Oversight Board

Public Comment Appendix for

PAO-2021-02
Case number

Case description

In this summary, Facebook refers to the social media platform while Meta refers to the company that owns and manages Facebook.

This policy advisory opinion request concerns Meta's policy on cross-check, a system that the company says helps it "ensure that enforcement decisions about [Facebook's] Community Standards are made accurately and with additional levels of human review."

Meta told the Board that the cross-check system plays an important role in protecting "Voice" and human rights. The company also stated that the system serves an important role in managing relationships with its business partners. "Incorrectly removing content posted by a page or profile with a large following, for instance, can result in negative experiences for both Facebook's business partners and the significant number of users who follow them."

Meta explained that Facebook's "primary review systems use technology to prioritize high-severity content, which includes 'viral' content that spreads quickly." When systems flag such content for escalation, Facebook reviewers decide if the content should remain on the platform. Meta also explained that although it "aim[s] to make the right decisions, it recognizes that false positives [erroneous removal of non-violating content] do occur." Cross-check is one of the systems Meta uses to prevent false positive mistakes on Facebook; it is not the only such system the company uses.

Meta stated that, historically, the company determined who should receive cross-check review by compiling lists of users or entities with higher associated risk of false positive actions against them. "'False positive risk' refers to the risk of incorrect enforcement against content or entities that do not actually violate"

Facebook's Community Standards. Meta applied a variety of criteria, including the type of user or entity (e.g., an elected official, journalist, significant business partner, human rights organization), the number of followers, and the subject matter of the entity. When users or entities on those lists posted content or took actions that Facebook's systems flagged, they would be added to a queue for cross-check review."

At the beginning of 2020, Meta: "made changes so that most content in the queue was prioritized using a risk framework, which assigned a level of false-positive risk that could result if Facebook incorrectly removed that content. This risk framework generally relied on three factors: (1) the sensitivity of the entity, (2) the severity of the alleged violation, and (3) the severity of the potential enforcement action. Based on those factors, the content would be assigned one of three tiers of review: low (reviewed by contract reviewers), medium (reviewed by our markets team who have specialized regional expertise), and high (reviewed by our markets team and Early Response team who have deeper policy expertise and the ability to factor in additional context). Within those review tiers, the content in the queue was then prioritized by potential policy violation severity."

Reviewers would then examine the content, confirm whether it violated its policies, and if so, enforce those policies. This review, depending on case complexity, potentially could escalate all the way to the company's leadership.

Meta rolled out additional changes to the cross-check system in 2021, after it conducted a holistic analysis of the system and identified opportunities for improvement. Meta implemented some changes, including breaking the cross-check into two components. "General Secondary Review" and "Early Response (ER) Secondary Review."

ER Secondary Review is the historical cross-check system, described above, and will continue to be maintained using lists of entities (i.e. users, pages, groups, etc.). However, Meta changed the process of compiling and revising cross-check lists. "Prior to September 2020, most employees had the ability to add a user or entity to the cross-check list. After September 2020, while any employee can request that a user or entity be added to cross-check lists, only a designated group of employees

have the authority to make additions to the list." The lists relevant to ER Secondary Review contained more than 660,000 entities as of October 16, 2021. Meta explained that the lists are not static and change as entities are added and removed.

The General Secondary Review represents the majority of cross-check and will continue to grow. By the end of 2021, Meta aims to make the system available to all Instagram and Facebook users and entities. General Secondary Review will operate using a dynamic prioritization system called "cross-check ranker." The "cross-check ranker ranks content based on false positive risk using criteria such as topic sensitivity (how trending/sensitive the topic is), enforcement severity (the severity of the potential enforcement action), false positive probability, predicted reach, and entity sensitivity (based largely on the compiled lists, described above)." The General Secondary review represents the majority of cross-checked content and entities.

In developing the cross-check ranker, Meta interviewed 14 internal stakeholders across the operations, policy, and product teams to better understand risks of over-enforcement. The company "chose internal stakeholders due to the complications of explaining how enforcement works, but [is] considering external engagement in the future."

Meta concedes that despite investing significant resources to improve cross-check it still has difficulties striking a balance between removing content that violates Facebook's polices "while ensuring that it continues to foster open communication and free expression." The company asked the Board three questions, which the Board is including in its call for public comments.

Questions posed by Meta to the Board:

- Because of the complexities of content moderation at scale, how should Meta balance its need to fairly and objectively apply Facebook's Community Standards with the need for flexibility, nuance, and context-specific decisions within cross-check?
- What improvements should Meta make to how it governs the Early Response ("ER") Secondary Review cross-check system to fairly enforce Facebook's

Community Standards while minimizing the potential for over-enforcement, retaining business flexibility, and promoting transparency in the review process?

- What criteria should Meta use to determine who is included in ER Secondary Review and prioritized as one of many factors by the company's cross-check ranker in order to help ensure equity in access to this system and its implementation?

Board requests for public comments on the following issues:

- Whether a cross-check system is needed and if it strengthens or undermines the protection of freedom of expression and other human rights.
- Cross-check is designed to be a "false positive" prevention mechanism. What are the checks and balances, if any, this system should contemplate to mitigate the risks of "false negatives" [erroneous lack of action on violating content]?
- Recommendations on what Meta should do to ensure that the cross-check system, including its escalation process, is neutral and free of political and other biases.
- What factors should Meta incorporate into the "cross-check ranker" system in addition to topic sensitivity, enforcement severity, false-positive probability, predicted reach, and the nature and importance of the entity? How should these factors be defined?
- The benefits and limitations of automated technologies used to prioritize review of high-severity content.
- Information on how the cross-check system should and can be improved for users and entities who do not post in English.
- Information on systems akin to cross-check used by other social media platforms and lessons learned that can be applicable to Meta.
- How Meta can improve transparency of the cross-check system.
- What additional research and resources should Meta dedicate to improving the cross-check system?

# Oversight Board

Public Comment Appendix for

PAO-2021-02

Case number

The Oversight Board is committed to bringing diverse perspectives from third parties into the case review process. To that end, the Oversight Board has established a public comment process.

Public comments respond to case descriptions based on the information provided to the Board by users and Facebook as part of the appeals process. These case descriptions are posted before panels begin deliberation to provide time for public comment. As such, case descriptions reflect neither the Board's assessment of a case, nor the full array of policy issues that a panel might consider to be implicated by each case.

To protect the privacy and security of commenters, comments are only viewed by the Oversight Board and as detailed in the Operational Privacy Notice. All commenters included in this appendix gave consent to the Oversight Board to publish their comments. For commenters who did not consent to attribute their comments publicly, names have been redacted. To withdraw your comment, please email contact@osbadmin.com.

To reflect the wide range of views on cases, the Oversight Board has included all comments received except those clearly irrelevant, abusive or disrespectful of the human and fundamental rights of any person or group of persons and therefore violating the Terms for Public Comment. Inclusion of a comment in this appendix is not an endorsement by the Oversight Board of the views expressed in the comment. The Oversight Board is committed to transparency and this appendix is meant to accurately reflect the input we received.

Oversight Board
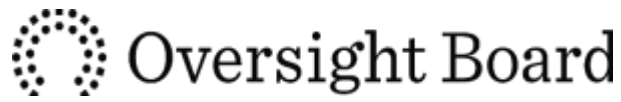
Public Comment Appendix for
Policy Advisory Opinion (2021-02)
Case number

87
Number of Comments

Regional Breakdown

9	2	12	3
Asia Pacific & Oceania	Central & South Asia	Europe	Latin America & Caribbean
3	3	55	
Middle East and North Africa	Sub-Saharan Africa	United States & Canada	

PAO-2021-02

PC-10303

United States and Canada

Andrew

Ronding

English

DID NOT PROVIDE

No

————

Short summary provided by the commenter

The cross-check program is unfair and should be abandoned.

Full Comment

Treating users differently leads to hurt feelings and confusion. If a popular user is allowed to post certain things with less review than others it will lead to regular users misunderstanding what rules are enforced and they are more likely to post content that violates content policies. It also leads to popular users not knowing what regular users experience with content moderation. Please discontinue use of the cross check program. Thank you for your consideration.

Link to Attachment
No Attachment

PAO-2021-02	PC-10310	United States and Canada
Case number	Public comment number	Region

Withheld	Withheld	
Commenter's first name	Commenter's last name	Commenter's preferred language

Withheld		No
Organization		Response on behalf of organization

––––

Short summary provided by the commenter

Facebook should add public tags to any people and/or pages exempted from standard moderation policies by virtue of "cross-check." Oversight Board should be disbanded.

Full Comment

Facebook's request for comment highlights a recurring framing strategy that it uses to undermine criticism of or oversight over its dysfunctional policies--it invents and arbitrarily and unnecessarily complex system and then claims the hopeless complexity of the system itself is the reason it can't be subject to oversight, or can't be transparent about what it is doing. The answer here is simple. The default should be that all users are subject to the same rules and processes for content moderation. Should Facebook decide that a certain individual or organization should receive, by virtue of its identity or notoriety, additional protections against its own notoriously and famously error-prone content moderation, it should so indicate on the person's public page--something akin to Twitter's "verified account" blue checkmark, so users interacting with that page or user will know that content on that page or from that user is not subject to the same review processes as everyone else. Unfortunately, Facebook's lying to the Oversight Board and to the world about this program shows that it is not truly invested in fixing the program and is rather simply interested in coopting the very respectable members of the Oversight Board into legitimizing or giving a veneer of respectability to Facebook's corrupt, broken business practices. Everyone involved in this sham process should resign.

Link to Attachment
No Attachment

PAO-2021-02

Case number

PC-10312

Public comment number

United States and Canada

Region

Withheld

Commenter's first name

Withheld

Commenter's last name

English

Commenter's preferred language

Withheld

Organization

No

Response on behalf of organization

––––

Short summary provided by the commenter

Disband cross check

Full Comment

Cross check must be immediately disbanded. Meta must treat its entire user base as equals and apply the same policies and privileges to every person and organization. No one should get unequal or otherwise special treatment from Meta by virtue of their perceived importance.

Link to Attachment
No Attachment

PAO-2021-02

Case number

PC-10317

Public comment number

United States and Canada

Region

Khadija

Commenter's first name

Gurnah

Commenter's last name

English

Commenter's preferred language

SumOfUS

Organization

Yes

Response on behalf of organization

————

Short summary provided by the commenter

As whistleblowers come forward to expose the harm caused by Facebook's business model, the mirage of accountability created by the Oversight Board is providing the platform cover to continue to evade responsibility. Consequently, over 50,000 of our members are calling on the Board, to either take meaningful action to protect civil society from the harms posed by the platform - or resign.

Full Comment

Dec 14, 2021 Dear Facebook Oversight Board, I'm writing on behalf of SumOfUs, a global nonprofit advocacy organization of over 19 million members that campaigns to hold corporations accountable. Our members are particularly alarmed at the harms caused by Facebook and other tech giants, with over 400,000 of our members signing petitions calling for tighter regulations on tech platforms. As whistleblowers come forward to expose the harm caused by Facebook's business model, the mirage of accountability created by the Oversight Board is providing the platform cover to continue to evade responsibility. Consequently, over 50,000 of our members are calling on the Board, to either take meaningful action to protect civil society from the harms posed by the platform - or resign. Specifically, we call on you to ensure that Meta/Facebook: *Discloses the full list of Facebook users protected under Facebook's Cross-Check programme, disaggregated by country. *Discloses the eligibility criteria for inclusion in the Cross-Check programme and for whitelisting certain users. *Discloses the management structure for oversight of this programme. *Discloses all the internal research that Facebook has carried out in relation to the Cross-Check programme. *Commits to an independent human rights impact assessment of Facebook's content moderation failures emerging from the Cross-Check programme. *Publishes in full any human rights impact assessments

with relation to specific countries. As Facebook's self appointed watchdog, your continued inability to rein in the platform undermines your legitimacy as a Board and tarnishes your reputations as individuals. It's time for you to stand up for civil society and our increasingly fragile democracies by taking immediate and impactful action - or resign. Sincerely, Khadija Gurnah (Campaigns Advisor), the SumOfUs team and membership. SumOfUs

Link to Attachment
No Attachment

PAO-2021-02

Case number

PC-10330

Public comment number

Europe

Region

REZALT

Commenter's first name

KASAJ

Commenter's last name

English

Commenter's preferred language

GO DIGITALMEDIA

Organization

Yes

Response on behalf of organization

————

Short summary provided by the commenter

Facebook and Instagram need to change policies in the way they treat users. For several weeks following the case of one of our clients, we have encountered an arrogance on the part of the company. When it comes to copyright infringement, Meta is very strict which is considered fair. But on the other hand, it does not apply equal rules to all. Sends automatic replies to appeals and is truncated in information. They refuse to give details, claiming to "preserve privacy" but on the other hand after many attempts give it.

Full Comment

Facebook and Instagram need to change policies in the way they treat users. For several weeks following the case of one of our clients, we have encountered an arrogance on the part of the company. When it comes to copyright infringement, Meta is very strict which is considered fair. But on the other hand, it does not apply equal rules to all. Sends automatic replies to appeals and is truncated in information. They refuse to give details, claiming to "preserve privacy" but on the other hand after many attempts give it. Recently our company is pursuing an issue for reactivating the Instagram account of an online magazine. What we have encountered is that Instagram gives automatic answers which are created by robots despite claiming that the issue goes to an operator. The answers are the same, and after each communication, none of them is personalized nor does it respond to a logical flow. On the other hand, in our case related to the blocking of a copyright Instagram account retains an unacceptable form of discretion. They claim to e-mail any details regarding the reports. Meanwhile, on the other hand, they do not give details. We have tried to get it through several applications in various online forms. After many attempts, we managed to get what we were looking for because the first

answers were the same "your account is disabled and can not be restored". To make the appeal mainly requires you to log in, while on the other hand does not allow you to log in. Agents can not provide solutions and the story goes on again from the beginning. Regarding Copyright, META in its two companies, Instagram and Facebook must implement a fair policy regarding the removal of content as is done on YouTube. The user is informed that he is using a video that has copyright and then he decides to remove it according to his right. On the other hand, in reporting copyrighted materials, Meta should provide a form of contact with the party who posted the content, allowing the reporter to have a personal dialogue with the publisher of the content and reach a solution.

Link to Attachment
No Attachment

PAO-2021-02

Case number

PC-10352

Public comment number

United States and Canada

Region

Mette

Commenter's first name

Bryans

Commenter's last name

English

Commenter's preferred language

N/A

Organization

No

Response on behalf of organization

————

Short summary provided by the commenter

I have no idea whether or not a cross check will improve compliance with Facebook's community standards. What I do know is that Facebook is a large platform for sharing information, and it must take a much firmer grip on the mis/disinformation shared by its users. American democracy literally depends on it.

Full Comment

I have no idea whether or not a cross check will improve compliance with Facebook's community standards. What I do know is that Facebook is a large platform for sharing information, and it must take a much firmer grip on the mis/disinformation shared by its users. American democracy literally depends on it.

Link to Attachment
No Attachment

PAO-2021-02

Case number

PC-10358

Public comment number

United States and Canada

Region

Chris

Commenter's first name

Riley

Commenter's last name

English

Commenter's preferred language

R Street Institute

Organization

Yes

Response on behalf of organization

————

Short summary provided by the commenter

This comment evaluates the context of cross-check, concluding that such a system can add value; but promoting trust requires substantially more transparency and a conscious focus on balance, rather than pursuing impossible standards such as a total absence of bias or the full implementation of often in-tension values of free expression and robust content policy enforcement.

Full Comment

See attached.

Link to Attachment
[PC-10358](PC-10358)

PAO-2021-02	PC-10364	Asia Pacific and Oceania
Case number	Public comment number	Region

Rita	Jabri Markwell	English
Commenter's first name	Commenter's last name	Commenter's preferred language

Australian Muslim Advocacy Network (AMAN)	Yes
Organization	Response on behalf of organization

————

Short summary provided by the commenter

Our comment responds to each of the questions provided by the Oversight Board. In addition to recommending what Meta should make transparent with its current cross-system system, we propose a model for part of a false-negatives cross-check system. Our proposal draws from a significant amount of original research and dialogue aimed at addressing the dehumanisation of minorities on Meta.

Full Comment

See attached as when I paste, it does not fully come across. The document is 6 pages.

Link to Attachment
[PC-10364](PC-10364)

PAO-2021-02	PC-10365	Central and South Asia
Case number	Public comment number	Region

Torsha	Sarkar	English
Commenter's first name	Commenter's last name	Commenter's preferred language

Centre for Internet and Society	Yes
Organization	Response on behalf of organization

————

Short summary provided by the commenter

Firstly, the Board should investigate the cross-check system as part of Meta's larger problems with algorithmically amplified speech, and how such speech gets moderated. Secondly, the Board must consider the cross-check system against Meta's commitments towards human rights, as outlined in its Corporate Human Rights Policy. Thirdly, the Board must urge Meta to adopt and implement the Santa Clara Principles on Transparency and Accountability to ensure that it is open about risks to user rights when there is involvement from the State in content moderation. Finally, the Board must investigate the potential conflicts of interest that arise when Meta's policy team has any sort of nexus with political parties.

Full Comment

Please refer to the attachment for the full text of the comments.

Link to Attachment
[PC-10365](PC-10365)

PAO-2021-02

PC-10366

Latin America and Caribbean

Case number

Public comment number

Region

Lucas Henrique

Muniz da Conceição

English

Commenter's first name

Commenter's last name

Commenter's preferred language

Institute for Technology and Society of Rio | ModeraLab

Yes

Organization

Response on behalf of organization

————

Short summary provided by the commenter

This public comment addresses the Board's investigation towards whether a system like cross-check on Meta's social media platforms is legitimate, enhancing freedom of speech and the coherent application of standards. The comment will specifically address the Early-Response Secondary Review system, and firstly consider its role in avoiding false positives. Secondly, it addresses the harms imbued in the false negatives the system may incur. Thirdly, the comment provides suggestions regarding possible measures the Board could advise the company into implementing to ensure that this system remains accountable and transparent, avoiding reliability issues regarding its neutrality.

Full Comment

After the Wall Street Journal disclosure of internal documents concerning the seemingly inconsistent way Meta makes content moderation decisions with the assistance of its cross-check system, the Oversight Board accepted Meta's request for a Policy Advisory Opinion. This public comment argues that, beyond Meta's concern with its business partners, the legitimacy of a cross-check system is inherently bound to its capacity to enhance freedom of speech and avoid the harms caused by false-positives. In addition, the comment tackles the issue of false-negatives, suggesting that further data could assist the Board in its investigation of the enforcement consistency of Facebook's and Instagram's standards for users enlisted in the cross-check system. Finally, the comment provides suggestions for transparency measures that could enhance the reliability of the system itself, specifically considering the perceived neutrality of Meta's content moderation

practices. Summary of Suggestions: We suggest that further enhancement to the system's reliability could be achieved by the deployment of (i) a clear parameter on the identification of false-positives and false-negative risks; (ii) independent and transparent procedures for the definition of the users and entities enlisted into the system, and (iii) express definition and limitation of the tiers of review to which content with an alleged violation should be reviewed. Regarding the first suggestion, the Board should consider recommending Meta to be more transparent with outcomes of its internal research concerning false positives, and correlate those results with further risk-based diagnoses of false negatives, especially in user accounts and entities that engage with a larger audience. Concerning our second suggestion, Meta has already conveyed that the compilation of the list of users assigned to Early Response secondary review is designated only to a group of employees with authority to make additions and removals of users and entities accounts. This information can be of significance for developing the Board's scrutiny over the content moderation decisions from Facebook and Instagram. The suggestion considers the need for higher transparency to the guidelines that inform the decisions to include certain accounts and pages into the system. Regarding the third suggestion, Meta describes the definition of three different tiers of review, considering contract reviewers (low), internal market teams with regional expertise (medium), and reviewers with deeper policy expertise (high), according to a risk framework in terms of content. The division of different tiers of review within the system appears to showcase a logic of contextualization and specialization, in which the more complex and sensitive the cases are, the higher the tier of scrutiny it goes. The suggestion is that escalation should follow as well in terms of understanding and knowledge of context. Company's leadership, although important for taking a final decision, may not necessarily be well equipped to have final say as in many cases, it may be detached from the expertise criteria and localized context that guided the original assessments. In this scenario, a further suggestion may be that the Board request Meta, where such escalation is inevitable, to request leadership to present the rationale for its final decision in lieu of that of its policy advisors. All in all, our suggestions to the Board is to consider the impact that a system of cross-check may have on freedom of expression online. This correlates with noted risks that false positives may pose on the content moderation process of users and pages with higher following. Nevertheless, we also suggest that a similar concern must be given to false negatives as the Early-Response Secondary review system may result in an inconsistent application of Facebook's and Instagram's community standards, impacting other human rights. We believe that with further transparency measures, the system's reliability can be strengthened, and other shortcomings may be addressed. [Please refer to the attachment for the full comment]

Link to Attachment
[PC-10366](PC-10366)

PAO-2021-02   PC-10367   United States and Canada

Case number   Public comment number   Region

Adam   Schiff   English

Commenter's first name   Commenter's last name   Commenter's preferred language

House of Representatives   No

Organization   Response on behalf of organization

––––

Short summary provided by the commenter

We are writing as Members of Congress to share our concern with Meta's Cross-Check program.

Full Comment

To the Facebook Oversight Board, We are writing as Members of Congress to share our concern with Meta's Cross-Check program. As you are aware, the Wall Street Journal first reported on the Cross-check, or XCheck program in September 2021, after reviewing internal Facebook, now Meta, documents. Meta published further information on the XCheck system in October, and we understand that the Facebook Oversight Board ("Board") has agreed to review this case and offer recommendations on the program to Meta. We now know that the XCheck program was allegedly intended to mitigate excessive reporting of content posted by higher risk accounts, including celebrities, politicians, and journalists. We also understand that these high-profile figures can play an outsized role in amplifying content and may require oversight at a higher level, but we believe the program clearly requires further regulation and has not been working effectively or as intended. In reality, the internal documents indicate that over 5.8 million users may have been exempted from Meta's enforcement process since the program began in 2013. These "VIP" individuals were allowed to post without the same rules or consequences as an average user. When protected by XCheck, users' posts can be shielded from moderators, even when the information in the post has been deemed false by Facebook's own fact checkers. The XCheck program has enabled the perpetuation of false claims and statements that are detrimental to our nation's health and security, including vaccine misinformation and incitements of violence. An internal summary of the XCheck program noted that the program allowed posts that violated its rules to be viewed at least 16.4 billion times. Additionally, the company found

that only 10% of XCheck content was being reviewed, and although Meta has said that number increased in 2020, the company did not provide further details. The internal documents also indicate that the XCheck system was in place during the 2020 elections. This is of great concern to Members of Congress, as we understand the effects that this lack of review and accountability can have on politicians and elections. We witnessed what the former president was able to say in direct violation of Facebook community standards because he was a "VIP." Moreover, not all candidates for public office were "VIPs," giving certain candidates a potential advantage on Facebook over others. We strongly recommend that the XCheck program treat all major candidates for office and office holders equally. Furthermore, we were alarmed to learn that when the Board investigated the case regarding President Trump's Facebook account, Meta left out important and necessary information about the XCheck system. In the Board's report, you stated that "given that the referral included a specific policy question about account-level enforcement for political leaders, many of whom the Board believes were covered by XCheck, this omission is not acceptable." We agree with the Oversight Board, and think it is deeply troubling that Meta withheld information, particularly at such an important moment in time. We ultimately believe that the program should be significantly reformed to ensure accountability. We believe that allowing powerful actors to be exempted from the rules is unacceptable. We did not find Meta's October explanation of the XCheck system to be sufficient and without further exculpatory details, we find the program to be deeply concerning and cannot support it. We believe that penalties for violating Meta's community standards should be the same for all, regardless of status or political office. We also believe that no one should be allowed to knowingly perpetuate dangerous falsehoods, especially when the consequences can be so drastic. Thank you for fully and fairly taking our opinions into consideration when you make your recommendations to Meta on the XCheck system. Sincerely, _____ _____
Adam B. Schiff Lori Trahan Member of Congress Member of Congress
_____ _____ _____ Jake Auchincloss Sean Casten Mark DeSaulnier Member of Congress Member of Congress Member of Congress _____ _____ _____ Anna G. Eshoo Jahana Hayes James P. McGovern Member of Congress Member of Congress Member of Congress _____ _____ _____
Mary Gay Scanlon Adam Smith Jackie Speier Member of Congress Member of Congress Member of Congress _____ _____ _____ Eric Swalwell Marc Veasey Peter Welch Member of Congress Member of Congress Member of Congress

Link to Attachment
[PC-10367](PC-10367)

PAO-2021-02

PC-10368

United States and Canada

Emma

Llanso

English

CDT

Yes

————

CDT recommends that Meta use the GSR program to help identify problems with its content moderation system more generally, and adapt its ERSR program to consider the risk that false-negative decisions on high-profile individuals' accounts pose to other users or the general public (for example, false-negative decisions on incitement to violence).

The Center for Democracy & Technology welcomes the opportunity to provide comments on Policy Advisory Opinion 2021-02, regarding Meta's policy on cross-checking the moderation decisions for certain high-profile or influential accounts. Below, we respond to two of the questions posed by the Oversight Board in its call for public comment. Whether a cross-check system is needed and if it strengthens or undermines the protection of freedom of expression and other human rights. In general, yes, a cross-check system can provide an important opportunity for close evaluation of the ways in which Facebook and Instagram's rules apply to powerful, high-profile, or at-risk users. Content moderation systems will always be prone to error, given the sheer volume of content uploaded to online services and the complexity of evaluating human communication against a single set of rules. Having a risk-based model that helps to prioritize certain content for additional review is, in general, a useful safeguard that can help avoid erroneous enforcement without obliging the user to take on the burden of appealing the company's decisions. Meta's cross-check program, however, has suffered from two fundamental flaws: its application only to high-profile individuals and its focus exclusively on evaluating "false positives". In combination, those flaws have led to

the creation of essentially two tiers of Facebook/Instagram usage, where already powerful or influential users were more likely to have their speech remain on the service than regular users who did not benefit from the cross-check system. More influential users were shielded from false-positive removal of their content, including in circumstances where Meta services' policies were unclear or the determination of whether to remove content was a close call, while regular users were not. Meta has begun to address this disparity by making cross-checking for false positives available to all users, in the General Secondary Review program. The kinds of errors that Meta's moderators or technical moderation tools may make will not be unique to high-profile or powerful users, and the harms of over-removal are not felt only by those with the largest audiences. It will be crucial for Meta to continuously evaluate that program to understand how it is being applied to and experienced by users from a variety of different backgrounds and who post a variety of types of content. This evaluation should include consultation with civil society organizations that represent the interests of users in different regions and cultural contexts, so that Meta can better understand the consequences and real toll of over-removal on regular users and whether the cross-check ranker is effectively identifying content that should undergo a General Secondary Review. Meta should make improvements to the General Secondary Review program based on this evaluation and also commit sufficient personnel and resources to ensure that the Secondary Review program provides a meaningful check for all users and does not function as a fig leaf that justifies the cross-check program for powerful users. In addition to this broadening of the scope of the false-positive review in its General Secondary Review program, Meta should also implement a false-negative review for its high-profile users (in what is now called the "Early Response Secondary Review" system). The focus on false-positive review in its cross-check system has shielded Meta from a risk of high-profile public criticism for overbroad content removal, but it does not address the risk of abuse of Meta's services by high-profile individuals. As discussed below, the issue of false-negative decisions is a vital element of how Meta moderates content posted by high-profile individuals and has significant consequences for the overall fairness of the content moderation on Meta services and the societal impact of high-profile individuals' speech. Cross-check is designed to be a "false positive" prevention mechanism. What are the checks and balances, if any, this system should contemplate to mitigate the risks of "false negatives" [erroneous lack of action on violating content]? Meta should incorporate a check for false-negative results, or erroneous decisions not to remove or otherwise action content, into its Early Response Secondary Review process. According to Meta, the 660,000 users and entities who currently undergo ER Secondary Review represents users in the following categories: "elected official, journalist, significant business partner, human rights organization", as well as those with a large number of followers or those who post about sensitive topics. The ER Secondary Review list presents a useful starting point for Meta to more carefully consider how high-

profile and powerful individuals may be abusing their services and causing real-world harm, even if there will not be perfect overlap between users facing a high risk of false-positive results and those facing a high risk of false-negative results. For example, as CDT discussed in our comments to the Oversight Board in its consideration of the suspension of Donald Trump's account, there are certain categories of policy violation in which false-negative moderation decisions pose a heightened risk to public safety. Specifically, in consideration of the risk of incitement to violence, we urged the Oversight Board to consider the six-part threshold test from the Rabat Plan of Action, which states that "the speaker's position or status in society" is a necessary element to the assessment of whether a speaker's statement is likely to incite violence. We urged that Facebook should " develop or improve their internal escalation process for moderation decisions about content posted by political leaders. . . . Given the risk and intensity of political violence around the world [This text box appears to cut off at ~6000 characters; full text of comments available in the attachment.]

Link to Attachment
PC-10368

PAO-2021-02

Case number

PC-10369

Public comment number

United States and Canada

Region

Matt

Commenter's first name

Bailey

Commenter's last name

English

Commenter's preferred language

PEN America

Organization

Yes

Response on behalf of organization

————

Short summary provided by the commenter

We argue for reforms to cross-check from a process standpoint and further outline specific reforms to establish meaningful transparency and accountability of cross-check to the public, including the establishment of routinized independent third-party audits and significantly increased public reporting and access to data for credentialed researchers.

Full Comment

Please see attached PDF.

Link to Attachment

[PC-10369](PC-10369)

PAO-2021-02

Case number

PC-10371

Public comment number

Europe

Region

Withheld

Commenter's first name

Withheld

Commenter's last name

English

Commenter's preferred language

Mnemonic

Organization

Yes

Response on behalf of organization

————

Short summary provided by the commenter

Stopgap solutions to bad content moderation **enforcement aren't good enough.** Meta should not invest significant financial and people resources in bandaids like **Meta's cross-check system (Xcheck) when wholescale improvement** of Meta's content moderation policies and systems is needed. This is particularly true for at-risk countries and non-English moderation, especially of Arabic and other non-**Latin languages. Fixing Meta's "language problem" would improve moderation for** all users, not just those lucky enough to make it on the Xcheck list. That being said, Xcheck can be improved with more transparency as well as significant outreach with civil society to ensure ranking factors and the data used to train any automated systems are high quality.

Full Comment

You don't appear to have corrected this form for the longer length of this comment, so I'm skipping this step and just attaching my comment, which is 13711 characters including spaces, and 5 pages long in 12 pt Times with regular margins.

Link to Attachment
[PC-10371](PC-10371)

PAO-2021-02

PC-10372

United States and Canada

Sharon

Kann

English

Media Matters for America

Yes

————

The Board has asked respondents for comments on Meta's cross-check system and recommendations to ensure that it is neutral and free of political and other biases. However, there will be no neutral cross-check system so long as Meta continues to prioritize and incentivize growth and positive publicity over solving serious issues that plague its platforms — including misinformation and hate speech that has caused real-world harm. Given the immense scale of Meta, with more than 3 billion people using its platforms, and its vast societal influence, the company should be dedicated to minimizing the risk of harm, both on and off the platform.

The Board has asked respondents for comments on Meta's cross-check system and recommendations to ensure that it is neutral and free of political and other biases. However, there will be no neutral cross-check system so long as Meta continues to prioritize and incentivize growth and positive publicity over solving serious issues that plague its platforms — including misinformation and hate speech that has caused real-world harm. Given the immense scale of Meta, with more than 3 billion people using its platforms, and its vast societal influence, the company should be dedicated to minimizing the risk of harm, both on and off the platform. The Board is seeking remedies to equalize a system that is built on a flawed foundation with an incentive structure that consistently places profit over people. Meta continues to perpetuate a culture of opacity, often misleading the public to avoid accountability and scrutiny. Neither the Board nor the public was even aware of Meta's cross-check program until disclosures from the Facebook Papers, and what is now known

suggests the company is ignoring the troves of data that it does have. Similarly, an internal memo from Meta employees detailed the amount of data the company had about the spread of the "Stop the Steal" campaign prior to the violence on January 6, including data on group membership overlap, super-inviters, direct coordination, growth and amplification of the campaign, and rates of hate speech and incitement. There is no effective cross-check system that would be neutral and free of bias so long as Meta continues to prioritize and incentivize growth and positive publicity. In its current form, the system is designed to be a "false positive" prevention mechanism, which completely ignores "false negatives" or violative content from public figures and politicians that evades moderation. While there is limited public knowledge of the cross-check system — which is consistent with Meta's usual lack of transparency — reporting about the program suggests that it has enabled public figures and politicians to push misinformation with impunity. The program reportedly grew to include at least 5.8 million users in 2020 that included "pretty much anyone regularly in the media or who has a substantial online following, including film stars, cable talk-show hosts, academics and online personalities with large followings." Some users were allowed to post rule-violating material pending Facebook employee reviews that often never came and some users were even "whitelisted," or rendered immune from enforcement actions. Facebook's concern for "false positives" but not "false negatives" allowed right-wing media outlets, media personalities, and politicians — all likely covered under Meta's cross-check program — to freely post misinformation, including about COVID-19, coronavirus vaccines, and 2020 elections. Many of the failures of Meta's cross-check system — and the company's broader approach to content moderation — are exemplified by its complete lack of action against former President Donald Trump's Facebook page, even as he repeatedly pushed misinformation, extreme rhetoric, and incitement of violence. In reality, Meta actually profited off of his misinformation and, thus, allowed him to violate policies with impunity. In fact, since Facebook started reporting advertiser spending on ads related to social issues, elections, and politics in May 2018, Trump and his campaign spent more than $113 million on ads. Trump also directed users to other misinformation spreaders that make Facebook money, putting him at the center of Facebook's lucrative right-wing misinformation ecosystem. In addition to pushing misinformation in ads, Trump used his Facebook page to share misinformation to his tens of millions of followers. Between November 3, 2020, and January 7, 2021, when Trump was banned from posting on Facebook, he posted 433 times with election-related keywords, earning over 176 million interactions, even as 75% of the posts featured at least one of Facebook's 18 different labels related to the election. Additionally, roughly a quarter of Trump's 6,081 Facebook posts that were created between January 1, 2020, and January 6, 2021, contained COVID-19 misinformation, election lies, or extreme rhetoric about his critics. In the 24 hours prior to the January 6 insurrection, Trump's top 10 most-engaged posts on Facebook earned more than 5.5 million total interactions, with

many of the posts containing misinformation about the election. After the Board upheld Meta's decision to restrict Trump's access to Facebook and Instagram, the company suspended him from the platform for two years. During this suspension, Facebook has left open loopholes that allow Trump to push his messaging on the platform, including letting him run ads for his joint fundraising committee and allowing his page to continue to earn interactions. As the Board considered the case involving Trump last year, the company failed to be forthcoming about the cross-check program, only providing information after the program was uncovered through public reporting. The lack of transparency, even with the Board, is part of Meta's broader culture that also fails to enforce its policies, capitulates to right-wing media and politicians to avoid false allegations of censorship and bias, uses an escalation process that enables executives to make sole moderation decisions, and allows right-wing outlets and conservative politicians to push misinformation with little moderation. Meta's broader culture Repeatedly capitulating to right-wing media and politicians to avoid false allegations of censorship and bias Meta's current systems are plagued with political biases that inevitably result in capitulation to right-wing media and conservative politicians in order to avoid their false allegations of censorship and bias. FULL TEXT IS IN THE ATTACHMENT

Link to Attachment
[PC-10372](#)

PAO-2021-02

Case number

PC-10373

Public comment number

Latin America and Caribbean

Region

Vladimir

Commenter's first name

Chorny

Commenter's last name

Commenter's preferred language

R3D: Red en Defensa de los Derechos Digitales

Organization

Yes

Response on behalf of organization

————

Short summary provided by the commenter

El trato desigualitario de usuarios dentro de la red social Facebook muestra una tensión significativa entre la moderación de contenidos y la libertad de expresión. La necesidad de gestión en la información y el manejo de las cuentas de usuarios dentro de la plataforma muestra una complejidad hasta ahora minimizada o ensombrecida que enfrenta el interés de la flexibilidad empresarial (como Meta lo denomina) con los intereses de los derechos humanos y las obligaciones que estos generan a los entes privados al relacionarse con los usuarios en Internet. Es en este sentido que damos respuesta a algunas de las cuestiones planteadas por el Consejo Asesor de Facebook, con la intención de contribuir a la solución del problema.

Full Comment

R3D: Red en Defensa de los Derechos Digitales Ciudad de México, México Asunto: **Comentarios para la Opinión Consultiva sobre el programa "Cross Check"** Introducción El trato desigualitario de usuarios dentro de la red social Facebook muestra una tensión significativa entre la moderación de contenidos y la libertad de expresión. La necesidad de gestión en la información y el manejo de las cuentas de usuarios dentro de la plataforma muestra una complejidad hasta ahora minimizada o ensombrecida que enfrenta el interés de la flexibilidad empresarial (como Meta lo denomina) con los intereses de los derechos humanos y las obligaciones que estos generan a los entes privados al relacionarse con los usuarios en Internet. El problema se vuelve particularmente relevante en el contexto de la pandemia de COVID-19 en los últimos años, en el que por razones de limitaciones de personal y de aislamiento domiciliario se hicieron claras las limitaciones de la moderación

automatizada y de la capacidad de respuesta humana a las apelaciones de usuarios que consideraron lesionado su derecho a la libertad de expresión. Esta situación pone en entredicho la capacidad de dar cumplimiento a principios fundamentales para el ejercicio de los derechos humanos en Internet tales como el debido proceso, la transparencia y la propia libertad de expresión. La preocupación de R3D surge de los hechos relacionados con las filtraciones del sistema Cross-Check que son de público conocimiento, pero también es resultado de las reuniones que Facebook ha llevado con un grupo de organizaciones de la sociedad civil defensoras de derechos digitales en los últimos años, en los que hemos dado seguimiento a las políticas desarrolladas por la plataforma y a sus limitaciones, en particular a partir de que reconociera abiertamente que durante la pandemia muchas de las personas usuarias perdieron la posibilidad apelar su moderación de contenidos por las circunstancias arriba mencionadas, mientras que los usuarios privilegiados por el sistema de revisión cruzada tuvieron a su alcance equipos especializados para atenderlos o simplemente no se les aplicaron las normas comunitarias, resultando en un trato discriminatorio a las mismas. Es en este sentido que damos respuesta a algunas de las cuestiones planteadas por el Consejo Asesor de Facebook, con la intención de contribuir a la solución del problema. Si se necesita un sistema de verificación cruzada, y si este incrementa o socava la protección de la libertad de expresión y otros derechos humanos. La compatibilidad de los sistemas automatizados de verificación cruzada con la libertad de expresión y otros derechos depende de su diseño institucional y de su implementación particular. En este caso concreto consideramos que es incompatible con los derechos humanos utilizar un sistema que incorpora factores comerciales que tiran en dirección contraria a la protección igualitaria de los derechos y que permite decisiones arbitrarias tomadas por el liderazgo de la empresa que pasan por alto sus propias normas comunitarias y que son contrarias al principio del debido proceso. La verificación cruzada se diseñó como un mecanismo de prevención contra los "falsos positivos". ¿Qué controles y contrapesos debería contemplar este sistema para mitigar el riesgo asociado a "falsos positivos" (ausencia errónea de medidas ante contenido infractor)? Consideramos que existen tres tipos de controles que pueden ayudar a la mitigación de riesgos en este sentido. El primero debería permitir el acceso al estudio externo de los sistemas automatizados con los que la plataforma construye sus procesos de revisión cruzada. La posibilidad de revisar posibles sesgos en su construcción que puedan derivar en tratos discriminatorios es fundamental y es a su vez una medida que fortalece la transparencia del sistema de revisión. Al mismo tiempo, Facebook debe permitir la fiscalización y análisis de los procesos de identificación de contenidos que infringen las normas comunitarias para poder evaluar su precisión y la calidad de los sistemas mismos. Si bien Facebook ha proporcionado los números agregados en sus informes periódicos y los ha presentado a la sociedad civil, estos números son solamente los resultados totales que la plataforma presenta, sin dar la posibilidad de conocer lo que los sistemas

hacen en primer lugar. Todo esto es insuficiente porque no permite evaluar la forma en que los sistemas automatizados funcionan ni ver los posibles sesgos y problemas en términos de derechos humanos que estos podrían tener. El segundo debe ser un control que garantice la inclusión de las posibles personas o grupos afectados por la moderación de contenidos. La garantía del debido proceso en la moderación de contenidos es básica y fundamental para que los derechos de las personas no sean vulnerados en las plataformas, particularmente ante la tensión irresoluble de la revisión automatizada sobre los contenidos y la necesidad de evaluar el contexto en el que una expresión o un contenido es difundido. Este requisito es particularmente importante a la luz de la variabilidad cultural que puede existir entre un contexto y otro, y en los distintos tipos de acciones con los que puede tomar forma una expresión determinada (por ejemplo, una protesta social y el uso de videos que den evidencia de represión policial de la misma). La inclusión también está relacionada de lleno con los procesos de apelación en las plataformas, en los que debe garantizarse que exista la revisión humana sobre "cualquier proceso automatizado, incluyendo la posibilidad de los usuarios de buscar la revisión humana de cualquier decisión de moderación de contenidos automatizada". Si bien un paso para incorporar el contexto en el que se emite una expresión pasa por tener equipos que hablen el idioma y estén familiarizados con el contexto cultural o regional, un elemento democratizador de las decisiones de moderación de contenidos puede encontrarse en la in

Link to Attachment
[PC-10373](PC-10373)

PAO-2021-02

Case number

PC-10374

Public comment number

Latin America and Caribbean

Region

Barthélémy

Commenter's first name

Michalon

Commenter's last name

English

Commenter's preferred language

Sciences Po / Tec de Monterrey

Organization

No

Response on behalf of organization

————

Short summary provided by the commenter

The recent evolutions in the cross-check system are welcome developments. This trend should continue and some major evolutions are still desirable: the recently introduced General Secondary Review should end up replacing fully the former system, false negatives detection should be added to it, the transparency requirements should fully apply to the whole mechanism and new language-specific cross-checking centers should be created as soon as a certain volume of users speaking this language is reached. The Oversight Board could also play a role in assessing the balance of the decisions taken through this system, from a randomly-defined and automatically-transmitted sample of cases that went through Cross Check.

Full Comment

My comments are in the attached document

Link to Attachment
[PC-10374](PC-10374)

PAO-2021-02

Case number

PC-10375

Public comment number

United States and Canada

Region

Ken

Commenter's first name

Buck

Commenter's last name

English

Commenter's preferred language

Rep. Ken Buck (R-CO)

Organization

Yes

Response on behalf of organization

————

Short summary provided by the commenter

Recognizing every person's inherent equality is at the cornerstone to every other right, including freedom of speech and expression. Rep. Buck's public comment demonstrates that Facebook's cross-check system unavoidably violates this principle of equality. While Facebook should rightly be concerned about curtailing the spread of illegal, exploitive, and abusive content on its platforms, violating the most foundational principles of human rights and dignity can never pave the path toward greater freedom. Rep. Buck urges the Oversight Board to ensure the expedient and complete termination of the cross-check system.

Full Comment

Dear Members of the Oversight Board, I write in response to the Board's request for public comment on Policy Advisory Opinion 2021-02 regarding Meta's cross-check system. These comments address the primary issue raised by the Board: 1. "Whether a cross-check system is needed and if it strengthens or undermines the protection of freedom of expression and other human rights." I recognize that moderating billions of users' posts is a herculean task. However, Facebook has failed to establish neutral standards and consistent enforcement practices for content posted on the site. This has increased public distrust toward Facebook and undermined the principles of free speech and access to information that Facebook claims to support. The basic concept of Facebook's cross-check system inherently contradicts the platform's core values. According to Facebook's website, the company's guiding principles include "Giving People a Voice," Serving Everyone," and "Keeping People Safe and Protecting Privacy." By recognizing some users'

content as more worthy of review, Facebook has distinguished these users as more important or valuable than the billions of other people who use the platform. Facebook appears to have gone even further than this, though, by establishing "whitelists" of users whose content skirts human and algorithmic scrutiny. In the face of these realities, Facebook's claim to give people a voice, "even when that means defending the right of people we disagree with," misses the mark. Instead, the cross-check system appraises people's worth, including whether their voices have a right to exist on Facebook, based on their ranking within the platform's dystopian class system. Further, while Facebook purports to serve everyone, this too is subject to an individual's status. The cross-check system unavoidably generates a caste system that applies arbitrary and biased criteria to determine users' ability to freely express themselves on the platform, the level of scrutiny applied to their content, and even their access to appeals mechanisms. Publicly, Facebook claims to adhere to and support human rights, but its internal policies and practices are distinctly misaligned with these principles. The Declaration of Independence articulates the most foundational natural right, the self-evident truth that all people are inherently created equal. Immediately following this assertion, the Declaration enumerates that the rights to life, liberty, and the pursuit of happiness are unalienable. Each of these rights are impossible to practice without the recognition of humanity's shared equality. However, while these rights are naturally endowed within every person, countless people globally are unable to express and practice these rights. Discrimination against any person because of some immutable characteristic is a violation of this principle. Irrespective of the logistical benefits that Facebook sees in its cross-check system, it is impossible to avoid the reality that the system cannot exist without adjudicating users – people – based on their number of followers, accomplishments, or some other non-objective criteria. Already, Facebook's cross-check system has left a trail of victims by protecting whitelisted users' abusive content from being flagged for removal, based solely on the individual's elite status. As the Wall Street Journal reported, Facebook's immunity program allowed revenge porn to not only be posted to the platform, but to remain online and accessible for tens of millions of viewers before Facebook finally removed the content. Tragically, Facebook has provided an oasis for criminal organizations, contributed to deteriorating mental health among child and teenage users, and undermined democracy in the United States and abroad. Despite knowing about each of these problems, Facebook failed to address them, and in some cases actively perpetuated them. While Facebook should rightly be concerned about curtailing the spread of illegal, exploitive, and abusive content on its platforms, violating the most foundational principles of human rights and dignity can never pave the path toward greater freedom. Instead of leveraging Facebook's influence to advance human rights and freedom, it has implemented programs like cross-check that sabotage the very values and systems that empowered the company's growth. Cross-check is built on a values system that is

incompatible with people's basic right to equality, and I strongly urge the Board to ensure the expedient and complete termination of this system.

Link to Attachment

[PC-10375](#)

Oversight Board
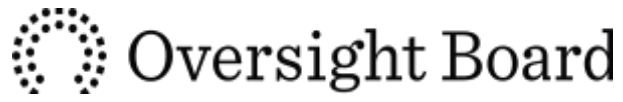
Public Comment Appendix for
PAO-2021-02
Case number

End of public comments