

To: Facebook Oversight Board

From: Centre for Democracy and Rule of Law, Ukraine

contact email: i.rozkladaj@cedem.org.ua (*Igor Rozkladaj*)

Independent Media Council, Ukraine

contact email: antonina.cherevko@gmail.com (*Antonina Cherevko*)

07 December 2020

Public consultation provided on the case 2020-003-FB-UA

Dear Colleagues,

Centre for Democracy and Rule of Law,[1] Ukraine (hereafter referred to as CEDEM) and the Independent Media Council ,[2] Ukraine (hereafter referred to as IMC) would like to use this opportunity to intervene in the OB case 2020-003-FB-UA.

Given that the OB is, among other functions, entitled with providing Facebook with the recommendations regarding its community rules and policies, we would like to use the above mentioned case to bring attention of the OB and Facebook to the following considerations which should be taken into account while deciding on removal of posts or any other moderation of online expression made in the course of international or non-international armed conflicts (IAC and NIAC respectively):

- While assessing any Facebook post or online expression related to ongoing conflict, it is important to differentiate between IAC and NIAC in terms of likeliness of harm inflicted on the community/nation/individuals targeted by such expression. In both IAC and NIAC heightened emotions, mutual accusations (sometimes with references to real or pseudo-historical facts and experiences) and aggressive tone characterise almost any communication between the conflicting parties, verbal insult becomes a “new normal” in a conflict situation. In IAC such communication in most cases does not produce an effect of direct incitement in civilian environments, rather it serves as a possibility for the conflicting nations to let off some steam or to express long-standing grievances online. The real harm in IAC is mostly inflicted at the battlefield rather than within civilian spaces. In NIAC, however, one of the parties is usually a minority group, often already marginalised and vulnerable, co-existing in joint space with the hostile majority, and harmful speech directed at such a group can have

serious consequences even in civilian spaces, outside of the battlefield, often subjecting those who do not take direct part in hostilities to abuse and violence.

- Due consideration should be given to a proper legal qualification of speech in terms of differentiation between “ordinary” hate speech (which can be expressed in the form of harsh political criticism, provocative artistic speech or controversial historical debates); incitement to discrimination, hostility or violence; and incitement to genocide, war crimes or crimes against humanity as the gravest form of hate speech. This graduated approach is important in assessing cases in IAC or NIAC settings. In IAC it is exactly the gravest form of hate speech – incitement to genocide, crimes against humanity or war crimes – that should be immediately addressed and moderated while less serious forms of harmful speech may not require such an urgent and decisive action. In NIAC, for the reasons of its internality and inherent difficulties in differentiation between combatants and civilian population, the threshold of harm should be lower since even milder forms of hate speech can result in actual violence against civilian population, especially where one of the conflicting parties is related to a minority group.

- Another aspect to be considered is a widespread tendency of weaponization of information in the course of the current conflicts, including various information operations. The ongoing IAC between Russia and Ukraine can provide plenty of examples. In this regard, complaints mechanisms on Facebook or other social media are weaponized too. Massive complaints against certain posts are sometimes initiated by bots and trolls rather than real people, and removal of posts of the opposing party to the conflict becomes a kind of “online victory” for the complainants. Pushing for the right narrative both internally and internationally and suppressing the narratives of the opponents is an indispensable element of modern conflicts.

- The most important success factor in achieving human rights compliant, reasonable and fair moderation of online expression on Facebook in terms of IAC or NIAC is employment of both knowledgeable and impartial moderators. While moderators should have sufficient knowledge about the local contexts, history, languages and specificities of a conflict situation, they should not be in any possible way affiliated with the conflicting parties: either politically or because of nationality/citizenship(s) (including previous citizenships), ethnicity or residence.

- Applying the aforementioned considerations to the present case, we would like to underline that people should be provided with a space to express their grievances, especially in the extremely tense social environment, such as a long-lasting IAC. Therefore, if impugned expression cannot reasonably lead to a serious violent consequence for civilians outside of the battlefield (as we can suggest in the present case), –such expression should not be qualified as a prohibited one. More specifically,

if an expression of harsh criticism towards cultural policies or socio-political decisions cannot directly incite civilian individuals to engage in violent actions against the group targeted by this expression, – it can hardly be interpreted as a serious form of online hate speech requiring urgent removal. In particular, political criticism normally warrants a higher degree of protection against moderation.

We are convinced that the points outlined above together with other possible related considerations justify elaboration (in consultations with civil society and legal experts) of a specific set of community standards/moderation rules covering online expression in IAC and NIAC. Both CEDEM and IMC would be glad to offer their assistance in developing such standards/rules.

Sincerely yours,

Olesia Kholopik, Director, CEDEM

Igor Rozkladai, Deputy Director, CEDEM

Antonina Cherevko, Head of IMC

[1] Centre for Democracy and Rule of Law is a civil society organization, set up in 2005 and based in Kyiv, Ukraine. It is an active think-tank pooling together media lawyers and experts specializing in media policy and human rights protection. CEDEM's mission is to encourage development of the society of active and responsible citizens through support to civic initiatives; impartial legal expertise; drafting and implementation of high-quality legislation; legal and civic education; protection of the freedom of speech and fundamental human rights. <https://cedem.org.ua/>

[2] Independent Media Council is an expert civil society initiative in Ukraine which provides legal and ethical assessment of the controversial cases in the Ukrainian media sector and aims at developing media self-regulation standards and practices in the country. <https://mediarada.org.ua/>

Кому: Наглядова рада Facebook

Від: Центр демократії та верховенства права, Україна

контактна електронна адреса: i.rozkladaj@cedem.org.ua (*Ігор Розкладай*)

Незалежна медійна рада, Україна

контактна електронна адреса: antonina.cherevko@gmail.com (*Антоніна Черевко*)

7 грудня 2020 року

Консультація громадськості у справі 2020-003-FB-UA

Шановні Колеги,

Центр демократії та верховенства права, [1] Україна (надалі - "ЦЕДЕМ") та Незалежна медійна рада, [2] Україна (надалі - "НМР") хотіли б скористатися цією нагодою, щоб долучитися до обговорення у справі Наглядової ради 2020-003-FB-UA.

Зважаючи на те, що серед інших функцій Наглядова рада має право надавати Facebook рекомендації щодо правил та політик спільноти, крізь призму вищезокресленої справи ми хотіли б звернути увагу Наглядової ради та Facebook на наступні міркування, на які зважати, приймаючи рішення про видалення постів чи будь-яку іншу модерацію онлайн-висловлювань у контексті збройних конфліктів міжнародного чи неміжнародного характеру (надалі - "ІАС" та "NІАС"):

- Оцінюючи будь-які публікації в Facebook чи онлайн-висловлювання стосовно триваючого конфлікту, важливо розрізняти ІАС та NІАС з точки зору ймовірності заподіяння шкоди громаді / нації / окремим особам, на яких спрямовані ці висловлювання. Як в ситуації з ІАС так і з NІАС, присутні сильні емоції, взаємні звинувачення (іноді з посиланнями на реальні чи псевдоісторичні факти та досвід) і агресивний тон, які є характерними для майже будь-якого спілкування між конфліктуючими сторонами, а словесна образа стає "новою нормальністю" в ситуації конфлікту. У контексті ІАС таке спілкування здебільшого не призводить до прямого підбурювання в громадянському середовищі, воно швидше є можливістю для конфліктуючих країн "випустити пару" чи висловити давні образи онлайн. Справжня шкода в контексті ІАС здебільшого завдається на полі бою, а не в цивільному середовищі. Однак у ситуації з NІАС однією із сторін

зазвичай є меншина, часто вже маргіналізована та вразлива, яка співіснує в спільному просторі з ворожою більшістю, і ненависницькі висловлювання в бік такої групи осіб можуть мати серйозні наслідки навіть у цивільному середовищі, поза полем бою, часто наражаючи на жорстоке поводження та насильство тих, хто не бере безпосередньої участі в бойових діях.

- Особливу увагу слід приділяти належній кваліфікації висловлювань з юридичної точки зору, розрізняючи “звичайну” мову ворожнечі (яка може набувати форми різкої політичної критики, провокативної художньої мови чи контроверсійних історичних суперечок); підбурювання до дискримінації, ворожості або насильства; і підбурювання до геноциду, військових злочинів чи злочинів проти людства як найтяжчої форми мови ворожнечі. Такий градуйований підхід є важливим для оцінки справ у контексті ІАС чи NІАС. В умовах ІАС саме найтяжча форма мови ворожнечі - підбурювання до геноциду, злочинів проти людства або військових злочинів - повинна негайно розглядатися та модеруватися, тоді як менш серйозні форми ворожих висловлювань можуть вимагати не таких термінових і рішучих дій. У ситуації з NІАС, через його внутрішній характер та неминучі труднощі в розрізненні учасників бойових дій та цивільного населення, поріг шкоди має бути нижчим, оскільки навіть більш м'які форми мови ворожнечі можуть призвести до актів насильства проти цивільного населення - особливо там, де одна з конфліктуючих сторін відноситься до групи меншин.

- Ще одним аспектом, який слід врахувати, є поширена тенденція вепонізації інформації під час триваючих конфліктів, зокрема різні інформаційні операції ІАС, що триває між Росією та Україною може надати безліч прикладів. У цьому сенсі механізми подання скарг у Facebook чи інших соціальних мережах також вепонізовано. Масові скарги на певні пости іноді ініціюються ботами й троями, а не реальними людьми, і видалення постів протилежної сторони конфлікту стає своєрідною "онлайн-перемогою" для скаргників. Проштовхування правильного нарративу на внутрішньому і на міжнародному рівні, та придушення нарративів опонентів є незамінним елементом сучасних конфліктів.

- Найважливішим фактором успіху в забезпеченні розумного та справедливого, з дотриманням прав людини, модерування онлайн-висловлювань на Facebook в контексті ІАС чи NІАС є залучення обізнаних і, водночас, неупереджених модераторів. Хоча модератори повинні мати достатні знання про місцевий контекст, історію, мови та особливості конфліктної ситуації, втім вони жодним чином не мають бути пов'язані з конфліктуючими сторонами: ні політично, ні через національність / громадянство(а) (зокрема попереднє громадянство), етнічну приналежність чи місце проживання.

- Застосовуючи вищезазначені міркування до даної справи, ми хотіли б підкреслити, що людям слід надати простір для викладення своїх наболілих думок, особливо в надзвичайно напруженому соціальному середовищі, такому як довготривалий ІАС. Тож, якщо оскаржувані висловлювання загалом не можуть призвести до тяжких наслідків для цивільних осіб поза полем бою (що, на нашу думку, має місце в даній справі), - такі висловлення думок не слід кваліфікувати як заборонені. Говорячи більш конкретно, якщо різкі критичні висловлювання щодо культурних політик чи соціально-політичних рішень не можуть безпосередньо підбурити цивільних осіб до насильницьких дій проти групи осіб, на яку спрямовані ці критичні висловлювання, це навряд чи можна трактувати як серйозну онлайн-форму мови ненависті, що вимагає термінового видалення. Зокрема, політична критика зазвичай вимагає вищого рівня захисту від модерації.

Переконані, що вищезазначені пункти разом з іншими можливими супутніми міркуваннями слугують підставою для розробки (в консультаціях з громадянським суспільством та експертами-юристами) особливого набору стандартів / правил модерації спільноти, що охоплюють онлайн-висловлювання в контексті ІАС та NІАС. І ЦЕДЕМ, і НМР із задоволенням запропонують свою допомогу в розробці таких стандартів / правил.

Щиро Ваші,

Олеся Холопик, директор, ЦЕДЕМ

Ігор Розкладай, заступник директора, ЦЕДЕМ

Антоніна Черевко, голова НМР

[1] Центр демократії та верховенства права є організацією громадянського суспільства, створеною у 2005 році, що базується в Києві, Україна. Це активний аналітично-адвокатський центр, який об'єднує медіа-юристів та експертів з питань медійної політики та захисту прав людини. Місія CEDEM - сприяти розвитку суспільства активних та відповідальних громадян шляхом підтримки громадських ініціатив; неупереджена юридична експертиза; розробка та впровадження високоякісного законодавства; правова та громадянська освіта; захист свободи слова та основних прав людини. <https://cedem.org.ua/>

[2] Незалежна медійна рада є експертною ініціативою громадянського суспільства в Україні, яка надає правову та етичну оцінку контроверсійним справам в українському медіа-секторі та має на меті розробку стандартів та практик саморегулювання у сфері ЗМІ в країні. <https://mediarada.org.ua/>