

**Mandate of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism**

REFERENCE:  
OL OTH 46/2018

24 July 2018

Dear Mr. Zuckerberg,

In my capacity as the United Nations (UN) Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, pursuant to Human Rights Council resolution 31/3, I write to express my mandate's concerns regarding Facebook's definition of terrorism and terrorist organizations and certain aspects of its policies governing access to and use of its platform and services as these may adversely impact on the protection of human rights, including but not limited to freedom of opinion and expression, and undermine the broad public interest in respecting and upholding these rights.

The Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism is an independent expert appointed by the UN Human Rights Council. As per my mandate, I have been invited to gather, request, receive and exchange information on alleged violations of human rights and fundamental freedoms while countering and preventing terrorism and violent extremism, and to report regularly and publicly to the Human Rights Council and General Assembly about *inter alia* identified good policies and practices, as well as existing and emerging challenges and present recommendations on ways and means to overcome them.

On 23 April 2018, Facebook published a blog post titled 'Hard Questions: How Effective Is Technology in Keeping Terrorists off Facebook?', co-authored by Monika Bickert, Vice President of Global Policy Management and Brian Fishman, Global Head of Counterterrorism Policy. The blog post stated the following:

We define terrorism as: "Any non-governmental organization that engages in premeditated acts of violence against persons or property to intimidate a civilian population, government, or international organization in order to achieve a political, religious, or ideological aim."

On 25 April 2018 the post was updated to clarify that the above definition applied to terrorist organizations. Neither the blog post nor Facebook's Community standards<sup>1</sup> provide details on the process for arriving at this definition, including whether the process has benefitted from consultations with stakeholders with expertise in counter-terrorism and international human rights law. The legal basis of the definition is also not specified.

In addition to spelling out Facebook's definition of terrorist organizations, the post also provides enforcement data "against ISIS, al-Qaeda, and their affiliates",

---

<sup>1</sup>[https://www.facebook.com/communitystandards/violence\\_criminal\\_behavior/dangerous\\_individuals\\_organizations/](https://www.facebook.com/communitystandards/violence_criminal_behavior/dangerous_individuals_organizations/)

highlighting that the platform has stepped up its takedown efforts, having removed twice as much content in the first quarter of 2018 than in the previous quarter; that it proactively searched for such content (with 99% of posts removed flagged by Facebook, not users or other third parties); as well as the speed at which such takedown takes place, with the median time for removing newly uploaded content that violates its terms and conditions of use being less than a minute.

Against the above, I wish to express my concern at the overly broad definition of terrorism and terrorist organizations used by Facebook as well as the seeming lack of a human rights approach to content moderation policies. I recommend that Facebook amend such policies with due consideration to and consultation with the human rights concerns outlined below.

The promotion and protection of human rights in the context of the challenges posed by terrorism is at the heart of my mandate. Terrorism poses a serious challenge to the very tenets of the rule of law, the protection of human rights and their effective implementation. Effectively combatting terrorism and ensuring respect for human rights are not competing but complementary and mutually reinforcing goals, as it has been unanimously recognized by the UN General Assembly in the Global Counter-Terrorism Strategy.<sup>2</sup> Moreover, relevant provisions of Security Council resolutions 1373 (2001), 1456 (2003), 1566 (2004), 1624 (2005), 2178 (2014), 2341 (2017), 2354 (2017), 2368 (2017), 2370 (2017), 2395 (2017) and 2396 (2017); as well as Human Rights Council resolution 35/34 and General Assembly resolutions 49/60, 51/210, 72/123 and 72/180 require that any measures taken to combat terrorism and violent extremism, including incitement of and support for terrorist acts, comply with States' obligations under international law, in particular international human rights law, refugee law, and international humanitarian law.

The role and influence of the private sector and, in particular, of tech companies, such as those owning and operating major social media platforms, in enabling and facilitating communication between a wide variety of stakeholders and, as a consequence, on the public's access to seek, receive and impart information cannot be ignored. Companies like Facebook exert considerable influence over individuals' access to information, freedom of opinion, expression, and association, and over interlinked political and public interest processes. Moreover, companies like Facebook are increasingly engaged in forms of regulation traditionally ascribed to States, a facet reflected in increased governmental reliance on Facebook to regulate content on media platforms. International organizations and governments have consequently called for enhanced cooperation between the public and private sectors to aid efforts to counter terrorism and violent extremism, "while respecting human rights and fundamental freedoms and complying with international law and the purposes and principles of the Charter."<sup>3</sup>

---

<sup>2</sup> A/HRC/60/288.

<sup>3</sup> See, for example A/RES/72/284.

The growing role of corporate actors and their increased impact on the enjoyment of human rights is addressed by the UN Guiding Principles on Business and Human Rights (UNGPs). These UN Principles provide an authoritative global standard for preventing and addressing adverse human rights impacts linked to business activity.<sup>4</sup> While the UNGPs have been endorsed by the Human Rights Council in Resolution 17/4 of 16 June 2011,<sup>5</sup> they are not formally legally binding. They represent however an important step towards matching the impact of businesses on human rights with corresponding levels of corporate responsibility. Moreover, they represent the direction of legal obligations, as soft law norms that may crystalize to hard law obligation over time and use. They are being recognized, accepted and implemented by a growing number of private companies.

Under the UNGPs the responsibility to respect internationally recognized human rights<sup>6</sup> implies that businesses must ‘[a]void causing or contributing to adverse human rights impacts through their own activities and address such impacts when they occur’ and ‘[s]eek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships’, even if they have not contributed to those impacts.<sup>7</sup>

Corporate responsibility under the UNGPs is independent of State obligations and as such ‘exists over and above compliance with national laws’ and irrespective of States’ abilities and/or willingness to fulfil their own duties under human rights law.<sup>8</sup>

In line with the UNGPs, businesses should adopt an explicit and public policy commitment to meet their responsibility to respect human rights and the commitment should be reflected in operational policies and procedures governing their activities.<sup>9</sup> To aid this process, the Principles recommend identifying the human rights the enterprise’s activities are most likely to have an impact on and effective ways to prevent and/or mitigate such impact.<sup>10</sup> For social media companies, human rights that are at particular risk to be affected through their operations would include the right to privacy and

---

<sup>4</sup> Parallel but complementary to the UNGPs, different interest groups have also developed guidance for the ICT sector, largely aligned with the Guiding Principles in their overall approach and expectations of companies. For example, the multi-stakeholder Global Network Initiative, an initiative that counts Facebook as a member, has adopted Principles on Freedom of Expression and Privacy, addressing some of the most salient human rights risks posed by the activities of ICT companies. See Global Network Initiative, ‘Principles on Freedom of Expression and Privacy’, available at: <http://globalnetworkinitiative.org/principles/index.php>.

<sup>5</sup> A/HRC/17/4.

<sup>6</sup> These are understood to include, at a minimum, the rights expressed in the International Bill of Human Rights and the principles concerning fundamental rights set out in the International Labour Organization’s Declaration on Fundamental Principles and Rights at Work. See UNGPs, Principle 12.

<sup>7</sup> UNGPs, Principle 13.

<sup>8</sup> See Office of the High Commissioner for Human Rights, *The Corporate Responsibility to Respect Human Rights. An Interpretative Guide* (2012), at 13.

<sup>9</sup> UNGPs, Principle 15.

The policy commitment should clearly set out expectations from employees of the company (including management) and their business relationships.

<sup>10</sup> Office of the High Commissioner for Human Rights, *The Corporate Responsibility to Respect Human Rights. An Interpretative Guide* (2012), at 28.

freedom of expression and opinion, encompassing the right to seek, receive and impart information, and interlinked rights such as the freedoms of association, assembly, and of religion or belief, among others. This endeavour should address whether and how business practices and policies relating to regulating access and use of social platforms may result in adverse human rights impact.

### **Definition of terrorism**

Facebook operates in numerous jurisdictions and as such has to comply with a plethora of domestic frameworks that set standards of diverging levels of human rights compliance in law and in practice. Moreover, due to the lack of a universally agreed upon definition of terrorism, these jurisdictions present significant differences in their approach to defining terrorism and terrorism-related offenses. I acknowledge the difficulties Facebook is confronted with in this context and the challenging implications such backdrop poses when it comes to adopting a human rights-based approach to govern its conduct and policies.

Notwithstanding these difficulties, I wish to emphasize the importance of ensuring that any relevant definitions adopted and employed by Facebook are compatible with standards set by international law, including international human rights law and international humanitarian law. In this respect, I wish to highlight that my predecessor on the mandate, Martin Scheinin, has formulated definitions of terrorism and incitement to terrorism. These definitions reflect best practice in countering terrorism, pursuant to an analysis undertaken on the basis of consultations and various forms of interaction with multiple stakeholders, including Governments and have since been affirmed by Mr. Scheinin's successors, including myself.<sup>11</sup>

In consequence, the mandate of the Special Rapporteur advances the following model definition of terrorism:

Terrorism means an action or attempted action where:

1. The action:

- (a) Constituted the intentional taking of hostages; or
- (b) Is intended to cause death or serious bodily injury to one or more members of the general population or segments of it; or
- (c) Involved lethal or serious physical violence against one or more members of the general population or segments of it;

And

The action is done or attempted with the intention of:

- (a) Provoking a state of terror in the general public or a segment of it; or
- (b) Compelling a Government or international organization to do or abstain from doing something;

and

(3) The action corresponds to:

---

<sup>11</sup> A/HRC/16/51.

- (a) The definition of a serious offence in national law, enacted for the purpose of complying with international conventions and protocols relating to terrorism or with resolutions of the Security Council relating to terrorism; or
- (b) All elements of a serious crime defined by national law.<sup>12</sup>

Furthermore, the model definition based on best practice for the offense of terrorist incitement is as follows:

It is an offence to intentionally and unlawfully distribute or otherwise make available a message to the public with the intent to incite the commission of a terrorist offence, where such conduct, whether or not expressly advocating terrorist offences, causes a danger that one or more such offences may be committed.<sup>13</sup>

My predecessors and I have consistently stressed the importance of restricting counter-terrorism measures to conduct that is truly terrorist in nature. In this vein, all counterterrorism laws “must be limited to the countering of offences within the scope of, and as defined in, the international conventions and protocols relating to terrorism, or the countering of associated conduct called for within resolutions of the Security Council, when combined with the intention and purpose elements identified in Security Council resolution 1566 (2001).”<sup>14</sup> Such laws must be adequately accessible and formulated with sufficient precision. In particular, “[c]rimes not having the quality of terrorism (...), regardless of how serious, should not be the subject of counter-terrorist legislation.”<sup>15</sup>

The definition employed by Facebook equates all non-state groups that use violence in pursuit of any goals or ends to terrorist entities.<sup>16</sup> However, as set out above, only a subset of violent acts committed by a non-state actor can be qualified as terrorism. The use of such a sweeping definition is particularly worrying in light of a number of governments seeking to stigmatize diverse forms of dissent and opposition (whether peaceful or violent) as terrorism. The definition is further at odds with international humanitarian law as it qualifies all non-state armed groups party to a non-international armed conflict as terrorists, even if these groups comply with international humanitarian law. Moreover, it is unclear how Facebook determines when a person belongs to a particular group and whether the respective group or person are given the opportunity to meaningfully challenge such determination.

Content produced and/or posted by groups encompassed by Facebook’s definition but that cannot genuinely be characterized as terrorist, may nonetheless be unlawful. In particular, such content may amount to advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. Such content should be

---

<sup>12</sup> A/HRC/16/51, Practice 7.

<sup>13</sup> A/HRC/16/51, Practice 8.

<sup>14</sup> E/CN.4/2006/98, para. 39.

<sup>15</sup> *Ibid.* para. 47.

<sup>16</sup> As spelled out by Facebook, the definition is “agnostic to the ideology or political goals of a group, which means it includes everything from religious extremists and violent separatists to white supremacists and militant environmental groups. It’s about whether they use violence to pursue those goals.”

addressed in line with Articles 20 and 19(3) of the International Covenant on Civil and Political Rights. In conducting relevant assessments, I recommend that Facebook be guided by the standards spelled out in the Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (Rabat Plan of Action).<sup>17</sup>

### **A human rights-compliant approach to providing access to the platform and regulating user content**

Companies functioning on the basis of business models centred around hosting third-party content have been facing mounting pressure from governments to proactively monitor and even police content generated or disseminated by users.<sup>18</sup> This trend is particularly salient in the context of the fight against terrorism.<sup>19</sup> In this respect, a number of states have enacted or are contemplating enacting legislation that may impose onerous obligations on social media platforms. I commend the important role Facebook plays in offsetting terrorist activity online. At the same time, I would like to underscore the importance that such activity is carried out in conformity with Facebook's responsibility to not unduly interfere with the human rights of its users.<sup>20</sup>

In line with the 'respect, protect, remedy' framework set out under the UNGPs, Facebook should provide users with clear guidance on the circumstances under which content may be blocked, removed or restricted or their access to a service restricted or terminated. Policies regulating conduct on the platform must be set out in clear and accessible language and be sufficiently detailed to allow users to foresee the consequences of their use of the platform and have the option to adjust their conduct accordingly. Users should be provided information in this regard prior to entering an agreement with the company and each time the business updates its terms and conditions of service. This information allows individuals to make informed decisions about their use of the services offered by the company. In this respect, I note with concern Facebook's statement that it avoided sharing details of their content moderation policies to discourage users from finding workarounds.<sup>21</sup>

Moreover, as set out above, I warn against the use of overly broad and imprecise definitions<sup>22</sup> as the basis for regulating access to and the use of Facebook's platform as these may lead to indiscriminate implementation, over-censoring and arbitrary denial of access to and use of Facebook's services. In this vein, I stress that removal and blocking efforts should be focused on content of genuinely terrorist nature and/or amounting to

---

<sup>17</sup> A/HRC/22/17/Add.4

<sup>18</sup> A/RES/72/284.

<sup>19</sup> See also United Nations Office on Drugs and Crime, *The use of the Internet for terrorist purposes*. (New York: United Nations, 2012), pp 3-13, available at [https://www.unodc.org/documents/frontpage/Use\\_of\\_Internet\\_for\\_Terrorist\\_Purposes.pdf](https://www.unodc.org/documents/frontpage/Use_of_Internet_for_Terrorist_Purposes.pdf).

<sup>20</sup> In this respect see also the latest report of the Special Rapporteur on the protection and promotion of the right to freedom of expression addressing the regulation of user-generated online content, A/HRC/38/35.

<sup>21</sup> Bickert, M., 'At Facebook we get things wrong – but we take our safety role seriously', *The Guardian*, 22 May 2017, available at <https://www.theguardian.com/commentisfree/2017/may/22/facebook-get-things-wrong-but-safety-role-seriously>.

<sup>22</sup> See also A/HRC/38/35, in particular para. 26.

advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.

Procedures set up to address user violations need to incorporate adequate safeguards against arbitrary or erroneous decisions. Facebook contends that it makes use of both artificial intelligence (AI) and the work of human moderators in this respect. Detailed information on such procedures and the criteria that determine which incidents will be dealt with by AI, human moderators, or both is however not publicly available. Facebook notes that the median removal time for newly uploaded content that it considers not compliant with its terms and conditions was under one minute. This clearly indicates that a significant percentage of content is removed by algorithms. I note that the use of artificial intelligence for such purposes comes with limitations as algorithms frequently have an inadequate understanding of context and many available tools, such as natural language processing algorithms do not have the same reliability rate across different contexts. They have at times also been shown susceptible to amplifying existing biases. Moreover, considering the volume of user content uploaded to Facebook, even the use of algorithms with a very high accuracy rate potentially results in hundreds of thousands of wrong decisions leading to screening that over- or under-inclusive.

I therefore emphasize the importance of integrating the human-in-the-loop principle for meaningful oversight of decisions made by AI tools. This is also consistent with requirements under the UNGPs calling for companies to set up internal accountability mechanisms for the implementation of policies<sup>23</sup> and have processes in place that enable the remediation of adverse human rights impacts that the company caused or contributed to.<sup>24</sup> In this sense, operational-level grievance mechanisms that are accessible, user-friendly and transparent, may be an effective means to ensure access to remedies to stakeholders whose legitimate interests have been infringed upon by the company, including users of the company's products and services.<sup>25</sup> In cases of content restriction or removal, affected users should be informed of the reasons for such measures. Users should also be informed of any available remedies, the right to challenge the removal, blocking or filtering of content or blocking or suspension of user accounts. Removed content and account information should be preserved. This is important both in case blocking, suspension or removal has been made erroneously as well as if such information may be needed as evidence of criminal conduct for a potential investigation. I welcome Facebook's decision to gradually expand its appeal process to cover individual content-related decisions<sup>26</sup> and invites Facebook to do so with due consideration to international human rights standards.

## **Recommendations**

---

<sup>23</sup> UNGPs, Principles 22, 29 and 31.

<sup>24</sup> See Office of the High Commissioner for Human Rights, *The Corporate Responsibility to Respect Human Rights. An Interpretative Guide* (2012), Section III.

<sup>25</sup> In accordance with Principle 31, such grievance mechanisms can be considered effective if they are 1) legitimate, 2) accessible, 3) predictable, 4) equitable, 5) transparent, 6) rights-compatible, 7) a source of continuous learning and 8) based on engagement and dialogue.

<sup>26</sup> <https://newsroom.fb.com/news/2018/04/comprehensive-community-standards/>

As the Special Rapporteur, I urge Facebook to incorporate a human rights approach into its policies, including those related to addressing violations of Facebook's terms and conditions. In this respect, I make the following recommendations:

1. Adopt the model definitions of terrorism and incitement to terrorism advanced and adopted for the past ten years by the mandate of the Special Rapporteur;
2. When handling content potentially amounting to advocacy of national, racial or religious hatred that may constitute incitement to discrimination, hostility or violence, Facebook should be guided by Articles 19(3) and 20 of the International Covenant on Civil and Political Rights and the standards spelled out in the Rabat Plan of Action, in particular the six-part threshold test proposed therein<sup>27</sup>;
3. To the extent needed, Facebook should refer to the Consolidated United Nations Security Council Sanction List<sup>28</sup> as a resource for determining entities that fall within the scope of terrorist groups;
4. Establish clear, detailed, foreseeable and publicly accessible rules and procedures governing access to and use of the platform;
5. Ensure that relevant procedures have adequate safeguards incorporated and respect the human-in-the-loop principle;
6. Ensure that there are easily accessible, user-friendly, non-onerous and transparent procedures for meaningfully challenging decisions taken by Facebook to block or restrict access to the platform as well as user content;
7. Facebook should be guided by the recommendations issued in this respect by the Office of the High Commissioner for Human Rights, the Working Group on the issue of human rights and transnational corporations and other business enterprises as well as the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.<sup>29</sup>

Thank you in advance for the consideration of my views. My mandate would be happy to continue the dialogue on the challenges faced by Facebook in the context of countering terrorist content on its platform.

Finally, I would like to inform you that this communication may be made available to the public and posted on the website page for the mandate of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, together with any response that you may consider sharing with me.<sup>30</sup>

Please accept the assurances of my highest consideration.

Fionnuala Ní Aoláin

---

<sup>27</sup> A/HRC/22/17/Add.4, para. 29.

<sup>28</sup> <https://www.un.org/sc/suborg/en/sanctions/un-sc-consolidated-list>

<sup>29</sup> See, in particular, A/HRC/38/35, A/HRC/35/22 and A/HRC/32/38.

<sup>30</sup> <https://www.ohchr.org/EN/Issues/Terrorism/Pages/SRTerrorismIndex.aspx>



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