SPECIAL COLLECTION OF THE CASE LAW ON FREEDOM OF EXPRESSION

The Decisions of the Oversight Board from the Perspective of International Human Rights Law

Joan Barata

Intermediary Liability Fellow

Program on Platform Regulation - Cyber Policy center

Stanford University

Global Freedom of Expression
COLUMBIA UNIVERSITY

Global Freedom of Expression
COLUMBIA UNIVERSITY

The Decisions of the Oversight Board from the Perspective of International Human Rights Law

Joan Barata

Intermediary Liability Fellow

Program on Platform Regulation - Cyber Policy center

Stanford University

I. OVERVIEW

1. Introduction

The Oversight Board (OSB) started its operations in 2020, as an initiative of the technology company Facebook (now named Meta, although the brand Facebook still designates one of its main social media platforms).

According to its own website, the OSB was created to help "answer some of the most difficult questions around freedom of expression online: what to take down, what to leave up and why"[1]. The four main principles of the Board are: independence (it is separated from the company and provides independent judgment), empowerment (it has authority to adopt binding decisions regarding allowing or removing content), accessibility (individuals will be able to appeal Facebook and Instagram content decisions to the board and to share a statement explaining their position), and transparency (the OSB publicly shares written statements about its decisions and rationale).

Fundamental attributes of the Board (article 4 of the Charter[2]) include the capacity to have a final say on whether material will be restored to the platform or not, or whether to confirm a decision to leave up a piece of content. The Board can also make policy recommendations for changes in the way that the company operates its community standards and practices. Besides these main powers, the OSB also counts on several instrumental competences and responsibilities, such as requesting information from Meta and issuing prompt and written explanations of the adopted decisions. It is important to note that this explanation will be made public within the context of the Board's decisions. This created, for the first time, a "normative" obligation for an online platform to provide public explanations regarding its content policies.

The four main principles of the Board are: independence, empowerment, accessibility and transparency.

A very important prerogative included in the Charter gives the Board the discretion to choose which requests it reviews and decides upon. This selection has to be guided by the need to consider cases that "have the greatest potential to guide future decisions and policies." This notion of "potential" is in any case particularly open to interpretation and is constrained by the need that cases available for review are only those submitted by users (either by the original poster of the content or by a person who had previously submitted the content to Meta for review) or by the company itself.

Article 2.2 of OSB Charter establishes the basis of the decision making of this body. According to this provision, decisions of the Board will focus on "content enforcement decisions [by Facebook or Instagram] and determine whether they were consistent with Facebook's content policies and values". Therefore, the main task of the Board is to assess consistency between content deci-

1 https://www.oversightboard.com

2 https://www.oversightboard.com/governance/

sions taken by the respective platform and their own internal (private) principles and rules. The last paragraph of the mentioned article also establishes that the Board "will pay particular attention to the impact of removing content in light of human rights norms protecting free expression". This reference to a human rights framework is connected to commitments in this field explicitly made by Meta, particularly since the adoption of its Corporate Human Rights Policy in 2021, as will be further explained in this paper.

With regards to this last particular reference, it is important to note that in his 2018 thematic report to the Human Rights Council[3], the United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and freedom of expression directly addressed platforms, requesting them to recognize that "the authoritative global standard for ensuring freedom of expression on their platforms is human rights law, not the varying laws of States or their own private interests, and they should re-evaluate their content standards accordingly". Therefore, independently from the detailed rules and standards that they may have in place (and which would not necessarily equate to the myriad of national legal norms governing speech), platforms would at least need to adhere to and use universal human rights law as a general guiding and interpretative framework for the establishment and enforcement of such norms.

In a letter sent to Mark Zuckerberg on 1 May 2019, the Special Rapporteur particularly welcomes the creation of the OSB and stresses the fact that international human rights law would provide the Board "with a set of tools and a common vocabulary for addressing and resolving hard questions around the moderation of online content".

There are, however, relevant challenges when it comes to determining the specific implications of requiring private corporations to adhere to human rights provisions. Needless to say, international human rights law is originally designed to govern the relationship of State authorities with individuals and groups. In 2011, however, the UN Human Rights Council adopted the UN Guiding Principles on Business and Human Rights (UNGPs)[4]. Despite their non-binding nature, they are currently seen as the most important and developed instrument to frame private corporations conduct vis-à-vis human rights (including protection, respect, and remedy of possible abuses).

While acknowledging the contributions that international human rights law can make to content moderation, Evelyn Douek has also warned about the actual limits of such a set of norms as a practical guide to what platforms should do in many difficult cases[5]. More specifically, she advocates for the creation of the institutions "necessary to ensure that [international human rights law] in content moderation serves the interests of users and society rather than being co-opted by

3 A/HRC/38/35. Available online at: https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx

4 Available online at: https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

5 See Evelyn Douek, "The Limits of International Law in Content Moderation", 6 UC Irvine Journal of International, Transnational, and Comparative Law 37 (2021).

platforms to their own ends". From a slightly different perspective, Jacob Mchangama, Natalie Alkiviadou and Raghav Mendiratta see human rights law as "a framework of first reference" and have articulated specific proposals on how a human rights approach may be implemented by online platforms to bring about rights-protective and transparent content moderation, particularly focusing on hate speech and disinformation. In addition, they recommend that major platforms formally commit to adopting an international human rights approach to content moderation by signing a non-binding Free Speech Framework Agreement (FSFA) administered by the Office of the UN High Commissioner for Human Rights (OHCHR) under the specific auspices of the Special Rapporteur on Freedom of Opinion and Expression[6].

Regarding possible intersections between legality and content review by the Board, article 1.2.1 of the Bylaws establishes that decisions on intellectual property or pursuant to legal obligations are not available for the former. In addition to this, and according to article 1.2.2, cases where the underlying content has already been blocked, following the sole receipt of a valid report of illegality, cases where the underlying content is criminally unlawful in a jurisdiction with a connection to the content and where a Board decision to allow the content on the platform could lead to criminal liability for Facebook, Facebook employees, the administration, or the Board's members, as well as cases where the underlying content is unlawful in a jurisdiction with a connection to the content and where a Board decision to allow the content on the platform could lead to adverse governmental action, are not eligible for the OSB to review.

In the already mentioned letter of the UN Special Rapporteur to Meta's CEO there is also a reference to this special limitation in the work of the OSB. The Special Rapporteur acknowledges that the latter is a private entity with no power or legitimacy to decide on the legality, according to national legislation, of a certain piece of content. This being said, he also recommends that the OSB plays a role in minimizing adverse human rights impacts in connection with State demands or requests. Considering Meta's responsibility vis-a-vis resolving any legal ambiguity in favor of respect for freedom of expression and other human rights, challenging overbroad or presumably illegal requests before the courts and disclosing the maximum possible amount of information concerning Governments' requests[7], the Special Rapporteur sees independent and external assessment from the Board regarding consistency of State demands with international standards as a potential tool to assist in the implementation of the company's commitment to international human rights law.

---

6 Jacob Mchangama, Natalie Alkiviadou, Raghav Mendiratta, A framework of first reference. Decoding a human rights approach to content moderation in the era of "platformization", Justitia, 2021. Available online at: http://justitia-int.org/wp-content/uploads/2021/11/Report_A-framework-o__f-first-reference.pdf

7 In particular, Meta's Corporate Human Rights Policy (as adopted in 2021), declares that: "We recognize the diversity of laws in the locations where we operate, and where people use our products. We strive to respect domestic laws. When faced with conflicts between such laws and our human rights commitments, we seek to honor the principles of internationally recognized human rights to the greatest extent possible. In these circumstances we seek to promote international human rights standards by engaging with governments, and by collaborating with other stakeholders and companies."

The object of this paper is to present a general overview of the decisions adopted by the OSB since its creation, particularly regarding the use of international human rights to interpret the meaning and scope of Meta's products' values and community standards.

2. Structure of the decisions of the OSB

Decisions adopted by the OSB are accessible, in full, from the body's website[8].

Up to the date of finalization of this paper, the OSB has issued 28 case decisions.

Every published decision includes a separated summary that provides with a general overview of the case, a description of the key findings and the content of the decision of the Board. This brief usually includes, as such, significant information regarding the most relevant aspects of the case and the decision.

The full decision has the following sections:

a) Decision summary.
b) Case description.
c) Authority and scope.
d) Relevant standards, including platform's content policies and values, and human rights standards.
e) Content creator/user statement.
f) Meta explanation of its own decision.
g) Third-party submissions.
h) OSB analysis, based on compliance with content policies, compliances with Meta's values, and compliance with Meta's human rights responsibilities.
i) OSB decision.
j) Possible policy advisory statements.

In light of the mentioned of the decisions, a few conclusions can be presented.

Firstly, when presenting and considering the facts of a case, the OSB does not only investigate the questioned content, its author and other contextual elements directly related to the post (comments, impact, etc.), but also other elements and circumstances taking place "outside" the platform. In this sense, decision 2021-01 on the *de-platformization* of US President Donald Trump pays considerable attention to events and statements not directly related to the posts under analysis, although definitely relevant in terms of context for the decision.

8 https://www.oversightboard.com/decision/

Secondly, the possibility of receiving third-party submissions has given NGOs, academic centers, and individuals the opportunity of submitting thousands of comments related to the different cases. Comments submitted are not incorporated into the decisions' text, although in some cases they are briefly mentioned. In any case, they are available to the public in a separate and accessible document.

Thirdly, in order to incorporate additional parameters, the Board also commissions, in many cases, independent research from academic centers with the capacity to draw scientific and expert teams from different geographical areas. As mentioned, the aim of such contributions is to provide additional information regarding the political, social, and cultural context to be considered when assessing the application of a certain concept or understanding of specific circumstances of a case. For example, in its decision 2021-04 the Board commissioned a wide research work to properly frame and discern the notions of bullying and harassment.

Fourthly, decisions briefly present in some cases the internal debates and criteria taken into consideration by different members of the Board as part of the process of reaching a final *judgement*. More specifically, despite the decisions not incorporating or directly reflecting "dissenting opinions", they occasionally reveal the differences between the point of view of a "majority" versus the position adopted by a "minority" or "other members". In the Decision 2021-05, which overturned a removal decision by Facebook, while supporting the majority's views on protecting satire on the platform, the minority did not believe that the piece of content under analysis (a meme/cartoon referring to the Armenian genocide) was a satire criticizing the Turkish government. The minority found that the user could be embracing the statements contained in the meme and thus engaging in discrimination against Armenians. Therefore, the minority held that the requirements of necessity and proportionality had been met and the post had been correctly removed.

The decisions briefly present in some cases the internal debates and criteria taken into consideration by different members of the Board as part of the process of reaching a *final judgement*.

It is also important to note that in some cases these differences do not necessarily refer to the basic appreciation of the facts or the final decision adopted by the Board, but to circumstantial elements and criteria considered during the process. For example, in the decision 2021-08, the Board endorsed Facebook's decision to leave up a post by a state-level medical council in Brazil, which claimed that lockdowns are ineffective, finding it consistent with its content policies. The Board noted that the content, in this case, was not used as a basis by the council for the adoption of public health measures that could create risks, as the council does not have authority to decide on these matters. Therefore, the whole body supported the idea that Facebook's decision to keep the content on the platform was justified, given that the threshold of imminent physical harm was not met. However, while the majority understood that "public authorities have a duty to verify information they provide to the public even when the false information disseminated is not directly related to its

statutory duties", a minority sustained that "despite the statement contained some inaccurate information, as a whole, it consisted of a fact-related opinion which is legitimate in public discussion".

Lastly, decisions also reflect the exchanges between the Board and Meta, particularly with regards to the provision of information relevant to the case by the later. It is remarkable to note how, on some occasions, the Board expresses a certain degree of frustration regarding the lack of proper collaboration and engagement by the platform in this area. In the context of the decision 2021-09 regarding the republication of a news item from Al-Jazeera, the Board considered it necessary to react to allegations that Facebook has censored Palestinian content due to the Israeli government's demands. Therefore, the OSB asked Facebook:

> "Has Facebook received official and unofficial requests from Israel to take down content related to the April-May conflict? How many requests has Facebook received? How many has it complied with? Did any requests concern information posted by Al Jazeera Arabic or its journalists?"

Facebook responded:

> "Facebook has not received a valid legal request from a government authority related to the content the user posted in this case. Facebook declines to provide the remaining requested information. See Oversight Board Bylaws, Section 2.2.2".

The Board finally concludes that the company did not indicate the specific reasons for the refusal under the Bylaws.

3. Use of international human rights standards

3.1 *Universal human rights standards included in the decisions*

As it was already mentioned, the OSB Charter establishes that the Board has authority to adopt binding decisions as to allowing or removing content based on consistency with Meta's content policies and values while paying particular attention to "the impact of removing content in light of human rights norms protecting free expression".

OSB decisions indeed reflect and take into consideration relevant human rights standards applicable to each specific case.

OSB decisions base their human rights assessment, firstly, on the main conceptual frame provided by the already mentioned UNGPs. It is also important to note that in March 2021, Meta

announced its Corporate Human Rights Policy[9], where it recommitted to respecting human rights in accordance with the UNGPs. This commitment encompasses internationally recognized human rights as defined by the Universal Declaration of Human Rights; the International Covenant on Civil and Political Rights; and the International Covenant on Economic, Social and Cultural Rights, as well as the International Labor Organization Declaration on Fundamental Principles and Rights at Work. The company also commits to utilize "depending on the circumstances" other widely accepted international legal instruments, including the International Convention on the Elimination of All Forms of Racial Discrimination; the Convention on the Elimination of All Forms of Discrimination Against Women; the Convention on the Rights of the Child; the Convention on the Rights of Persons with Disabilities; the Charter of Fundamental Rights of the European Union; and the American Convention on Human Rights. It is important to note, however, that this is a declaration made by Meta as a corporation, which does not necessarily need to be seen as prescriptive or limitative for the OSB in terms of the international standards to consider.

Regarding freedom of expression, decisions usually refer to the most relevant principles, rules and standards deriving from the universal human rights system. These include both international human rights legal provisions as well as standards falling under the category of *soft law*.

The OSB usually takes as the main direct legal reference regarding freedom of expression the provisions included in the International Covenant on Civil and Political Rights (ICCPR): article 19 (freedom of expression) and article 20 (propaganda for war and hate speech). These articles are particularly considered based on their most relevant and authoritative interpretation criteria, provided by General Comment number 34 of the UN Human Rights Committee[10]. In addition to this, approaches to the human right to freedom of expression also take in special consideration the standards contained in the different reports and other documents elaborated by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and freedom of expression, as well as the joint declarations made by this body together with the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. Lastly, the Board also incorporates as part of their assessment criteria regarding the specific right to freedom of expression other relevant soft law instruments and documents. It is perhaps worth mentioning that when it comes to hate speech matters, the OSB has been using the important and widely recognised standards established within the context of the so-called Rabat Plan of Action, elaborated under the auspices of the Office of the UN Human Rights Commissioner and endorsed by the UN Human Rights Council Resolution 16/18 of 12 April 2011[11].

9 Available online at: https://about.fb.com/wp-content/uploads/2021/03/Facebooks-Corporate-Human-Rights-Policy.pdf

10 Available online at: https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf

11 A/HRC/RES/16/18. Available online at: https://www2.ohchr.org/english/bodies/hrcouncil/docs/16session/a.hrc.res.16.18_en.pdf

Besides the right to freedom of expression, other human rights protected under the ICCPR are also mentioned in the decisions: right to non-discrimination (articles 2 and 26), right to effective remedy (article 2), right to life (article 6), right to security of person (article 9), right to be informed in the context of access to justice (article 14), right to privacy (article 17), right to peaceful assembly (article 21), and the right to participation in public affairs and to vote (article 25). The content and scope of these rights is usually considered in light of the interpretative criteria provided by the respective General Comments adopted by the UN Human Rights Committee.

OSB decisions also include references to another pillar of international human rights law, the International Covenant on Economic, Social and Cultural Rights (ICESCR), including the right to physical and mental health (article 12 and General Comment No. 14 of the Committee on Economic, Social and Cultural Rights), and the right to take part in cultural life (article 15), among others.

These general provisions included in the so-called International Bill of Human Rights are presented in connection with the documents elaborated by the already mentioned Committees on Human Rights and on Economic, Social and Cultural Rights, as well as the Human Rights Council. Reports and documents from other relevant Committees and consultative bodies and experts are also considered, such as those from the Committee on the Elimination of Racial Discrimination, the Committee on the Rights of the Child, the Working Group of Experts on People of African Descent, the Independent Expert on Minority Issues, the UN Special Rapporteur on Minority Issues, the UN Special Rapporteur on the situation of human rights in the Palestinian territories occupied since 1967, the UN Independent Expert on protection against violence and discrimination based on sexual orientation and gender identity, the UN Special Rapporteur on freedom of peaceful assembly and of association, and the UN Special Rapporteur in the field of cultural rights.

Besides the right to freedom of expression, other human rights protected under the ICCPR are also mentioned in the decisions.

Aside from the mentioned human rights instruments, OSB decisions also incorporate references to other international treaties and conventions, such as those mentioned by Meta's Corporate Human Rights Policy, as well as other texts that refer to issues particular and specific to the case in question; for example the UN Convention on Psychotropic Substances and the UN Convention Against Illicit Traffic in Narcotic Drugs and Psychotropic Substances within the context of a case involving a post describing in a positive way traditional or religious uses of non-medical drugs such as ayahuasca (decision 2021-13). Additional international documents considered also include the UN Declaration on the Rights of Indigenous People.

The OSB furthermore uses an extensive range of international legal instruments and standards to frame a proper understanding and application of the different rights, principles and values directly or indirectly affected by each of the cases. It also uses the language and approach that can commonly be found in decisions and reports elaborated by a wide range of international commit-

tees, councils, and consultative bodies, as well as international human rights courts such as the European Court of Human Rights (ECtHR), the Inter-American Court of Human Rights (IACtHR) and the African Court on Human and People's Rights (ACHPR).

The most relevant example of the latter would be the application of the so-called three-part test to assess the imposition of possible limitations to the right to freedom of expression. This test requires that: 1) any interference must be provided by law, 2) the interference must pursue a legitimate aim included in such provision, and 3) the restriction must be strictly needed, within the context of a democratic society, in order to adequately protect one of those aims, according to the idea of proportionality. This matter will be further analyzed in the next section.

It is important to underscore that, independently from the region where cases originate, the approach of the OSB is almost exclusively based on universal standards. In other words, the existence of specific instruments, standards, or jurisprudence at the regional level (which would be particularly the case of Europe, Africa, and the Americas) is not considered by the Board, thus using a legal approach that is consistent across all the decisions. There are, however, a few exceptions to mention.

In the already brought up case about the positive approach to ayahuasca consumption, some members of the Board use a comparative reference to a decision of the ECtHR (case of *Animal Defenders International v the United Kingdom*[12]) to sustain that a broad allowance for "traditional and religious" drugs would not be administrable and would likely be subject to users attempting to "game" the system[13]. Important to note that the post was published by a user in Brazil. In the Decision 2021-05, regarding a post containing a cartoon on the Armenian genocide, the Board refers to the decision of the ECtHR in the case of *Dink v Turkey* to illustrate the assertion that Turkish authorities have specifically targeted expression denouncing the atrocities committed by the Turkish Ottoman Empire against Armenians from 1915 onwards.

In any case, it must be concluded that regional jurisprudence has not been so far utilized by the Board to sustain the basis and the central arguments of their decisions but to incorporate supportive criteria in terms of context assessment.

3. 2 Human rights responsibilities, Community Standards and Values

It has already been explained that a wide range of human rights legal instruments and standards are included in the decisions of the Board. As it will be further elaborated in the next section, these parameters are considered in particular detail and utilized by the OSB in order to frame their assessment of each case and determine whether a specific piece of content must be left up or removed.

12 Animal Defenders International v. the United Kingdom. Application no 48876/08. Judgment of 22 April 2013.

13 In this case, Meta had justified their removal decision before the Board by referring to judgements from the Supreme Court of the Netherlands and the ECtHR.

In this context, another set of principles and standards also governs Meta's decision and their review by the OSB: Community Standards and Values, which represent Meta's internal and fundamental content policies. The decisions of the Board articulate a special relationship or dialogue between these norms and human rights principles.

Firstly, we can find cases where a divergence between Community Standards and human rights criteria prompts the Board to establish the need to amend the former to adapt them to the latter. In the decision 2020-06, regarding a post that criticized the lack of a health strategy in France and stated that hydroxychloroquine combined with azithromycin was being used elsewhere to save lives, the Board finds Facebook's misinformation and imminent harm rule, which this post is said to have violated, to be "inappropriately vague and inconsistent with international human rights standards". Once again regarding the decision 2021-13, the Board agrees with the company that the content (the already mentioned ayahuasca post) violates the Facebook Regulated Goods Community Standard, as incorporated by reference in the Instagram Guidelines. However, it also concludes that although the content violates the Regulated Goods Community Standard, "Meta's values and international human rights standards support the Board's decision to restore the content". The Board also makes policy recommendations to bring the Community Standard in line with Meta's values and international human rights standards. Therefore, the Board does not only use human rights standards as authoritative and prevailing assessment criteria, but also refers to the general "Values" that the company also commits to use in its content decisions.

The Board does not only use human rights standards as authoritative and prevailing assessment criteria, but also refers to the general "Values" that the company also commits to use in its content decisions.

Meta's values include the notions of "Voice", "Safety", and "Dignity". Obviously, they do not point in the same directions. However, the Board balances them in each decision and precisely uses human rights criteria to decide which one outweighs the other in the respective context. In the decision 2021-02, on the publication in the Netherlands of caricatures of Black people in the form of blackface, the Board notes that Facebook has explicitly prohibited this type of content as part of its Hate Speech Community Standard. However, while the majority argued that such caricatures "are inextricably linked to negative and racist stereotypes and are considered by parts of Dutch society to sustain systemic racism in the Netherlands", a minority of the Board saw insufficient evidence to directly link this piece of content to the harm supposedly being reduced by removing it. They noted that "Facebook's value of 'Voice' specifically protects disagreeable content and that, while blackface is offensive, depictions on Facebook will not always cause harm to others".

In the decision 2021-14 on a post in Ethiopia containing allegations that the Tigray People's Liberation Front (TPLF) killed and raped women and children and looted the properties of civilians in Raya Kobo and other towns in Ethiopia's Amhara region, the Board found that "removing the post was consistent with Meta's human rights responsibilities as a business" and that "unver-

ifiable rumours in a heated and ongoing conflict could lead to grave atrocities, as was the case in Myanmar". However, the Board recommends that Meta rewrites its value of "Safety" to reflect that "online speech may pose risk to the physical security of persons and the right to life, in addition to the risks of intimidation, exclusion and silencing"; as well as reflect in the Facebook Community Standards that "in the contexts of war and violent conflict, unverified rumours pose higher risk to the rights of life and security of persons". In its decision 2020-003, the Board already stated that Meta must not neglect the fact that "in situations of armed conflict in particular, the risk of hateful, dehumanising expressions accumulating and spreading on a platform, leading to offline action impacting the right to security of person and potentially life, is especially pronounced".

Secondly, there are cases where the Board uses human rights law to refine and frame the way compliance with Community Standards must be interpreted by Meta. In the decision 2021-01 on President Trump, the Board does not only agree with Facebook's decision that the two posts by Mr. Trump on 6 January violated Facebook's Community Standards and Instagram's Community Guidelines on dangerous individuals and organizations, but also understands that other Community Standards may have been violated in this case, including the Standard on violence and incitement. In this case, a significant component of the Board's rationale is based upon the six factors from the Rabat Plan of Action to assess the capacity of speech to create a serious risk of inciting discrimination, violence, or other lawless action (context, status of the speaker, intent, content and form, extent and reach, and imminence of harm).

The decision 2021-06 refers to a post in Turkey encouraging people to discuss the solitary confinement of Abdullah Öcalan. After the user appealed the removal decision and the Board selected the case for review, Facebook concluded that the content was removed in error and restored it. A piece of internal guidance elaborated in 2017 allows discussion on the conditions of confinement for individuals designated as dangerous. The Board expresses concern that "Facebook misplaced this exception for three years and that this may have led to many other posts being wrongly removed". However, it also underscores that even without the discovery of the misplaced guidance, the content should never have been removed. The user did not advocate violence in their post and did not express support for Öcalan's ideology or the PKK. Instead, "he sought to highlight human rights concerns about Öcalan's prolonged solitary confinement, which have also been raised by international bodies". As the post was unlikely to result in harm, its removal "was not necessary or proportionate under international human rights standards". The Board also adds that "the breadth of the term "support" in the Community Standards combined with the misplacement of internal guidance on what this excludes, meant that an unnecessary and disproportionate removal occurred".

In the decision 2021-10, the Board overturned another decision to remove a post shared on the Facebook Page of a regional news outlet in Colombia by another Facebook Page without adding any additional caption. This shared post shows protesters using slurs against the President of the country. The post was qualified by the company to violate Facebook's Hate Speech Community Standard, which does not allow content that "describes or negatively targets people with slurs

based on protected characteristics such as sexual orientation". It needs to be noted, however, that the slur in question ("marica"), despite its most common meaning, was not used in this case to refer to the sexual orientation of the mandatary, but as a mere derogatory term. The Board concluded that the company "should have applied the internal newsworthiness allowance in this case, which requires Facebook to assess the public interest of allowing certain expression against the risk of harm from allowing violating content". The Board particularly notes how "the content was posted during widespread protests against the Colombian government at a significant moment in the country's political history". Based on the international human rights three-part test, the restriction was thus unnecessary and disproportionate.

In the decision 2022-02, upholding Meta's decision to restore a Facebook post depicting violence against a civilian in Sudan, the Board appreciates an internal contradiction in Meta's policy on Violent and Graphic Content: while the policy rationale states that Meta allows users to post graphic content "to help people raise awareness" about human rights abuses, the policy itself prohibits all videos (whether shared to raise awareness or not) "of people or dead bodies in non-medical settings if they depict dismemberment." The Board concludes "that Meta's policies should be amended to better respect the right to freedom of expression for users seeking to raise awareness of or document abuses". In addition to this, the Board also notes that "while it was used in this case, the newsworthiness allowance is not an effective means of allowing this kind of content on Facebook at scale". This is based on the fact that, based on data provided by the company, this allowance was used in an extremely limited number of cases in the prior 12 months.

In the decision 2022-05, overturning Meta's original decision to remove a Facebook post from a news outlet page reporting a positive announcement from the Taliban regime in Afghanistan on women and girls' education, the Board notes that this was inconsistent with Facebook's Dangerous Individuals and Organizations Community Standard, which permits reporting on terrorist groups, and Meta's human rights responsibilities. On this occasion, the Board is particularly concerned that Meta's systems and policies may interfere with freedom of expression when it comes to reporting on terrorist regimes. It notes how the company's Community Standards and internal guidance for moderators are not clear on how the praise prohibition and reporting allowance apply, or the relationship between them. In this case, two reviewers found the post was in violation of content moderation policies, which makes the Board to conclude that internal standards on these important matters may not be properly understood. The Board is particularly concerned that "Meta's default is to remove content under the Dangerous Individuals and Organizations policy if users have not made it clear that their intention is to report." The Board is also concerned that the content was not reviewed within the HIPO system, that is, a system Meta uses to identify cases where it has acted incorrectly, for example, by wrongly removing content. The apparent reason is the fact that the company counted on less than 50 Urdu-speaking reviewers allocated to HIPO at the time, and therefore the post was not deemed high priority and was never reviewed in the system.

Thirdly, human rights standards are also applied to assess internal rules used by moderators to interpret and enforce Community Standards. In the decision 2021-12 the OSB outlines how certain

internal policy rules given to moderators "may instruct them to err on the side of removing content that includes hate speech where the user's intent is not clear". In particular, the Board found in this case that "internal guidance provided limited instruction to moderators on how to properly distinguish prohibited hate speech from counter speech that quotes hate speech to condemn it or raise awareness". Aside from this, the Board denounces the "absence of any guidance on how to assess evidence of intent in artistic content quoting or using hate speech terms, or in content discussing human rights violations, where such content is covered by the policy allowances". In this context, the OSB reminds Meta of the fact that "it has a responsibility to perform human rights due diligence under UNGPs, and such responsibility includes identifying any adverse impacts of content moderation on artistic expression and the political expression of Indigenous peoples countering discrimination (as in the case under review)". It also tasks Meta to further identify how it will prevent, mitigate, and account for its efforts to address those adverse impacts.

In the decision 2022-01, which overturns Meta's original decision to leave a post on Facebook which depicted ethnic Serbs as rats, the Board requests Meta to clarify the Hate Speech Community Standard and the guidance provided to reviewers to guarantee that "even implicit references to protected groups are prohibited by the policy, when the reference would be reasonably understood". In this sense, the Board criticizes the fact that in the evolution of this case, "moderators consistently interpreted the Hate Speech policy as requiring them to identify an explicit, rather than implicit, comparison between ethnic Serbs and rats before finding a violation". The Board also disagrees with Meta's conclusion that this content did not violate Facebook's Violence and Incitement Community Standard. The Board particularly takes into consideration the remembrance of a past conflict and existence of an actual violent threat. In addition to all the above, this decision also contains an interesting reference to internal moderation processes, particularly when it comes to content escalation. In this sense, the Board points to the fact that in January 2022, when the Board identified the case for full review, "Meta decided that, while the post did not violate the letter of its Hate Speech policy, it did violate the spirit of the policy, and removed the post from Facebook. Later, when drafting an explanation of its decision for the Board, Meta changed its mind again, concluding that the post violated the letter of the Hate Speech policy, and all previous reviews were in error". The Board also observes that the content was not sent to Meta's specialized teams for assessment before it reached the Board. This shows that "the company's processes for escalating content are not sufficiently clear and effective" and urges Meta "to provide more information on how it escalates content".

Fourthly, in the decision 2020-03, a very specific international standard is used, one of the very few particularly crafted to be used by platforms when assessing speech. In his report to the UN Human Rights Council of 9 October 2019, the Special Rapporteur on freedom of opinion and freedom of expression[14] acknowledges that entities engaged in content moderation such as Facebook can "regulate" hate speech according to the scale, complexity, and long-term challenges that such forms of speech present on these platforms. Restrictions could thus be imposed "even if it is not

14 A/74/486. Available online at: https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/A_74_486.pdf

clearly linked to adverse outcomes (as hateful advocacy is connected to incitement in Article 20(2) of the ICCPR)". Based on this, the Board considers that in the case in question (use of the word "tazis" to refer to Azerbaijanis) the slur used was "hateful and dehumanizing". Although the expression could not be qualified as incitement, "the potential for adverse outcomes was nevertheless present". The use of dehumanizing language in a context of armed conflict may create "a discriminatory environment that undermines the freedom of others to express themselves". Therefore, the Board concludes that "the presence of these risks and Facebook's human rights responsibility to avoid contributing to them meant it was permitted to remove the slur."

4. Most relevant aspects of the legal reasoning of the OSB

4.1 Context matters

Context analysis is a central element of many decisions of the OSB. The Board has overturned several removal decisions adopted by Meta's services based on the company's alleged wrong appreciation of the context in which the content was posted and disseminated. As already expressed in one of its early decisions (2020-03), for the Board "context is key". This context assessment by the Board must not be seen as solely belonging to the area of evaluation of "facts", but as part of the criteria to determine the necessity and proportionality of certain restrictions, as required by the already mentioned three-part test.

To point just at a few examples, the Board considers that Meta's platforms failed to properly assess (or did not assess) relevant contextual elements in cases such as the following:

> a) A Facebook user in Myanmar published in Burmese a post that includes two widely shared photographs of a Syrian toddler of Kurdish ethnicity who drowned in the Mediterranean Sea in September 2015. The accompanying text begins by stating that "there is something wrong with Muslims". The post was removed by Facebook. The Board "acknowledges that it is difficult for Facebook to evaluate the intent behind individual posts when moderating content at scale and in real time". However, it also establishes that "even in circumstances where discussion of religion or identity is sensitive and may cause offense, open discussion remains important". Therefore "while some may consider the post offensive and insulting towards Muslims" the Board also declares that "removing this content is unlikely to reduce tensions or protect persons from discrimination". It is important to note that the particular intention of this post, as assessed by the Board, was to criticize the lack of reaction, from Muslims in general, vis-à-vis the abuses against the Uygur community in China, in comparison with responses to the publication of Mohammad cartoons in Europe. (2020-02).
>
> b) A user in Brazil posted a picture on Instagram with a title in Portuguese indicating that

it was to raise awareness of signs of breast cancer. Eight photographs within the picture showed breast cancer symptoms, with five of them including visible and uncovered female nipples, while the remaining three photographs included female breasts, with the nipples either out of shot or covered by a hand. The post was removed by an automated system enforcing Facebook's Community Standard on adult nudity and sexual activity. The Board acknowledges the fact that "automated technologies are essential to the detection of potentially violating content". However, it also warns that "enforcement which relies solely on automation, in particular, when using technologies that have a limited ability to understand context, leads to over-enforcement that disproportionately interferes with user expression." (2020-04)

c) A user posted a quote (incorrectly) attributed to Joseph Goebbels, the Minister of Propaganda in Nazi Germany. The quote, in English, stated that, rather than appealing to intellectuals, arguments should appeal to emotions and instincts. Facebook removed the post for violating its Community Standard on dangerous individuals and organizations. The Board affirms that "context is key for assessing necessity and proportionality" and that in this case "the content of the quote and other users' responses to it, the user's location and the timing of the post during an election campaign are all relevant". Therefore, "Facebook's approach requiring content moderators to review content without regard to these contextual cues resulted in an unnecessary and disproportionate restriction on expression." (2020-05)

d) A Facebook user posted in a public group a meme with the text: "if the tongue of the kafir starts against the Prophet, then the sword should be taken out of the sheath." The post also included hashtags referring to President Emmanuel Macron of France as the devil and calling for the boycott of French products. Facebook removed the post under its Community Standard on violence and incitement. The board emphasized that "just as people have the right to criticize religions or religious figures, religious people also have the right to express offence at such expression" and concluded that the post must have been interpreted as a criticism of Macron's response to religiously motivated violence, rather than a credible threat to violence. It was also found that Facebook "did not accurately assess all contextual information" and the specific piece of content had to be restored based on international human rights standards on freedom of expression. (2020-07)

e)A user shared a video post from Punjabi-language online media company Global Punjab TV. This featured a 17-minute interview with Professor Manjit Singh who is described as "a social activist and supporter of the Punjabi culture." In text accompanying the post, the user claimed that the right-wing organization Rashtriya Swayamsevak Sangh (RSS) was threatening to kill Sikhs, a minority religious group in India, and to repeat the "deadly saga" of 1984 when Hindu mobs massacred and burned Sikh men, women and children. The user alleged that Prime Minister Modi himself is formulating the threat of "Genocide of the Sikhs" on advice of the RSS President, Mohan Bhagwat. The user also claimed that

Sikh regiments in the army have warned Prime Minister Modi of their willingness to die to protect the Sikh farmers and their land in Punjab. After being reported by one user, a human reviewer determined that the post violated Facebook's Dangerous Individuals and Organizations Community Standard and removed it. The Board noted that "the post highlighted the concerns of minority and opposition voices in India that are allegedly being discriminated against by the government". It also establishes that "the political context in India when this post was made, with mass antigovernment farmer protests and increasing governmental pressure on social media platforms to remove related content, underscores the importance of getting decisions right" and the fact that "dominant platforms should avoid undermining the expression of minorities who are protesting their government and uphold media pluralism and diversity" (2021-04)

f) A Facebook user who appeared to be in Myanmar posted in Burmese on their timeline. The post discussed ways to limit financing to the Myanmar military following the coup in Myanmar on 1 February 2021. It proposed that tax revenue should be given to the Committee Representing Pyidaungsu Hluttaw (CRPH), a group of legislators opposed to the coup. Facebook translated the supposedly violating part of the user's post as "Hong Kong people, because the fucking Chinese tortured them, changed their banking to UK and now (the Chinese), they cannot touch them" and removed the post under its Hate Speech Community Standard. The Board considered that the case "highlights the importance of considering context when enforcing Hate Speech Policies, as well as the importance of protecting political speech". The Board also warns about the fact that "Facebook's policy of presuming profanity mentioning national origin refers to states and people may lead to disproportionate enforcement in some linguistic contexts, such as this one, where the same word is used for both." (2021-07).

g) In the very straightforward decision 2022-03, the Board overturns Meta's decision to remove an Instagram post from an account that describes itself as a space for discussing queer narratives in Arabic culture. The content included pictures showing words that can be used in a derogatory way towards men with "effeminate mannerisms" in the Arabic-speaking world, including the terms "zamel", "foufou" and "tante/tanta". The user stated that the post intended "to reclaim [the] power of such hurtful terms". The Board finds removing this content to be "a clear error which was not in line with Meta's Hate Speech Policy": while the post does contain slur terms, the content is covered by an exception for speech "used self-referentially or in an empowering way", as well as an exception which allows the quoting of hate speech to "condemn it or raise awareness". The Board expresses its concern regarding the fact that "reviewers may not have sufficient resources in terms of capacity or training to prevent the kind of mistake seen in this case".

h) A relevant and specific last example would refer to the use by the Board of what can be considered as "cumulative context". In the decision 2021-02 on caricatures of Black people in the form of blackface this body did not only consider the specific circumstances of the

case or even the intention of the user, but the systemic and cumulative effect of this kind of speech within the context of certain societies, concluding that "allowing such posts to accumulate on Facebook would help create a discriminatory environment for Black people that would be degrading and harassing."

4.2 Legality, legitimacy, and necessity and proportionality

As mentioned earlier, the Board has incorporated the three-part test as the basic and most solid tool to assess the possible restrictions to the right to freedom of expression submitted to its consideration.

Most decisions incorporate a very clear and comprehensive description of the three main elements of the test: legality (clarity and accessibility of the rules applied by the platform), legitimacy (pursuit of one of the legitimate aims listed in the ICCPR), and necessity and proportionality (restrictions on freedom of expression should be appropriate to achieve their protective function and should be the least intrusive instrument).

It has already been shown how in several decisions the Board has used the principle of legality to recommend Facebook to provide more information to users regarding the way Community Standards are interpreted and enforced, as well as better and more specific references for users as to the specific policies violated in a concrete case. The main objective of this principle is for users/citizens to be able to understand the rules that may constrain their speech and regulate their behavior accordingly. Besides this, a very characteristic application of the principle of legality within the context of Facebook's policies is the recommendation, in some cases, to provide moderators with better and more precise criteria in order to perform their activities in a more accurate and informed manner.

The board has considered, in certain cases, that although a community standard may not fully satisfy the principle of legality, when the content is undoubtedly covered by the respective prohibition the removal decision will be legitimate insofar as it meets the other parts of the test.

It is also worth noting that the Board has considered, in certain cases, that although a community standard does not fully satisfy the principle of legality, when the content is undoubtedly covered by the respective prohibition – and regardless of its more diffuse borders or its vagueness – the removal decision will be legitimate insofar as it meets the other parts of the test. For example, in the decisions 2020-05 and 2021-01 (President's Trump posts), the Board indicates that the Standard against praise and support of dangerous individuals and organizations "leaves much to be desired", in line with a criticism already expressed by the UN Special Rapporteur when providing examples of vague rules established by

online platforms[15]. However, in these decisions, which follow the precedent established by decision 2020-03, although the Standard may be considered as vague from a general perspective, the requirement of legality (in the sense of clarity and accessibility) is satisfied in light of the specific circumstances of the case.

Regarding legitimacy, the Board consistently refers to the rights and areas of public interest included in the last paragraph of article 19. This being said, it is also important to underscore that in the decision 2020-02 on use of the word "tazis" to refer to Azerbaijanis, the legitimacy of the restriction is appreciated by the Board according to the specific standards that the UN Special Rapporteur has established regarding hate speech on online platforms. As already mentioned, the determination of the Board is not purely based on criteria established by human rights law (adverse outcomes in terms of incitement connected to advocacy) but according to the scale, complexity, and long-term challenges that such form of speech presents on these platforms, "even if it is not clearly linked to adverse outcomes". This new basis for legitimacy, connected in any case with international human rights soft law, may open the door for the Board to consider other possible values and interests not directly contemplated by the ICCPR.

Similarly, in the decision 2022-01, and regarding an implicit comparison of ethnic Serbs with rats, the Board acknowledges that while prohibiting such kind of content would raise concerns if imposed by a government at a broader level, "particularly if enforced through criminal or civil sanctions, Facebook can regulate such expression, demonstrating the necessity and proportionality of the action."

The principle of necessity and proportionality (as it generally happens as well in decisions adopted by international human rights bodies and courts) has been used by the Board as a key criterion to assess and determine the legitimacy of content removal decisions. According to what has already been presented in the previous epigraph, it is common for the Board to establish a connection between the way a certain post needs to be understood and analyzed according to a series of contextual elements, and the Board's determination regarding the necessity and proportionality of the platform's decisions. As expressed in the 2021-01 decision, while Meta has the responsibility to create "necessary and proportionate penalties that respond to severe violations of its content policies", the Board has to ensure that "Facebook's rules and processes are consistent with its content policies, its values and its human rights commitments". Likewise, in its decision 2021-04, the Board clearly states that "context is key for assessing necessity and proportionality".

Also referring to necessity and proportionality, the Board has acknowledged and been sensitive to the existence of a big difference between decisions taken by Meta regarding the use of its services and determinations made by State bodies and courts: while the latter can devote extensive time to adopt a judgement related to an "isolated" case, decisions taken by Meta and

15 A/HRC/38/35 9, para 26, footnote 67.

sent for review are just a small sample of all the determinations made at scale on a constant basis covering a wide range of topics. As mentioned above, the second decision adopted by the Board (2020-02) already acknowledges that "it is difficult for Facebook to evaluate the intent behind individual posts when moderating content at scale and in real time".

A few examples include:

a) In the decision 2021-06 the very vague notions included under the Dangerous Individuals and Organizations policy combined with a proper assessment of the context and potential consequences of the post under analysis, lead to conclude that the removal decision was not necessary or proportionate under international human rights standards.

b) In the decision 2021-09 the Board considers that the removal of a re-published post from Al-Jazeera was not necessary as "it did not reduce offline harm and instead resulted in an unjustified [unnecessary] restriction on freedom of expression on a public interest issue". In this case, the Board assessed the necessity of the removal in connection with the internal value of "Voice" also considering the broader media and information environment in the Middle East region.

c) In the decision 2021-13, the Board concluded that international standards on necessity and proportionality pointed in the direction of restoring the contested content to Instagram, based on the consideration that there was no direct and immediate connection between the content, which primarily discussed the use of ayahuasca in a religious context, and the possibility of harm.

d) In a very similar way, in the decision 2021-15, the Board clearly states that Meta's decision to remove a post on how to talk to a doctor about the prescription medication Adderall to be unnecessary and disproportionate, as there was no direct or immediate connection between the content and the possibility of harm.

e) In the decision 2022-04, the Board overturns Meta's original decision to remove a Facebook post of a cartoon depicting police violence in Colombia. 16 months after the user posted the content, the company removed the content as it matched with an image in a Media Matching Service bank. The Board finds the removal of the content in this case particularly concerning as the content did not violate any Meta policy but contained criticism of human rights violations which is protected speech, in line with the determinations contained in the decision 2021-10 on a similar case in the same country. Restrictions thus imposed by Meta did not meet the requirements of legitimacy and necessity. The Board also notes with concern that the design of Meta's Media Matching Service banks enabled reviewers to mistakenly add content to a bank that resulted in the automatic removal of identical content, despite it being non-violating. This is considered to entail a flagrant violation of the principle of proportionality "even considering Meta's scale of

operation". In this sense, and from a broader perspective, the Board warns that "the use of Media Matching Service banks to remove content with limited or flawed feedback mechanisms raises concerns of disproportionate erroneous enforcement, where one mistake is amplified to a much greater scale". It particularly recommends that Meta ensures that content with high rates of appeal and high rates of successful appeal is re-assessed for possible removal from its Media Matching Service banks.

4.3 Beyond freedom of expression

It was noted in the introduction that according to its rules, the fundamental mission of the Board is to scrutinize content decisions "in light of human rights norms protecting free expression".

However, it needs to be noted that in some of its decisions, the Board has explicitly referred to the possible violation of other human rights as part of its assessment.

In the decision of the case 2021-05 on a meme referring the Armenian genocide, aside from its freedom of expression analysis, the Board notes that the incorrect notice given to the user of the specific content rule violated "implicates the right to be informed in the context of access to justice (Article 14, para. 3(a) ICCPR)." This body subsequently stresses that when limiting a user's right to expression, "Facebook must respect due process and inform the user accurately of the basis of their decision, including by revising that notice where the reason is changed (General Comment No. 32, para. 31)." It is also important to note that, as part of its final decision, the Board establishes that the mentioned right had been violated by Meta.

Similarly, in the decision 2021-06 on a post encouraging people to discuss the solitary confinement of Abdullah Öcalan, the Board took the opportunity to express concern regarding the fact that "neither users whose content is removed on the basis of the Community Standards, nor the Board, are informed where there was government involvement in content removal". This takes the Board to conclude that Meta did not respect the right to remedy, in contravention of its own Corporate Human Rights Policy (Section 3).

Decision 2021-15 refers to a case where Meta removed content under Facebook's Restricted Goods and Services Community Standard. Following the removal, Meta restricted the user's account for 30 days. As a result of the Board selecting this case, Meta identified its removal as an "enforcement error" and restored the content. The Board thus expressed dismay regarding the fact that the removal decision was not reversed until the case was brought to Meta's attention following the Board's selection, nor it was remedied. It is particularly stated that "Meta failed its responsibility to provide an effective remedy" and indicates that in the future, Meta "should make sure that user appeals are reviewed in a timely fashion when content-level enforcement measures also trigger account-level enforcement measures".

In the decision 2021-02 the right to non-discrimination plays a fundamental role in the conformation of the Boards judgement. As already mentioned, this body considers that "blackface" caricatures are inextricably linked to negative and racist stereotypes" and are considered to sustain systemic racism and racial discrimination, and thus endorses Facebook's removal decision.

There are, however, other cases where possible conflicting rights, particularly privacy and data protection, seem to be confronted with the right to freedom of expression in a more expeditious or even superficial manner. This approach would not be consistent with the way international human rights courts and other bodies generally aim at carefully considering and pondering the different rights at stake in a specific case. In the decision 2022-02 on a Facebook post depicting violence against a civilian in Sudan, the Board considers, as already mentioned, that placing a warning label on the content was a "necessary and proportionate restriction on freedom of expression" which did not place an undue burden on those who wish to see the content "while informing others about the nature of the content and allowing them to decide whether to see it or not". The Board also considers that such measure is also adequate to protect "the dignity of the individual depicted and their family". How a mere warning (and not other measures such as pixelating the face of the victim) may serve the purpose of protecting their dignity and prevent their identification is left unexplained.

There are, however, other cases where possible conflicting rights, particularly privacy and data protection, seem to be confronted with the right to freedom of expression in a more expeditious or even superficial manner.

Similarly, in the case 2021-16 the Board overturs Meta's decision to remove a post describing incidents of sexual violence against two minors, based on the assessment that the context of the post makes it clear that the user was reporting on an issue of public interest and condemning the sexual exploitation of a minor. The post contains a photo of a young girl sitting down with her head in her hands in a way that obscures her face. According to the description of the case, the photo has a caption in Swedish describing incidents of sexual violence against two minors. The post contains details about the rapes of two unnamed minors, specifying their ages and the municipality in which the first crime occurred. The user also details the convictions that the two unnamed perpetrators received for their crimes. The decision states that the Board was "unable to determine whether the pieces of information provided, along with links to media reports, could increase the possibility that the victims will be identified", while at the same time acknowledges the fact that some of its members "emphasised that when there is doubt about whether a specific piece of content may lead to functional identification of a child victim". This appears to be a very sensitive topic that requires a proper consideration of the different human rights at stake, particularly bearing in mind that both freedom of expression and the rights of the child (particularly in cases of sexual abuse) are subject to very strong protections under international law. The very descriptive and succinct paragraphs devoted to this very important question at the very end of the Board's analysis leave the general matter unresolved and fail to provide Meta with valuable criteria to be used in the future.

5. Conclusion

As mentioned in the text, the OSB Charter establishes that this body is entrusted with the task of analyzing content enforcement decisions by Facebook or Instagram and determine whether they were consistent with Meta's content policies and values. In fulfilling its remit, the Board "will pay particular attention to the impact of removing content in light of human rights norms protecting free expression".

The overview of the decisions adopted so far by the OSB shows that the Board has executed this task from a very specific angle. Instead of focusing on content policies and values as the main judgement criteria, this body has rather taken a human rights-based approach, which puts such international legal standards at the center of its internal debates and determinations. This means that every piece of content is indeed analyzed in light of internal policies, but at the same time these are subjected to a thorough human rights-based scrutiny and interpretation. This scrutiny has even taken the Board, in some cases, to the point of dismissing Community Standards and other moderation documents as the basis for the final decision, thus recommending their repeal or reform. It is obvious that such rationale goes beyond "paying particular attention" to human rights, as it rather puts them at the very center of the Board's set of deciding rules. Besides this, it is important to note, as it has already been stressed in different sections of this paper, that the human rights norms used by the Board in its decisions do not only refer to freedom of expression (although it occupies an obvious principal position) but to all the possible human rights and international values at stake within the context of each case.

This approach does not violate the remit of the Board as established in the Chapter, although it obviously expresses a very explicit statement with regards to the way this institution sees itself within Meta's content moderation machinery.

In terms of possible future developments of the role of the OSB, it needs to be noted that so far it has been performing its activities in a very much court-inspired manner. Decisions particularly focus on the content itself and analyze it still without thoroughly considering the characteristics of the different platforms (Facebook and Instagram) and the particularities and technicalities associated to their services. In the same vein, the Board has not been able so far to fully incorporate into its analysis the structural elements that determine at scale the way Meta polices content. This weakness must not necessarily be ascribed to the OSB. Meta's resistance, as already described in this paper, to provide to Board with comprehensive and complete information in this area is a significant factor to consider. In any case, following Evelyn's Douek interesting suggestions included in her very recent writings and presentations, a proper understanding and improvement of content moderation requires moving beyond the "constitutional" scope of freedom of expression debates to pursue meaningful accountability of content moderation systems as a whole and encourage necessary innovation. In other words, debates and regulatory proposals must not fixate on the application of a fixed rule to a specific piece of content but on a comprehensive understanding of the task involved[16].

16 Evelyn Douek, "Content Moderation as Administration", forthcoming at Harvard Law Review Vol. 136. Available online at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4005326


Global Freedom of Expression
Columbia University