

## Case Decision 2020-006-FB-FBR

### Case Summary

The Oversight Board has overturned Facebook’s decision to remove a post which it claimed, “contributes to the risk of imminent... physical harm.” The Board found Facebook’s misinformation and imminent harm rule (part of its Violence and Incitement Community Standard) to be inappropriately vague and recommended, among other things, that the company create a new Community Standard on health misinformation.

### About the case

In October 2020, a user posted a video and accompanying text in French in a public Facebook group related to COVID-19. The post alleged a scandal at the Agence Nationale de Sécurité du Médicament (the French agency responsible for regulating health products), which refused to authorize hydroxychloroquine combined with azithromycin for use against COVID-19, but authorized and promoted remdesivir. The user criticized the lack of a health strategy in France and stated that “[Didier] Raoult’s cure” is being used elsewhere to save lives. The user’s post also questioned what society had to lose by allowing doctors to prescribe in an emergency a “harmless drug” when the first symptoms of COVID-19 appear.

In its referral to the Board, Facebook cited this case as an example of the challenges of addressing the risk of offline harm that can be caused by misinformation about the COVID-19 pandemic.

### Key findings

Facebook removed the content for violating its misinformation and imminent harm rule, which is part of its Violence and Incitement Community Standard, finding the post contributed to the risk of imminent physical harm during a global pandemic. Facebook explained that it removed the post as it contained claims that a cure for COVID-19 exists. The company concluded that this could lead people to ignore health guidance or attempt to self-medicate.

The Board observed that, in this post, the user was opposing a governmental policy and aimed to change that policy. The combination of medicines that the post claims constitute a cure are not available without a prescription in France and the content does not encourage people to buy or take drugs without a prescription. Considering these and other contextual factors, the Board noted that Facebook had not demonstrated the post would rise to the level of imminent harm, as required by its own rule in the Community Standards.

The Board also found that Facebook’s decision did not comply with international human rights standards on limiting freedom of expression. Given that Facebook has a range of tools to deal with misinformation, such as providing users with additional context, the company failed to demonstrate why it did not choose a less intrusive option than removing the content.

The Board also found 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. A patchwork of policies found on different parts of Facebook’s website make it difficult for users to understand what content is prohibited. Changes to Facebook’s COVID-19 policies announced in the company’s Newsroom have not always been reflected in its Community Standards, while some of these changes even appear to contradict them.

### **The Oversight Board’s decision**

The Oversight Board overturns Facebook’s decision to remove the content and requires that the post be restored

In a policy advisory statement, the Board recommends that Facebook:

- Create a new Community Standard on health misinformation, consolidating and clarifying the existing rules in one place. This should define key terms such as “misinformation.”
- Adopt less intrusive means of enforcing its health misinformation policies where the content does not reach Facebook’s threshold of imminent physical harm.
- Increase transparency around how it moderates health misinformation, including publishing a transparency report on how the Community Standards have been enforced during the COVID-19 pandemic. This recommendation draws upon the public comments the Board received.

\*Case summaries provide an overview of the case and do not have precedential value.

### **Full Case Decision**

#### **1. Decision Summary**

The Oversight Board has overturned Facebook’s decision to remove content that it designated as health misinformation that “contributes to the risk of imminent . . . physical harm.” The Oversight Board found that Facebook’s decision did not comply with its Community Standards, its values, or international human rights standards.

#### **2. Case Description**

In October 2020, a user posted a video and accompanying text in French in a Facebook public group related to COVID-19. The video and text alleged a scandal at the Agence Nationale de Sécurité du Médicament (the French agency responsible for regulating health products) which refused to authorize hydroxychloroquine combined with azithromycin for use against COVID-19,

but authorized and promoted remdesivir. The user criticized the lack of a health strategy in France and stated that “[Didier] Raoult’s cure” is being used elsewhere to save lives. Didier Raoult (who is mentioned in the post) is a professor of microbiology at the Faculty of Medicine of Marseille, and directs the “Institut Hospitalo-Universitaire Méditerranée Infection” (IHU) in Marseille. The user’s post also questioned what society had to lose by allowing doctors to prescribe in an emergency a “harmless drug” when the first symptoms of COVID-19 appear. The video claimed that the combination of hydroxychloroquine and azithromycin was administered to patients at early stages of the disease and implied this was not the case for remdesivir. The post was shared in a public group related to COVID-19 with more than 500,000 members and received about 50,000 views, about 800-900 reactions (the majority of which were “angry” followed by “like”), 200-300 comments on the post made by 100-200 different people and was shared by 500-600 people. Facebook removed the content for violating its Community Standard on Violence and Incitement. In referring its decision to the Oversight Board, Facebook cited this case as an example of the challenges of addressing the risk of offline harm that can be caused by misinformation about the COVID-19 pandemic.

### **3. Authority and Scope**

The Board has authority to review Facebook’s decision under Article 2 (Authority to Review) of [the Board’s Charter](#) and may uphold or reverse that decision under Article 3, Section 5 (Procedures for Review: Resolution of the Charter). Facebook has not presented reasons for the content to be excluded in accordance with Article 2, Section 1.2.1 (Content Not Available for Board Review) of [the Board’s Bylaws](#), nor has Facebook indicated that it considers the case to be ineligible under Article 2, Section 1.2.2 (Legal Obligations) of the Bylaws. Under Article 3, Section 4 (Procedures for Review: Decisions) of the Board’s Charter, the final decision may include a policy advisory statement, which will be taken into consideration by Facebook to guide its future policy development.

### **4. Relevant Standards**

The Oversight Board considered the following standards in its decision:

#### *I. Facebook’s Community Standards:*

The introduction to Facebook’s Community Standards includes a link titled “COVID-19: Community Standards Updates and Protections” that states:

*As people around the world confront this unprecedented public health emergency, we want to make sure that our Community Standards protect people from harmful content and new types of abuse related to COVID-19. We’re working to remove content that has the potential to contribute to real-world harm, including through our policies prohibiting coordination of harm, sale of medical masks and related goods, hate speech, bullying and harassment and misinformation that contributes to the risk of imminent violence or physical harm.*

Facebook stated that it relied specifically on the prohibition on “misinformation and unverifiable rumors that contribute to the risk of imminent violence or physical harm,” which is contained

within the Community Standard on Violence and Incitement (referred to as the “misinformation and imminent harm rule” from this point on). The rule appears under the qualification that it “require[s] additional information and/or context to enforce.”

Facebook’s policy rationale for Violence and Incitement states that it aims “to prevent potential offline harm that may be related to content on Facebook.” Facebook further states that it removes content “that incites or facilitates serious violence” and “when it believes there is a genuine risk of physical harm or direct threats to public safety.”

Although Facebook did not rely on its Community Standard on False News in this case, the Board notes the range of enforcement options besides removal under this policy.

## *II. Facebook’s Values:*

The introduction to the Community Standards notes that “Voice” is Facebook’s paramount value. The Community Standards describe this value as:

*The goal of our Community Standards has always been to create a place for expression and give people a voice. [...] We want people to be able to talk openly about the issues that matter to them, even if some may disagree or find them objectionable.*

However, the platform may limit “Voice” in service of several other values, including “Safety”. Facebook defines its “Safety” value as:

*We are committed to making Facebook a safe place. Expression that threatens people has the potential to intimidate, exclude or silence others and isn’t allowed on Facebook.*

## *III. Relevant Human Rights Standards:*

The UN Guiding Principles on Business and Human Rights ( [UNGPs](#)), endorsed by the UN Human Rights Council in 2011, establish a voluntary framework for businesses’ human rights responsibilities. The Board’s analysis in this case was informed by UN treaty provisions and the authoritative guidance of the UN’s human rights mechanisms, including the following:

- International Covenant on Civil and Political Rights ( [ICCPR](#)), Article 19;
- Human Rights Committee General Comment No. 34 on freedom of opinion and expression (2011) ( [General Comment No. 34](#));
- UN Special Rapporteur on freedom of opinion and expression, report on Disease Pandemics and the Freedom of Opinion and Expression, [A/HRC/44/49](#) (2020), [Research Paper 1/2019](#) on Elections in the Digital Age (2019), and reports [A/74/486](#) (2019) and [A/HRC/38/35](#) (2018).

## **5. User Statement**

Facebook referred this case to the Oversight Board. Facebook confirmed to the Oversight Board that the platform sent the user notification of the opportunity to file a statement with respect to this case, but the user did not submit a statement.

## **6. Explanation of Facebook's Decision**

Facebook removed the content for violating its misinformation and imminent harm rule under its Violence and Incitement Community Standard. According to Facebook, the post contributed to the risk of imminent physical harm during a global pandemic.

Facebook explained that it removed this content because (1) the post claimed a cure for COVID-19 exists, which is refuted by the World Health Organization (WHO) and other credible health authorities, and (2) leading experts have told Facebook that content claiming that there is a guaranteed cure or treatment for COVID-19 could lead people to ignore preventive health guidance or attempt to self-medicate. Facebook explained that is why it does not allow false claims about cures for COVID-19.

Facebook elaborated that in cases involving health misinformation, the company consults with the WHO and other leading public health authorities. Through that consultation, Facebook has identified different categories of health misinformation about COVID-19, such as false claims about immunity (e.g., "People under age thirty cannot contract the virus"), false claims about prevention (e.g., "Drinking a gallon of cold water gives you about an hour of immunity"), and false claims about treatments or cures (e.g., "Drinking a tablespoon of bleach cures the virus").

Facebook considered this case as significant because it concerns a post that was shared within a large public Facebook group related to COVID-19, and therefore had the potential to reach a large population at risk of COVID-19 infection. Also, Facebook considered this case to be difficult because it creates tension between Facebook's values of "Voice" and "Safety." Facebook observed that the ability to discuss and share information about the COVID-19 pandemic and to debate the efficacy of potential treatments and mitigation strategies must be preserved while the spread of false information that could lead to harm must be limited.

## **7. Third party submissions**

The Board received eight public comments: one from Asia Pacific and Oceania, three from Europe and four from United States and Canada. Seven of these public comments have been published with this case, while one comment was submitted without consent to publish. The submissions covered a number of themes, including the importance of meaningful transparency and less intrusive measures as alternatives to removal; general critique on censorship, bias and Facebook's handling of misinformation related to the pandemic, as well as feedback for improving the public comment process.

## **8. Oversight Board Analysis**

### **8.1 Compliance with Community Standards**

Facebook removed the content on the basis that it violated its misinformation and imminent physical harm rule. Facebook stated the post constituted misinformation because it asserted there was a cure for COVID-19 whereas the WHO and leading health experts had found there is no cure. Facebook noted that leading experts had advised the platform that COVID-19 misinformation can be harmful because, if those reading misinformation believe it, then they may disregard precautionary health guidance and/or self-medicate. Facebook relied on this general expert advice to assert that the post in question could contribute to *imminent* physical harm. In addition, Facebook noted someone had died after ingesting a chemical that is commonly used to treat aquariums because of COVID-19 related misinformation.

The Board finds that Facebook has not demonstrated how this user's post contributed to *imminent* harm in this case. Instead, the company appeared to rely on equating any misinformation about COVID-19 treatments or cures as necessarily rising to the level of imminent harm. Facebook's Community Standards state that additional information and context is needed before Facebook removes content under its misinformation and imminent harm rule. However, the Community Standards do not explain what contextual factors are considered and Facebook did not discuss specific contextual factors in its rationale for this case.

Deciding whether misinformation contributes to Facebook's own standard of "*imminent*" harm requires an analysis of a variety of contextual factors, including the status and credibility of the speaker, the reach of his/her speech, the precise language used, and whether the alleged treatment or cure is readily available to an audience vulnerable to the message (such as the misinformation noted by Facebook about resorting to water or bleach as a prevention or cure for COVID-19).

In this case, a user is questioning a government policy and promoting a widely known though minority opinion of a medical doctor. The post is geared towards pressuring a governmental agency to change its policy; the post does not appear to encourage people to buy or take certain drugs without a medical prescription. Serious questions remain about how the post would result in *imminent* harm. While some studies indicate the combination of anti-malarial and antibiotic medicines that are alleged to constitute a cure may be harmful, experts the Board consulted noted that they are not available without a prescription in France. Moreover, the alleged cure has not been approved by the French authorities and thus it is unclear why those reading the post would be inclined to disregard health precautions for a cure they cannot access. The Board also notes that this public group on Facebook could have French speaking users based outside of France. Facebook did not address particularized contextual factors indicating potential imminent harm with respect to such users. The Board remains concerned about health misinformation in France and elsewhere (see Policy Recommendation II. b.). In sum, while the Board acknowledges that misinformation in a global pandemic can cause harm, Facebook failed to provide any contextual factors to support a finding that this particular post would meet its own *imminent harm* standard. Facebook therefore did not act in compliance with its Community Standard.

The Board also notes that this case raises important issues of distinguishing between opinion and fact; along with the question of when "misinformation" (which is undefined in the Community

Standards) is an appropriate characterization. It also raises the question of whether an allegedly factually incorrect claim in a broader post criticizing governmental policy should trigger the removal of the entire post. While we need not consider these issues in deciding whether Facebook acted consistently with its misinformation and imminent harm rule in this case, the Board notes such issues could be critical in future applications of the rule.

## **8.2 Compliance with Facebook Values**

The Oversight Board finds that the decision to remove the content was not consistent with Facebook's values. Facebook's rationale did not demonstrate the danger of this post to the value of "Safety" in a manner sufficient to displace "Voice" to the extent of justifying removal of the post.

## **8.3 Compliance with Human Rights Standards on Freedom of Expression**

This section examines whether Facebook's decision to remove the post from its platform is consistent with international human rights standards. Article 2 of our Charter specifies that we must "pay particular attention to the impact of removing content in light of human rights norms protecting free expression." Under the UNGPs companies are expected "to respect international human rights standards in their operations and address negative human rights impacts with which they are involved" (UNGPs, Principle 11.). International human rights standards are defined by reference to UN instruments, including the ICCPR (UNGPs, Principle 12.). In addition, the UNGPs specify that non-judicial grievance mechanisms (such as the Oversight Board) should deliver outcomes that accord with internationally recognized human rights (UNGPs, Principle 31.). In explaining its rationale for removing the content, Facebook acknowledged the applicability of the UNGPs and ICCPR to its content moderation decision.

Article 19 para. 2 of the ICCPR provides broad protection for expression of "all kinds." The UN Human Rights Committee has highlighted that the value of expression is particularly high when discussing matters of public concern (General Comment No. 34, paras. 13, 20, 38). The post in question is a direct critique of governmental policy and appears aimed at getting the attention of the Agence Nationale de Sécurité du Médicament. The user raises a matter of public concern, albeit by including the invocation and promotion of a minority opinion within the medical community. The fact that an opinion reflects minority views does not make it less worthy of protection. The user questions why doctors should not be allowed to prescribe a particular drug in emergency situations and does not call on the general public to independently act on Raoult's minority opinion.

That said, ICCPR Article 19, para. 3 permits restrictions on freedom of expression when a speech regulator can prove three conditions are met. In this case Facebook should show that its decision to remove content met the conditions of legality, legitimacy and necessity. The Board examines Facebook's removal of the user's post in light of this three-part test.

### *I. Legality*

Any restriction on expression should give appropriate notice to individuals, including those charged with implementing the restrictions, of what is prohibited. (See General Comment No. 34, para. 25). In this case, the legality test requires assessing whether the misinformation and imminent harm rule is inappropriately vague. To begin with, this rule contains no definition of “misinformation.” As noted by the UN Special Rapporteur on Freedom of Opinion and Expression, “vague and highly subjective terms-such as ‘unfounded,’ ‘biased,’ ‘false,’ and ‘fake’-do not adequately describe the content that is prohibited” (Research Paper 1/2019, p. 9). They also provide authorities with “broad remit to censor the expression of unpopular, controversial or minority opinions” (Research Paper 1/2019, p. 9). Further, such vague prohibitions empower authorities with “the ability to determine truthfulness or falsity of content in the public and political domain” and “incentivize self-censorship” (Research Paper 1/2019, p. 9). The Board also notes that this policy falls under a heading that states additional information and/or context is necessary to determine violations, but no indication is given of what type of additional information/context is relevant to this assessment.

Moreover, Facebook has announced multiple COVID-19 policy changes through its Newsroom without reflecting those changes in the current Community Standards. Unfortunately, the Newsroom announcements sometimes appear to contradict the text of the Community Standards. For example, in the Newsroom post “[Combating COVID-19 Misinformation Across Our Apps](#)” (March 25, 2020) Facebook specified it will “remove COVID-19 related misinformation that *could* contribute to imminent physical harm,” implying a different threshold than the misinformation and imminent harm rule, which addresses misinformation that “contributes” to imminent harm. In its mid-December 2020 Help Desk article, “[COVID-19 Policy Updates and Protections](#),” Facebook states that it would:

*remove misinformation that contributes to the risk of imminent violence or physical harm. In the context of a pandemic such as COVID-19, this applies to (...) claims that there is a ‘cure’ for COVID-19, until and unless the World Health Organization or other leading health organization confirms such cure. This does not prevent people from discussing medical trials, studies or anecdotal experiences about cures or treatments for the known symptoms of COVID-19 (e.g. fever, cough, breathing difficulties).*

This announcement (which was made after the post in question was removed) reflects the constantly evolving nature of both scientific and governmental stances on health issues. However, it was not integrated into the Community Standards.

Given this patchwork of rules and policies that appear on different parts of Facebook’s website, the lack of definition of key terms such as “misinformation,” and the differing standards relating to whether the post “could contribute” or actually contributes to imminent harm, it is difficult for users to understand what content is prohibited. The Board finds the rule applied in this case was inappropriately vague. The legality test is therefore not met.

## *II. Legitimate aim*



The legitimacy test provides Facebook's removal of the post should serve a legitimate and specified public interest objective in Article 19, para. 3 of the ICCPR (General Comment No. 34, paras. 28-32). The goal of protecting public health is specifically listed in this Article. We find that Facebook's purpose of protecting public health during a global pandemic satisfied this test.

### *III. Necessity and proportionality*

With regard to the necessity test, Facebook should demonstrate that it has selected the least intrusive means to address the legitimate public interest objective (General Comment No. 34, para. 34).

Facebook should show three things:

- (1) the public interest objective could not be addressed through measures that do not infringe on speech,
- (2) among the measures that infringe on speech, Facebook has selected the least intrusive measure, and
- (3) the selected measure actually helps achieve the goal and is not ineffective or counterproductive (A/74/486, para. 52).

Facebook has a range of options available to deal with false and potentially harmful health-related content. The Board asked Facebook whether less intrusive means could have been deployed in this case. Facebook responded that for cases of imminent harm, its sole enforcement measure is removal, but for content assessed by external partners as false (but not linked to imminent harm), it deploys a range of enforcement options short of content removals. This response essentially re-stated how its Community Standards work but did not explain why removal was the least intrusive means of protecting public health.

As noted in its Community Standard on False News, Facebook's tools to address such content include the disruption of economic incentives for people and pages that promote misinformation; the reduction of the distribution of content rated false by independent fact checkers; and the ability to counter misinformation by providing users with additional context and information about a particular post, including through [Facebook's COVID-19 Information Center](#). The Board takes note of Facebook's False News policy - not to imply that it should be used to judge opinions, but to note that Facebook has a range of enforcement options beyond content removals to deal with misinformation.

Facebook did not explain how removal of content in this case constituted the least intrusive means of protecting public health because, among other things, it did not explain how the post related to imminent harm; it merely asserted imminent harm to justify removal. The removal of the post therefore failed the necessity test.

## **9. Oversight Board Decision**

## 9.1 Content Decision

The Oversight Board decides to overturn Facebook's decision to remove the post in question.

## 9.2 Policy Advisory Statements

*I. Facebook should clarify its Community Standards with respect to health misinformation, particularly with regard to COVID-19.*

The Board recommends that Facebook set out a clear and accessible Community Standard on health misinformation, consolidating and clarifying existing rules in one place (including defining key terms such as misinformation). This rule-making should be accompanied with "detailed hypotheticals that illustrate the nuances of interpretation and application of [these] rules" to provide further clarity for users (See report [A/HRC/38/35](#), para. 46 (2018)). Facebook should conduct a human rights impact assessment with relevant stakeholders as part of its process of rule modification (UNGPs, Principles 18-19).

*II. Facebook should adopt less intrusive enforcement measures for policies on health misinformation.*

a.) To ensure enforcement measures on health misinformation represent the least intrusive means of protecting public health, the Board recommends that Facebook:

- Clarify the particular harms it is seeking to prevent and provide transparency about how it will assess the potential harm of particular content;
- Conduct an assessment of its existing range of tools to deal with health misinformation;
- Consider the potential for development of further tools that are less intrusive than content removals;
- Publish its range of enforcement options within the Community Standards, ranking these options from most to least intrusive based on how they infringe freedom of expression,
- Explain what factors, including evidence-based criteria, the platform will use in selecting the least intrusive option when enforcing its Community Standards to protect public health; and
- Make clear within the Community Standards what enforcement option applies to each rule.

b.) In cases where users post information about COVID-19 treatments that contradicts the specific advice of health authorities and where a potential for physical harm is identified but is not imminent, the Board strongly recommends Facebook to adopt a range of less intrusive measures. This could include labelling which alerts users to the disputed nature of the post's content and provides links to the views of the World Health Organization and national health authorities. In certain situations it may be necessary to introduce additional friction to a post - for example, by preventing interactions or sharing, to reduce organic and algorithmically driven amplification. Downranking content, to prevent visibility in other users' newsfeeds, might also be considered. All enforcement measures, including labelling or other methods of introducing friction, should be clearly communicated to users, and subject to appeal.

*III. Facebook should increase transparency of its content moderation of health misinformation.*

The Board recommends that Facebook improves its transparency reporting on health misinformation content moderation and drawing upon public comments received:

- Publish a transparency report on how the Community Standards have been enforced during the COVID-19 global health crisis. This should include:
  - data in absolute and percentage terms on the number of removals, as well as data on other enforcement measures, on the specific Community Standards enforced against, including on the proportion that relied entirely on automation;
  - a breakdown by content type enforced against (including individual posts, accounts, and groups);
  - a breakdown by the source of detection (including automation, user flagging, trusted partners, law enforcement authorities);
  - a breakdown by region and language;
  - metrics on the effectiveness of less intrusive measures (e.g., impact of labelling or downranking);
  - data on the availability of appeals throughout the crisis, including the total number of cases where appeal was withdrawn entirely, and the percentage of automated appeals;
  - conclusions and lessons learned, including information on any changes Facebook is making to ensure greater compliance with its human rights responsibilities going forward.

**\*Procedural note:**

The Oversight Board's decisions are prepared by panels of five Members and must be agreed by a majority of the Board. Board decisions do not necessarily represent the personal views of all Members.

For this case decision, independent research was commissioned on behalf of the Board. An independent research institute headquartered at the University of Gothenburg and drawing on a team of over 50 social scientists on six continents, as well as more than 3,200 country experts from around the world, provided expertise on socio-political and cultural context.