriminatory manner. We further wish to highlight the importance of archiving all content that can potentially constitute evidence of war crimes in case such content is removed from the platforms.

Further, the proliferation of hate speech, incitement to violence and disinformation hinders the enjoyment of a variety of human rights, including the rights to life (article 6 of the ICCPR), security of person, including bodily and mental integrity (article 9 of the ICCPR), the right to the highest attainable standard of physical and mental health (article 12 of the ICESCR), non-discrimination (article 2 of the ICCPR), honour and reputation (article 17 of the ICCPR), and to protection from propaganda for war and advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law (article 19 of the ICCPR). There is an increasing recognition of the deep impact of systemic oppressions on mental health. As the Secretary General of the United Nations indicated in the organization's Strategy and Plan of Action on Hate Speech, hate speech is a menace to democratic values, social stability and peace.

Electronic communication services, social media platforms and search engines provide an ideal environment both for the anonymous delivery of targeted threats, and for the mass dissemination of intimidating, defamatory, degrading, deceptive or discriminatory narratives. Individuals or groups systematically targeted by cyber-harassment are generally left without any effective means of defense, escape or self-protection and, at least in this respect, often find themselves in a situation of "powerlessness" comparable to physical custody. Likewise, the generalized shame inflicted by public exposure, defamation and degradation can be just as traumatic as direct humiliation by perpetrators in a closed environment. As various studies on cyberbullying have shown, harassment alone in comparatively limited environments can expose targeted individuals to extremely elevated and prolonged levels of anxiety, stress, social isolation and depression and significantly increases the risk of suicide, which may amount to psychological torture (A/HRC/43/49, paras 74-75).

Broadly speaking, online racism not only affects targeted groups of people, but exercises greater influence on society at large, emphasizing divisions, exacerbating fractures and strengthening polarization within the society. The above-mentioned elements become more meaningful when we consider the rising importance that young people attach to cyberspace, and the latter's potential to influence their choices and values. The latter, coupled with the lack of an internationally agreed definition of terrorism and the ingrained and systematic abuse of counter-terrorism and security laws and practices, the use of digital technologies creates a compounded and intersectional set of human rights challenges (A/HRC/52/49, para. 17).

In OTH 212/2021, we welcomed the adoption by Meta of a Corporate Human Rights Policy, which makes a direct reference to human rights including the UN Guiding Principles on Business and Human Rights, the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights. We continue to urge strongly that Meta integrates international human rights standards into all stages of its operations. In practice, this means that all content restrictions on the right to freedom of expression on the platform should pursue a legitimate aim as prescribed under international human rights law and be necessary and proportionate to achieve the legitimate aim.

We further urge Meta to put international human rights standards at the centre of your business model, including in the management of content on your platforms and your recommender systems. This can be done through carrying out regular human rights impact assessments of products and services, and ensuring that policies defining and prohibiting hate speech on your platform, in accordance with international human rights standards, including on freedom of expression, of peaceful assembly and of association. We underline the importance of contextual, conflict-sensitive analysis, grievance mechanisms, crisis protocols and evaluation in the context of content moderation, and the importance to allocate sufficient capacity in this regard, including through human oversight and expertise, including linguistic and cultural expertise, and the involvement of communities and individuals most affected by hate speech. This is particularly important when the situation is plausibly amounting to a real and imminent risk of genocide.

Specifically, we urge Meta to adopt a definition of incitement, in line with articles 19(3) and 20 of the ICCPR and the standards spelled out in the Rabat Plan of Action, when addressing advocacy of national, racial or religious hatred that may constitute incitement to discrimination, hostility or violence (A/HRC/22/17/Add.4). We also urge Meta to adopt the model definition of incitement to terrorism advanced by the mandate of the United Nations Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism (A/HRC/16/51, practice 8).

As we mentioned, we welcome the fact that Meta has published its policies relating to hate speech and violence and incitement. However, in the light of the above international standards, we urge that these terms are clearly defined and to ensure that it does not lead to unnecessary or disproportionate interference with freedom of expression (see general comment No. 34 of the UN Human Rights Committee, para. 46). Meta should publish its entire list of designated Dangerous Organisations and Individuals (DOI) and put in place due process standards empowering users to understand how assessments are made and appeal decisions regarding the inclusion of individuals or organisations. We are concerned that poorly defined concepts or overly broad application may result in the suppression of minority or dissenting views (See e.g. A/HRC/31/65). Further, we would like to underscore the need for transparency, procedural fairness and effective means for an individual to challenge a decision once made. Moreover, we are concerned that Meta's DOI policy consists of three key actions that are prohibited on Meta's platforms (support, representation and glorification of designated individuals and organizations), which are vague and overly-broad metrics. The UN Human Rights Committee confirmed in its general comment No. 34 that "such

offenses as “encouragement of terrorism” and “extremist activity” as well as offenses of “praising”, “glorifying”, or “justifying” terrorism, should be clearly defined to ensure that they do not lead to unnecessary or disproportionate interference with freedom of expression.” Further, we note that the expressions that Meta prohibits for constituting praise or glorification of terrorism are not public, and therefore, users are not aware of the rules under which their speech is restricted.

When faced with problematic requests, companies should seek clarification from the relevant authorities and provide users with full transparency about the volume and type of private requests they receive. Businesses have the responsibility to take every measure at their disposal to resist the implementation of requests that are likely to contravene their human rights responsibilities under the UN Guiding Principles on business and human rights (UNGPs). Social media companies in particular should contribute to facilitate the exercise of the rights of expression, association and peaceful assembly, to public advocacy and education campaigns, to advance respect and non-discrimination and combat the root causes of racist and gender-based hate speech, including discriminatory social norms and harmful racial and gender stereotypes, and actively to partner and show public support for feminists and women human rights defenders working on all anti-discrimination issues. We strongly urge Meta to apply the UNGPs into all stages of its operations. In practice, this means that all content restrictions on the right to freedom of expression on the platform should pursue a legitimate aim as prescribed under international human rights law and be necessary and proportionate to achieve the legitimate aim, and be applied in a non-discriminatory manner. A rigorous human rights due diligence process, conducted in line with the UNGPs would allow Meta to achieve this goal and avoid repeat failures from the past.

We urge Meta to review and revise its policies and practices to prevent or remediate infringements of the right to freedom of opinion and expression through censorship, shadow-banning, blocking and deleting accounts, and removing or restricting content, comments and live-streams. We urge Meta to undertake human rights due diligence to ensure that your platforms are removing incitement to violence, dehumanization and calls for genocide against Palestinians in a fair and balanced way, in line with the UNGPs and the requirements and conditions of article 19 (3) ICCPR and article 20 of the ICCPR – in the same way incitement to violence against Israelis is, and must be, prohibited.

In this regard, and as mentioned above, we encourage Meta to conduct a human rights impact assessment that includes a strict and thorough analysis of local contexts where groups at heightened risk of vulnerability and marginalization live, such as minorities or people living under occupation as well as in situations of armed conflict. This is particularly important when the situation is plausibly amounting to a real and imminent risk of genocide. We draw your attention in particular to the reports of the Special Rapporteur for Freedom of Expression on the issue of disinformation in times of conflict (A/77/288), and of the Working Group on the issue of human rights and transnational corporations and other business enterprises clarifying the practical measures that States and business enterprises should take to prevent and address business-related human rights abuse in conflict and post-conflict contexts, focusing on heightened human rights due diligence and access to remedy (A/75/212). To that end, multi-stakeholder engagement with civil society and other international and regional actors is critical to ensure the voices of those already marginalised are not further

suppressed. We stress the need for enhanced transparency, human rights due diligence, accountability and remediation by Meta to ensure human rights are not unduly interfered.

Finally, we wish to highlight A/HRC/55/L.30 on the human rights situation in the occupied Palestinian territory and the obligation to ensure accountability and justice. This resolution includes a call for an end to all ongoing policies of harassment, threats, intimidation and reprisals, detention and expulsion against human rights defenders, journalists, media workers and civil society actors who peacefully advocate for the rights of the Palestinian people, and underscores the need to investigate all such acts and to ensure accountability and effective remedies. The resolution also includes an expression of concern at the spread of disinformation and propaganda, including on the Internet, which can be designed and implemented so as to mislead, to violate human rights, including the right to freedom of expression, to spread hatred, racism, xenophobia, negative stereotyping or stigmatization and to incite violence, discrimination and hostility, and emphasizes the important contribution by journalists in countering this trend.

We share these concerns in the hope to pursue our dialogue with Meta on the best ways to advance international human rights norms and standards pertaining to freedom of opinion and expression, especially in a context where your company works to assess and improve its content review operations in conflict situations, particularly in situations that are plausibly amounting to a real and imminent risk of genocide.

In connection with the above alleged facts and concerns, please refer to the **Annex on Reference to international human rights law** attached to this letter which cites international human rights instruments and standards relevant to these allegations.

As it is our responsibility, under the mandates provided to us by the Human Rights Council, to seek to clarify all cases brought to our attention, we would be grateful for your observations on the following matters:

1. Please provide any additional information and/or comment(s) you may have on the above-mentioned allegations.
2. Please provide information on the number of accounts and amount of content that have been restricted in relation to the events since 7 October 2023, the reasons thereof, and the remedy measures in place. Please also provide information about accounts and content which were eventually reinstated.
3. Please provide detailed information as to the measures, including human rights due diligence, that your company has taken in line with the UN Guiding Principles on Business and Human Rights to identify, prevent, mitigate and account for adverse human rights impacts caused by your company's products and services, or to which they may have contributed or be directly linked.
4. Please provide information on your company's efforts, if any, to ensure transparent communication on human rights issues, including the

responsibility to conduct heightened human rights due diligence processes in conflict-affected contexts (see Working Group on business and human rights report A/75/212). Please indicate whether you take any specific steps when the situation is plausibly amounting to a real and imminent risk of genocide.

5. Please indicate how you conduct stakeholder engagement, both in general and specifically in conflict-affected contexts to mitigate the lack of information, the polarization and the high level of mistrust which usually exist among groups and communities (see Working Group on business and human rights report A/75/212). Please indicate whether you take any specific steps when the situation is plausibly amounting to a real and imminent risk of genocide.
6. Please provide information on the steps Meta is taking to address potential bias in its training datasets and ensure that automated content moderation is applied in a non-discriminatory manner.
7. Please provide information about the restrictions imposed on pro-Palestinian human rights content on Meta's platforms, and whether Meta informed the affected individuals of the measures taken against their accounts or posts.
8. Please describe the measures that your company has taken or is considering taking to screen advertisements and/or posts that disseminate disinformation causing polarization and racial discrimination within society, while ensuring that these measures meet international human rights standards, including for the promotion and protection of freedom of expression.
9. Please enumerate research and studies that you are conducting to assess the intended and unintended impacts that your platforms may have on minorities.
10. Please indicate how human rights and international standards for freedom of expression, association and peaceful assembly are protected and guaranteed in your content moderation policies aimed at countering hate speech, disinformation and incitement to violence.
11. Please provide information about the strategies that your company has adopted to prevent or mitigate account or content removal requests that are inconsistent with international human rights standards, including when removal requests come from States.
12. In relation to the events described in this communication, did your company consider options other than removing content and accounts to protect the right of users to freedom of expression? Did your company inform the affected individuals of the measures taken against their accounts or posts?

13. Please provide information about specific measures taken to address instances of advocacy of hatred that constitutes incitement to violence and discrimination on the platforms. Please indicate specific remedial measures that your company has taken or is considering taking to prevent racialized hate speech.
14. Please provide information on whether your company has put in place human rights training for personnel in relevant business functions.
15. Please state the steps that your company has taken following the findings of the independent report on the Human Rights Due Diligence of Meta's Impacts in Israel and Palestine that Meta had negatively affected the human rights of Palestinian users in areas such as freedom of expression, freedom of assembly, political participation and non-discrimination.
16. Please provide information on any action taken by your company to prevent recurrence of the kind of allegations specified in this letter. As part of this response, please explain what monitoring and evaluation systems you have put in place to ensure the effectiveness of actions taken to mitigate and prevent the allegations mentioned in this letter.
17. Please provide information on steps taken by your company to establish, or participate in effective operational-level grievance mechanisms, in line with the UN Guiding Principles on Business and Human Rights, to address adverse human rights impacts caused by your company throughout your operations globally.

We would appreciate receiving a response within 60 days. Past this delay, this communication and any response received from you will be made public via the communications reporting [website](#). They will also subsequently be made available in the usual report to be presented to the Human Rights Council.

We may publicly express our concerns in the near future as, in our view, the information upon which the press release will be based is sufficiently reliable to indicate a matter warranting immediate attention. We also believe that the wider public should be alerted to the potential implications of the above-mentioned allegations. The press release will indicate that we have been in contact with you to clarify the issue/s in question.

Please note that letters expressing similar concerns will also be sent to the chief executive officers of other relevant social media companies and the States where they are domiciled.

Please accept, Mr. Zuckerberg, the assurances of our highest consideration.

Robert McCorquodale
Chair-Rapporteur of the Working Group on the issue of human rights and
transnational corporations and other business enterprises

Irene Khan
Special Rapporteur on the promotion and protection of the right to freedom of opinion
and expression

Clement Nyaletsossi Voule
Special Rapporteur on the rights to freedom of peaceful assembly and of association

Margaret Satterthwaite
Special Rapporteur on the independence of judges and lawyers

Nicolas Levrat
Special Rapporteur on minority issues

Francesca Albanese
Special Rapporteur on the situation of human rights in the Palestinian territory
occupied since 1967

Ana Brian Nougrères
Special Rapporteur on the right to privacy

Ben Saul
Special Rapporteur on the promotion and protection of human rights and fundamental
freedoms while countering terrorism

Annex

Reference to international human rights law

In connection with the above-alleged facts and concerns, we would like to draw your attention to the applicable international human rights norms and standards, as well as authoritative guidance on their interpretation.

We would like to highlight the UN Guiding Principles on Business and Human Rights, which were unanimously endorsed in 2011 by the Human Rights Council in its resolution (A/HRC/RES/17/31) after years of consultation with governments, civil society and the business community. The Guiding Principles have been established as the authoritative global standard for all States and businesses to prevent and address business-related adverse human rights impacts. These Guiding Principles are based on the recognition of:

- a) "The existing obligations of States to respect, protect and fulfil human rights and fundamental freedoms;
- b) The role of business enterprises as specialised bodies or corporations performing specialised functions, which must comply with all applicable laws and respect human rights;
- c) The need for rights and obligations to be matched by appropriate and effective remedies when they are violated".

The Guiding Principles also make clear that businesses have an independent responsibility to respect human rights. Principles 11-24 and 29-31 provide guidance to businesses on how to meet their responsibility to respect human rights and to provide remedies where they have caused or contributed to adverse impacts. The Guiding Principles have identified two main components of the corporate responsibility to respect human rights, which require "business enterprises to:

- a) Prevent their own activities from causing or contributing to adverse human rights impacts and address those impacts when they occur;
- b) Seek to prevent or mitigate adverse human rights impacts directly related to operations, products or services provided through their business relationships, even where they have not contributed to those impacts".
(Guiding principle 13)

The commentary to guiding principle 13 notes that businesses can be affected by adverse human rights impacts, either through their own activities or as a result of their business relationships with other parties (...) The 'activities' of business enterprises are understood to include both actions and omissions; and their 'business relationships' include relationships with business partners, entities in their value chain and any other non-State or State entities directly linked to their business operations, products or services.

To meet their responsibility to respect human rights, businesses should have in place policies and procedures appropriate to their size and circumstances:

- a) A political commitment to uphold their responsibility to respect human rights;
- b) A human rights due diligence process to identify, prevent, mitigate and account for how they address their human rights impact;
- c) Processes to redress any adverse human rights impacts they have caused or contributed to (guiding principle 15).

According to guiding principles 16-21, human rights due diligence involves:

- a) Identifying and assessing actual or potential adverse human rights impacts that the enterprise has caused or contributed to through its activities, or that are directly related to the operations, products or services provided by its business relationships;
- b) Integrate the results of impact assessments into relevant business functions and processes, and take appropriate action in accordance with their involvement in the impact;
- c) Monitor the effectiveness of the measures and processes adopted to address these adverse human rights impacts in order to know whether they are working;
- d) Communicate how adverse effects are addressed and demonstrate to stakeholders - particularly those affected - that appropriate policies and processes are in place to implement respect for human rights in practice.

This process of identifying and assessing actual or potential adverse human rights impacts should include substantive consultation with potentially affected groups and other stakeholders (guiding principle 18).

Where an enterprise causes or is likely to cause an adverse human rights impact, it should take the necessary steps to end or prevent that impact. “The establishment of operational-level grievance mechanisms for those potentially affected by corporate activities can be an effective means of redress provided they meet certain requirements listed in principle 31 (guiding principle 22).

Furthermore, business enterprises should remedy any actual adverse impact that they cause or to which they contribute. Remedies can take a variety of forms and may include apologies, restitution, rehabilitation, financial or non-financial compensation and punitive sanctions (whether criminal or administrative, such as fines), as well as the prevention of harm through, for example, injunctions or guarantees of non-repetition. Procedures for the provision of remedy should be impartial, protected from corruption and free from political and other attempts to influence the outcome (commentary to guiding principle 25).

In addition, principle 18 highlights the essential role of civil society and human rights advocates in helping to identify potential negative human rights impacts of business. We also note that in its recent guidance on how to ensure respect for human rights defenders (A/HRC/47/39/Add.2), the Working Group on Business and Human Rights highlighted the urgent need to address the negative impacts of business activities on human rights defenders in particular. The commentary on principle 26 of the Guiding Principles on Business and Human Rights stresses that States, in order to ensure that the legitimate activities of human rights defenders are not impeded. Businesses have an independent responsibility to take the necessary steps to cease or prevent their contribution and to use their influence to mitigate as far as possible any remaining impact that contributes or may contribute to a negative impact on human rights (commentary on guiding principle 19) and should remediate any actual negative impact they cause or contribute to. Redress procedures must be impartial, protected from corruption and free from political or other attempts to influence the outcome (commentary on guiding principle 25).

We would also like to draw your attention to the report of the Special Rapporteur on minority issue hate speech, social media and minorities (A/HRC/46/57). The report specifically addresses the targeting of minority women, and the obligations of businesses, including by encouraging social media companies to:

- a) Put in place in procedures and mechanisms for the mandatory collection of data on hate speech, and at a minimum on incitement to genocide and advocacy that constitutes incitement to discrimination, hostility or violence. Such data should be disaggregated according to the basis of the hatred, whether national, ethnic, racial or religious hatred, which all relate to vulnerable minorities, or hatred affecting other vulnerable groups, on the basis of such factors as gender and sexual orientation.
- b) Manage hate speech on their platforms with reference to the human rights implications of their products, including algorithms and artificial intelligence programs such as chatbots. They must therefore have in place human rights review processes that refer to and focus specifically on the groups most susceptible to hate speech in the States concerned, and in particular minorities.
- c) Commit clearly, in the social media's content moderation systems and community standards, to protecting vulnerable and marginalized minorities and other groups. Minorities should specifically be identified as priorities.
- d) Apply the UNGPs in their own operations. Human rights standards should be integrated systematically into the content policies and decision mechanisms of social media platforms, so that, as specified in the UNGPs, the companies "comply with all applicable laws and respect internationally recognized human rights wherever they operate", and "treat the risk of causing or contributing to gross human rights abuses as a legal compliance issue wherever they operate", for which they could be liable.

- e) Comprehensively address distortion and systemic bias against Jews and Muslims, as evidence suggests that antisemitism and Islamophobia are pressing challenges for minority rights.
- f) In order to improve mechanisms and content vetting policies for the handling of hateful content, and to ensure incorporation of the concerns of the main targets of hate speech in social media, the Special Rapporteur urges that minorities, as the most targeted and vulnerable groups, be represented in advisory and other relevant boards.
- g) Look into innovative, educational and preventive strategies that focus on protection of and respect for diverse communities in order to counter hate speech.

We would also like to draw your attention to the report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression on online hate speech (A/74/486). The report provides a set of recommendations to business, especially to those active in the ICT sector:

- a) Evaluate how their products and services affect the human rights of their users and the public, through periodic and publicly available human rights impact assessments;
- b) Adopt content policies that tie their hate speech rules directly to international human rights law, indicating that the rules will be enforced according to the standards of international human rights law, including the relevant United Nations treaties and interpretations of the treaty bodies and special procedure mandate holders and other experts, including the Rabat Plan of Action;
- c) Define the category of content that they consider to be hate speech with reasoned explanations for users and the public and approaches that are consistent across jurisdictions;
- d) Ensure that any enforcement of hate speech rules involves an evaluation of context and the harm that the content imposes on users and the public, including by ensuring that any use of automation or artificial intelligence tools involve human-in-the-loop;
- e) Ensure that contextual analysis involves communities most affected by content identified as hate speech and that communities are involved in identifying the most effective tools to address harms caused on the platforms;
- f) As part of an overall effort to address hate speech, develop tools that promote individual autonomy, security and free expression, and involve de-amplification, de-monetization, education, counter-speech, reporting and training as alternatives, when appropriate, to the banning of accounts and the removal of content.

In report A/74/486 to the UN General Assembly, the then Special Rapporteur on freedom of expression provides important guidance on how businesses can address hate speech on their platforms, while respecting international standards for freedom of expression. Businesses should evaluate how their products and services affect the human rights of their users and the public, through periodic and publicly available human rights impact assessments. Furthermore, businesses should adopt content policies based on international human rights law and enforcing the rules in line with international human rights standards, including relevant UN treaties, guidance from treaty bodies and special procedures mandate holders and the Rabat Plan of Action. Furthermore, it is crucial that any enforcement of hate speech rules involves an evaluation of context and the harm that the content imposes on users and the public, including by ensuring that any use of automation or artificial intelligence are subject to human oversight to minimize the risk of legitimate content being removed. Communities most affected by content identified as hate speech must be involved in the contextual analyses and in identifying tools to address harms caused on the platforms.

In her report A/HRC/47/25, the Special Rapporteur on freedom of expression called on social media companies to, in line with the UNGPs, review their business models and ensure that their business operations, data collection and data processing practices are compliant with international human rights standards, including article 19 of the International Covenant on Civil and Political Rights, as well as data protection principles and relevant national consumer protection standards. They should also conduct human rights impact assessments of their products, particularly of the role of algorithms and ranking systems in amplifying disinformation or misinformation. Such assessments should be conducted regularly and ahead of and following significant events such as national elections or major crises like the COVID-19 pandemic.

The Special Rapporteur on the rights to freedom of peaceful assembly and of association noted that international human rights norms should guide digital technology companies' governance and further recommends them "to prevent or mitigate the adverse human rights impacts of their involvement" (A/HRC/41/41, para. 84). In the same report, he called on them to "integrate early warning systems within business processes to identify human rights risks, and respond in a timely fashion (...); support the research and development of appropriate technological solutions to online harassment, disinformation and propaganda, including tools to detect and identify State-linked accounts and bots" (A/HRC/41/41, para. 86 (b)).

In addition, the Special Rapporteur on the rights to freedom of peaceful assembly and association has emphasized in various reports the importance of digital technology to exercise the mentioned rights, and in his report on freedom of assembly and association in the digital age, he detailed that those "(...) technologies are important tools for organizers who seek to mobilize a large group of people in a prompt and effective manner, and at little cost, and also serve as online spaces for groups of people that are marginalized by society and are confronted with restrictions when operating in physical spaces" (A/HRC/41/41 para. 11).

Additionally, we would like to refer to article 17 of the ICCPR, which protects the right to privacy and provides that no one shall be subjected to arbitrary or unlawful interference with his or her privacy. The Human Rights Committee stated in its general

comment 32 (para. 8) that the positive obligations on States Parties under ICCPR are only fully discharged if individuals are protected by the State, not just against violations of their rights by its agents, but also against acts committed by private persons or entities.

We would also like to highlight the reports to the Human Rights Council on the practical application of the UNGPs to the activities of technology companies (A/HRC/50/56) and its Addendum (A/HRC/50/56/Add.1). The report indicates that “Leaks of personal information, dissemination of hate speech, the undermining of democratic processes and “algorithmic discrimination” – to name but a few of the documented risks – can all negatively affect the ability of people to enjoy their human rights” and provides recommendations to technology companies to ensure respect for human rights.