Safety at Stake:

How to Save Meta's Trusted Partner Program

Internews

July 2023

About Internews

Internews is an international non-profit that supports independent media in 100 countries — from radio stations in refugee camps, to hyper-local news outlets, to filmmakers and technologists. Internews trains journalists and digital rights activists, tackles disinformation, and offers business expertise to help media outlets thrive financially. For 40 years, it has helped partners reach millions of people with trustworthy information that saves lives, improves livelihoods, and holds institutions accountable.

Contents

Background

"The Trusted Partner Channel saves lives. It is literally a lifeline. And right now it is broken."
– Meta Trusted Partner

"I wish we'd never joined to be honest. It is worse than before. Often, we never receive any response, and when we do it can be five months later. It is worse than nothing."
– Meta Trusted Partner

"Trust? What Trust? They don't trust us, and so we don't trust them. There is no trust."
– Meta Trusted Partner

Meta's Trusted Partner program is a critical tool for keeping Meta's products safe and protecting users from harm. According to Meta, the Trusted Partner program comprises of "expert organisations that represent the voices and experiences of at-risk users around the globe and are equipped to raise questions and concerns about content on Facebook and Instagram." Meta's Trusted Partners help alert the company to dangerous content, harmful trends, and other online risks, providing invaluable contextual, cultural, and linguistic expertise in countries and regions where Meta often lacks deep local knowledge. Meta also relies on its Trusted Partner network to provide insights and feedback into content moderation policies and practices, stating that "our trusted partners subject matter and regional expertise help strengthen our policies by enabling us to consider a range of perspectives to inform our content moderation efforts."

As of February 2023, there are 465 organizations included in Meta's Trusted Partner program, representing 122 countries. The profile of Trusted Partner participants varies greatly, ranging from grassroots civil society organizations operating in repressive or conflict affected states, to international human rights groups. Trusted Partners participate in the program voluntarily and are not compensated for their time by Meta.

Trusted Partners report issues to Meta via a designated channel, separate from the reporting channels available to regular users. The range of issues that partners report through the Trusted Partner channel vary greatly, and can include death and rape threats, hacked accounts, impersonation of activists or journalists, deactivation of detained individuals accounts, and harmful content such as hate speech or incitement to violence. The Trusted Partner channel is also used to report violations of Meta's 'escalation only' policies, such as the 'misinformation and harm' policy, which relies explicitly on Trusted Partners to assess when content is both false and has the potential to cause imminent harm. Policies such as that around 'misinformation and harm' allow Meta to take action on content that may otherwise be deemed non-violating of Meta's content policies, based in part on the expert judgement of context experts such as Trusted Partners. Important but potentially less urgent issues raised through the Trusted Partner channel include matters such as incorrect enforcement of Meta policies, or account recovery for at-risk users or key pages.

The Trusted Partner program is just one element of Meta's broader content enforcement, which includes reporting functions available to general users and automated detection of harmful content. Programs such as the Trusted Partner program are of the greatest importance in countries and contexts where general user reporting is infrequent or of poor quality, and where Meta's automated systems lack language capacity. In addition to the Trusted Partner program Meta receives escalations through other channels, including a 'human rights defender response channel', or third party channels such as Access Now's Digital Security Helpline. In some cases, government and law enforcement agencies may also be able to escalate issues to Meta staff. Each of these channels also warrant scrutiny and review, however they fall outside of the scope of this report.

Since at least 2018 (when Internews was first onboarded as an early participant) Meta's Trusted Partner program has had consistent issues with responsiveness, accessibility, transparency and accountability to its partners. ome Trusted Partner reports never receive any response from Meta. When responses are received, response times vary from a single day to as long as eight months for ostensibly similar reports, with no explanation or acknowledgement given by the company. Typically, no explanation is provided to partners for Meta's decision making on Trusted Partner reports, including when complex, 'escalation only' policies such as those relating to misinformation and harm may apply. Other significant and unresolved issues relating to the program include safety-risks and resource-burdens placed on Meta's Trusted Partners for the work they undertake, in some cases whilst operating in authoritarian or conflict affected countries.

Given the potentially life-and-death nature of the issues reported through the Trusted Partner program, it is notable that Meta has failed to meaningfully address these issues or consult widely and formally with the Trusted Partner organizations participating in the program. Internews has raised these issues with Meta regularly since 2019. In 2021 Internews proposed to conduct a joint review of the program in active collaboration with Meta, a proposal which Meta initially agreed to and then eventually (in 2022) declined.

After consultation with other partner organizations Internews chose to complete the review without Meta's formal participation, eventually conducting interviews and survey work with 24 of Meta's Trusted Partners (or just over 5% of the total partner base) representing every major global region. To ensure safety and protect privacy participating partners have been anonymized in the review, however they include representatives from multiple priority at-risk and conflict affected countries, as well as multiple countries which had major elections during the review period of 2022 and early 2023. For more information on how the Review was conducted see the Methodology section below.

This review is intended to be a positive and constructive step towards improving Meta's Trusted Partner program, ensuring it is resourced appropriately, and best placed to protect global platform user's safety into the future. Internews is an active partner of Meta's beyond the Trusted Partner program, and it is our intention to continue to work productively with the company to address the issues raised in this review (and beyond). The range of experiences shared by Trusted Partners during the review process was wide, reflecting the global and diverse nature of the participants, however there were some clear consistencies that emerged. These are outlined in the Key Findings and Recommendations sections below.

Key Findings

- Meta's Trusted Partner program is significantly under-resourced and understaffed, and has been further impacted by recent layoffs. Many of the most severe operational failures of the Trusted Partner program appear to relate directly to this lack of resourcing and could be remedied with sufficient investment and adequate staffing. This lack of resourcing undermines a critical program focused on user safety and platform integrity.

- Meta's response times to Trusted Partner reports are erratic and regularly run to weeks and often months. In some cases, no response is ever received. No explanation is provided for these extended wait times, which can sometimes apply even to highly time sensitive content such as incitement of violence. Meta has declined requests to provide information on average response times or internal targets.

- There is a significant disparity of service between Ukraine and other countries, including countries that are also experiencing major armed conflict, internal displacement, and political disinformation. Whilst Ukrainian partners can expect a response within 72 hours, in Ethiopia equivalent reports relating to the Tigray War can go unanswered for several months.

- The Trusted Partner channel is currently functioning as an emergency channel, even though it is not designed or resourced as one, and is clearly not fit for this purpose. This is because no dedicated emergency channel is available to Trusted Partners. Every single partner interviewed for this report who had previously used the Trusted Partner channel stated that they had used the channel to report an issue they believed would lead to imminent harm and required immediate action.

- Many Trusted Partners choose to supplement or bypass the official Trusted Partner channel, communicating directly with personal contacts at the company via WhatsApp or ignal, or CCing them into Trusted Partner reports to ensure that they are read. Partners who can leverage their personal connections in this way receive a much better response than those forced to rely on the dedicated Trusted Partner reporting mechanism. This is unfair, inefficient, and unsustainable.

- Meta's communication with Trusted Partners is impersonal and often alienating. Combined with the lack of responsiveness, this poor communication has contributed to an absence of trust and resulted in multiple partners ceasing to engage with the program or make use of the reporting channel.

- Participating in the Trusted Partner program places a substantial burden on partners. Partner organizations take on significant extra work, reporting issues they have identified, as well as serving as a conduit to Meta for many other organizations, human rights defenders, media outlets, or humanitarian actors who are unable to directly communicate with the company. In many contexts participation in the program increases the risk profile of the partner organization, as they may be reporting extremist content, operating in a conflict zone, or in an authoritarian or semi-authoritarian state.

- The design of Meta's reporting mechanism forces partners to use their personal platform accounts and restricts partner organizations to a limited number of whitelisted email addresses. This causes unnecessary frustrations and burdens for participating organizations and creates complexities when staff turnover or change roles. Currently there is no email bounce-back when reports are made from non-whitelisted email reports, which is likely also resulting in lost reports.

- There is a lack of clarity around operationalization of critical policies that rely on Trusted Partners, such as those relating to misinformation and harm.

- All partners agreed that given the added workload partners are voluntarily accepting to keep Meta platforms safe, they expect the Trusted Partner program to be sufficiently resourced and highly responsive to partner needs. Meta does not currently provide compensation for Trusted Partners.

- There is an urgent need for an overhaul of the Trusted Partner channel, and this should be done through a transparent co-design process with the partner organizations who make up the program. More detailed recommendations are below.

Methodology

This review was conceived in 2021 as a collaboration with Meta. Meta initially indicated agreement to this proposal, before eventually withdrawing their participation in 2022. After consultation with other Trusted Partners Internews made the decision to continue with this 'partner led' review without Meta's participation.

This review was conducted through a mixture of formal key informant interviews, surveys, and informal discussions. A total of 24 Trusted Partners contributed to the review, representing just over 5% of Meta's 465 global partners. Internews is a member of Meta's Trusted Partner program, and Internews' own experience with the program has been included in the review.

In order to ensure review participants felt comfortable speaking freely, all review contributors have been anonymized. Any names have also been removed from quotes or email correspondence included in the below review. While the lack of identification has significant drawbacks, the frankness of the resulting discussions justifies the decision.

To capture the global nature of Meta's Trusted Partner program, participants were included from every major global region. While the inclusion of two specific priority countries have been highlighted for the sake of comparison, Internews has elected not to otherwise list the countries included as this would compromise the anonymity of review participants, given Meta has at most a handful of partners in each country.

For important privacy and security reasons Meta does not publicly share its list of Trusted Partners. To reach a more representative sample, Internews requested Meta assist with notifying its partners of the review and providing them with Internews' contact information should partners wish to participate. Meta declined to notify its Trusted Partners of the review.

For this reason participation in this review was restricted to organizations which Internews has contacts with and knows to be part of Meta's Trusted Partner program.

In June 2022 Internews sent Meta a list of questions for this review, and Meta committed to providing a response. Meta eventually provided these answers in February 2023. The extended timeframe of the response significantly delayed the finalization of this report. Both the questions submitted by Internews and the full answers provided by Meta can be found in ANNEX 1.

In April 2023 Internews shared a draft version of this report with participating Trusted Partners as well as with Meta to gather comments and feedback and ensure accuracy to the best of our ability. In May 2023 Meta provided written feedback to the draft report, which is included in full in ANNEX 2 at the end of this report. Some feedback from both Meta and Trusted Partner reviewers was incorporated into the final version of this document.

Analysis

Response times and rates

While partners interviewed for this review raised a wide range of issues, the most urgent feedback is related to Meta's responsiveness to Trusted Partner reports.

When a report is submitted through the official Trusted Partner channel, Meta's response times are erratic. Partners reported that while in some cases a report was resolved within the same day, in others no response was received for weeks, or even months.

"It is so frustrating because sometimes you escalate something and it takes them more than a month to respond to you, and sometimes even more than that! And then sometimes they are responding in the same day to a similar report, it depends!" **– Meta Trusted Partner**

Almost all partners interviewed for this report said that they had received responses more than a month after submitting a report, while more than a third said that they had received responses multiple months after the initial submission.

"Two months plus. And in our emails we tell them that the situation is urgent, people are dying, the political situation is very sensitive, and it needs to be dealt with very urgently. And then it is months without an answer."
– Meta Trusted Partner

"Sometimes it is three to six months without any reply at all. By the time you get the response sometimes you can't remember what it is in reference to. It's too late for it to make any difference, even if they do something, which often they don't. Just an email six months later saying they reviewed and didn't do anything. Great, thanks!" **– Meta Trusted Partner**

Almost every partner interviewed experienced receiving no response or resolution beyond an automated email confirming the report had been submitted. The only partner who said they had never failed to receive a response from Meta to a Trusted Partner report was a Ukrainian partner who was only added to the program after the Russian invasion in 2022.

"When you hear nothing you think, are they still reviewing? Did it get lost? Are they just ignoring us? You doubt yourself, but it is them who are at fault."
– Meta Trusted Partner

"If it is a case involving a human rights defender or an activist we usually get a response in under two weeks. That is what we consider fast. If it is something like gender-based violence it is usually much longer than that to get a response. Often, we never get anything back at all."
– Meta Trusted Partner

This is consistent with Internews' own experience. To provide a concrete example, on the 8th of February 2023 Internews received a response to a report submitted on 21st of June 2022. There was nothing in either the report or the response provided by Meta which would indicate why 232 days would be required to provide a response. Typically, Meta do not address or even acknowledge the slow response in these instances.

From: [redacted]facebook.com>
Sent: 08 February 2023 16:47
To: [redacted]@internews.eu>
Cc: [redacted]@internews.org>; [redacted]@internews.org>; [redacted]@meta.com; [redacted]@meta.com
Subject: Lebanon: A Christian group incites against LGBT community

Update on your investigation request

Hi,

Thank you for bringing this content to our attention.

Our team has done an in depth investigation of the content in question, but has found that it does not violate our Community Standards:
www.facebook.com/communitystandards

If there is more information you can provide or if you would like us to review specific elements of the content, please reply to this message.

Kind regards,
Meta Team

From: [redacted]@internews.eu
Date: Tuesday, June 21, 2022, 5:18 AM
To: [redacted]facebook.com
Subject: Lebanon: A Christian group incites against LGBT community

Dear TPC,

A screenshot of a response received by Meta 232 days after the report was submitted to Meta's Trusted Partner Channel. The subject of the report was harmful misinformation, hate speech, and incitement to violence directed towards the GBT population. Meta found the posts to be non-violating and the content remains on Facebook today.

Given the nature of the issues raised through the Trusted Partner channel, this lack of responsiveness has the very real potential to cost lives. Every single partner interviewed for this review who had experience submitting reports stated that they had previously used the Trusted Partner channel to report issues they believed would lead to imminent harm. Two partners interviewed for this report said that regular failure to receive a response, or unreasonable response times, had led to them to cease submission all together.

"Initially we were using the channel as often as possible because we thought they would take it seriously. We were using it very frequently because we were hopeful they would take action. But after a while we are almost not using it at all. We are frustrated as much as you can imagine. It is just draining our energy… It takes time, it takes energy, it takes your emotions, you have to go through so much hateful content, so it is not easy for us. If they were taking action, immediately we would have been taking that as an incentive, because we are doing this with an objective of contributing to our country and our people."
– Meta Trusted Partner

Other interviewees continued to use the channel, but stated that the lack of responsiveness made them less likely to report and eroded trust in the program. In all cases Meta's failure to respond to reports led to significant additional stress and frustration for partners, many of whom are already dealing with crisis and risks to their personal safety.

Ukraine as outlier

In a context of increased attention to Ukraine after the Russian invasion in 2022, Ukraine appears as an exception when it comes to Meta's responsiveness.

Ukraine partners shared that since the Russian invasion in 2022 they typically receive a response to reports submitted through the Trusted Partner channel within 24 to 72 hours – although this was not the case before the invasion.

One partner shared that for most content reported for removal, Meta usually responds within 24 hours, whilst for more complicated cases that involve restoring accounts, or official pages that have been taken over by occupying forces, the response time is usually a little longer – 48 to 72 hours.

The faster response times in Ukraine are obviously warranted and should be celebrated. However, it must also be noted that they stand in stark contrast with other parts of the world, including other conflict zones.

In Ethiopia for example, the Tigray War has resulted in the death of around 600,000 civilians in 2021 and 2022, making it the deadliest war of the 21st century. While the conflicts cannot be meaningfully compared, for a sense of scale, civilian deaths in the Ukraine war are estimated at around 8000 by February 2023. et in Ethiopia, Meta's Trusted Partners say that it regularly takes them weeks or months to receive any response to the reports they share with the company. The reports that are being submitted in Ethiopia are similar in nature to those submitted in Ukraine – threats, disinformation, incitement, incorrect enforcement, compromised accounts – and are being reported through the same channel. Why is one country seemingly guaranteed a quick response, while equivalent reports elsewhere can take months to be dealt with?

The successful experience of the Trusted Partners program in Ukraine shows that improvements regarding responsiveness and response times are possible when appropriate resources are allocated by Meta to the program. When Meta prioritises and allocates resources appropriately it can provide a consistent response within a 24 to 72 hour timeframe, even for the most complicated cases. If they can meet these targets in Ukraine they can meet them anywhere, should they choose to do so and invest accordingly.

In Ukraine, Meta's Trusted Partner program is providing a potentially lifesaving service. Users in Ethiopia deserve the same standard. As do users in emen, yria, Palestine, Myanmar or wherever else they may be.

"I would say that they should invest resources based on the needs! Not only based on the political agenda and their economic interests… With the global south countries they don't care so they are not investing. I know they are investing more than before, because we are hearing that in every single meeting with Meta – 'we are investing more and more!' But the kind of investment they are doing is not anywhere near enough in the kinds of contexts we are operating in."

– Meta Trusted Partner

Lack of transparency around responsiveness

In response to a draft version of this report Meta's statement included the following comments (Meta's full response can be found in Annex 2):

> We acknowledge the variety of Partner experiences documented in the Report, and we are committed to continue improving training resources and ingestion systems to address these outliers and strengthen the program. However, the reporting issues of the small sample of Trusted Partners who contributed to the Report do not, in our view, represent a full or accurate picture of the program.

Meta have, thus far, declined requests to share any data that supports that assertion.

Meta currently provides no transparency around response rates or response times to Trusted Partners, nor does it share any targets for response times that may help manage expectations. In questions submitted to Meta for this report (see ANNEX 1) Internews requested concrete data on response rates and times for the Trusted Partner channel, however Meta declined to provide this information. Questions Meta did not answer included the following:

1. What percentage of Trusted Partner reports received any response from Meta (not including an automated response to say that the report has been received).

a. In the most recent month?
b. In the last 12 months?

2. For those reports that do receive a response, what is the average time between the moment that the report is submitted by the Trusted Partner to the time at which they receive a response from Meta (not including an automated response to say that the report has been received)?

a. In the most recent month?
b. In the last 12 months?

3. Does Meta have internal targets for response rates and times to Trusted Partner reports?

a. If so, can these targets be shared with partners?
b.If targets cannot be shared, why not?

Given that the Trusted Partner program utilizes a ticketing system it is assumed that Meta has this information available internally. Response rates and times with participating partners are aggregated data, so there is no obvious breach of data privacy or security in sharing these numbers, or reason for Meta not to do so.

In response to a draft version of this report Meta shared the following statement in relation to this section on transparency:

> We recognize the value of increased transparency, both with respect to clarity on shared goals and performance, and to recognize the significant impact of Trusted Partner reporting. We note your Report also recognizes the need for strong operational security protocols; this accounts for some of our program design and structures. While data protection laws prevent us from sharing information about actions taken with regard to other users, we strive to provide high-level feedback to our partners through group consultations and one-on-one debriefs.
>
> We are indeed working to develop new methods of sharing information about the overall impact and performance of the Trusted Partner program, consistent with security, confidentiality preferences, and data protection of the many hundreds of organizations who participate.

Meta's reference to data protection laws in this response is either confusion or deliberate misdirection. To reiterate, there is no legal, privacy or safety reason why Meta cannot share aggregated and averaged figures about the company's own response times or response rates, which is the only information currently being requested. Meta is under no obligation to share this information with partners, however doing so would help rebuild trust and set clear benchmarks around performance.

While Meta chose not to provide any concrete data addressing these questions on response rates, the company did provide a statement (see ANNEX 1) which blames delayed response times largely on staffing issues during the Covid pandemic and indicates they are now 'steadily increasing' their responsiveness:

"We recognize that the Covid pandemic severely impacted our operations and resulted in poor reporting experiences for our partners from 2019 - 2021. During this period our content review teams operated at limited capacity and were unable to respond as quickly to trusted partner channel reports as we would like and as they have done in the past. Under these difficult circumstances, we prioritized the most harmful content for our teams to review, such as risk of imminent physical harm or violence.

In 2022, we were able to improve our overall operational resources for content review teams and are steadily increasing our ability to respond to Trusted Partner reports in a timely manner.

We generally expect reports to be reviewed and actioned within 1 to 5 days, though especially complex cases may take longer. All Trusted Partner reports receive an automated response acknowledging receipt, though we acknowledge there have been issues with this in the past."

Internews asked participating partners if they had noticed an improvement in response rates and response times since 2021, and whilst almost all said that there had been some improvement, all noted that the problems persisted in 2022 and 2023.

"About a year ago till January 2022, the response and resolution time was overwhelmingly delayed, sometimes by 3 to 6 months even for reports that required a time sensitive approach. Since then, the response rate has gotten better, but just for about half the reports that our organization makes."
– Meta Trusted Partner

"The average has definitely come down since the beginning of 2022. The majority of reports we submit now get a response within five days. But there are still something like 20% of reports that take weeks, or we don't get a response. We report a lot, so it is still a lot of reports with an unexplained delay."
– Meta Trusted Partner

Internews' own experience demonstrates that poor response rates and response times of multiple months continue to be an issue in 2023, which can no longer be blamed on Covid. Notably, Internews has continued to receive response times stretching to weeks and months despite raising this issue directly with managers of the Trusted Partner program and requesting to work with the company on a review of the service.

While anecdotally there have been improvements to average response times since 2021, Meta's refusal to share concrete data makes any potential improvement impossible to quantify or verify. While averages may have improved, the persistence of non-responses or months-long response times with no explanation indicates that something remains broken in Meta's case management system.

Communication and partner engagement

Closely following responsiveness, the issue that most consistently frustrated Trusted Partners interviewed for this review was communication and engagement with Meta. Overall most partners feel that Meta's communication is often perfunctory and dismissive.

"They treat it like a privilege to have this communication with them. They explicitly said that it is a 'privilege' that we have this connection with them. They said that to us. It is because they view it as a privilege they feel they don't have to respond."
– Meta Trusted Partner

Generally when Meta has made a decision on a Trusted Partner report the partner will receive a response notifying them if Meta has acted or not. In most cases these responses are pro forma and contain no further information relating to the issue in question. Even when the company has acted on a report, this lack of detail or acknowledgement can be frustrating. When Meta provides a response saying that it has reviewed the report and decided not to act, the pro forma response often comes across as dismissive or even insulting.

"You get a response that just says 'this does not violate our community standards' and no other information. Sometimes they link to their policies – but we know the policies, that's why we reported the post! Because we think it does violate!"
– Meta Trusted Partner

"Sometimes the response is not at all relevant to the case, perhaps reflecting that they didn't actually understand the report."
– Meta Trusted Partner

Cases that take an extended time to receive any response and then receive impersonal or pro forma replies are particularly galling.

"Waiting for someone to respond a month or so, and then they ask for more details on the report that we sent a month earlier. Which means we have to go through every link that we sent a month ago. We have to face every kind of emotional feeling again. Meanwhile people have been killed or harmed in multiple ways while the links are there and being spread."
– Meta Trusted Partner

As with poor responsiveness, the way that Meta communicates with its partners is highly demotivating, and this is especially felt by partners who are operating under the most difficult circumstances or dealing with the most urgent crises.

"Our motivation is entirely dependent on the response we get from Facebook."
– Meta Trusted Partner

Bypassing the official Trusted Partner Channel

Outside of Ukraine, Trusted Partners who reported the best experiences in relation to responsiveness, clear communication, and action on reports, were those who routinely supplemented or entirely bypassed the official reporting channel, communicating directly with known, personal contacts at Meta to escalate the issues raised in the report. It is important to note that these personal relationships are not directly tied to the Trusted Partner program, and many Trusted Partner participants do not have access to this kind of side-channel. Ad-hoc side-reporting significantly advantages partners who have good relationships within Meta, or have responsive local policy staff, and further alienates partners who do not have these relationships.

"If it is something urgent I just WhatsApp REDACTED. If I submit a [formal] report I copy in our policy person to follow up. But if it is urgent I don't even bother to send a report because it will take too long, I just message REDACTED."
– Meta Trusted Partner

Another partner stated that while they had previously experienced extremely poor responsiveness and action-rates from Meta's Trusted Partner channel, their experience had changed dramatically when the Meta public policy focal point for their country changed:

"The previous policy person just ignored me and never responded, but we're very happy with this new person. Even if we message them late at night they write back to us."
– Meta Trusted Partner

In this case, the previous Meta focal point had been from another country and lacked contextual understanding of the issues that the partner was reporting. When the partner notified this person via WhatsApp that they had submitted a Trusted Partner report they often received no acknowledgement or follow up. In 2022 when the new Meta country focal point took over they were immediately more responsive, understanding the issues that were reported and offering to chase them up internally at Meta and requesting further information via WhatsApp when needed. In this way the partner was able to supplement the official reporting channel. At this point both the action rates and response times to the partner's reports improved dramatically.

"When I read about the layoffs at Meta my first thought was that if my contact there was fired we would no longer get any response to our reports. I thought I should write to Mark Zuckerberg and say 'please do not fire this man!'"
– Meta Trusted Partner

For some partners these informal channels have supplanted the official Trusted Partner channel.

"We use Signal primarily. It isn't perfect but it means we can exchange information in real time, and see when they have read and responded. Beyond the response times it is about communication and having a personal interaction. We know the people we are reporting to, and there is an element of trust there, built up by working together over time."
– Meta Trusted Partner

This partner, who represents a high priority and high-profile country, is largely satisfied with Meta's responsiveness, regularly receiving same-day resolutions to issues raised with personal contacts at Meta via Signal. It is unclear if these 'reports' submitted outside of the official Trusted Partner channel receive case numbers in Meta's Trusted Partner system, how they are documented, or if they count towards Trusted Partner statistics (discussed below). Internews' own experience is largely in line with other partners' interviewed for this review, and serves to highlight that the particular individuals within Meta who are alerted to a report have a significant impact on whether or not that report will receive a timely resolution.

From: [redacted]@INTERNEWS.ORG>
Date: Friday, June 10, 2022 at 8:13 AM
To: [redacted]@fb.com>, [redacted]@fb.com>, [redacted]@fb.com>
Cc: [redacted]@fb.com>, [redacted]@fb.com>
Subject: RE: Hacked Liberian Newspaper Page

The [redacted] still has not been able to regain access to their Page. They finally received an email communication from the Community Operations team on June 7th and immediately responded with all of the requested information. It has now been three more days since that information was provided and no further follow up communication has been received. It has been 28 days since the initial hack, and 22 days since I first reported the incident through the Trusted Partner channel.

For clarity, I would like to give a timeline of the process so far:

May 12th – [redacted] falls victim to a phishing attack and gives admin rights to a hacker.
May 13th to 15th – [redacted] team reports the problem through in-platform tools but is unable to figure out any way to report a phishing attack.
May 16th – The [redacted] team lodge a complaint using the copyright process and receive an automated response telling them this is not the appropriate path for this issue, but with no further advice.
May 17th – The [redacted] request assistance from Internews.
May 18th – Internews reports the issues through Meta's Trusted Partner channel. No answer is received.
May 26th – Internews sends a follow up to the previous report questioning why no action has been taken.
May 30th – Internews sends a second follow to the report, this time copying in [redacted].
May 31st – [redacted] responds saying will circle back.
June 1st – [redacted] responds saying that the team has yesterday contacted [redacted] requesting a new email but has not received anything. The [redacted] team has not received any such communication. It is unclear via what channel this request was made.
June 1st – The hacker loses admin rights to the [redacted] Facebook page.
June 3rd – The hacker somehow *regains* admin rights to the [redacted] page.
June 6th – [redacted] team sets up new email address which is provided to Meta by Internews. [redacted] has still not received any direct communication from Facebook via any channel at this point.
June 7th – [redacted] team receives an email from Facebook Community Operations requestion information on the incident as well as photo ID. [redacted] team responds with all requested information.
June 10th – No further communication has been received to date. [redacted] team still has no access to their page.

Throughout these 28 days the [redacted] has been unable to post any updates clarifying the situation to its followers, and pornography has continued to be visible on the page.

Screenshot of 2022 email chain with Meta staff escalating an urgent Trusted Partner report which did not receive a response through the official channel after 12 days. The issue was eventually resolved 40 days after the initial Trusted Partner report was submitted.

Such reliance on informal channels is presumably not hIn some cases where long wait times were experienced, Internews' staff escalated the report directly with personal contacts at Meta, eventually receiving a resolution. At other times Meta staff advised Internews staff that the official Trusted Partner email was the only appropriate channel for escalation, resulting in some reports that never received any response.ow the Trusted Partner channel is supposed to work. Even when such avenues are available, the need to bypass the official reporting channel to receive an adequate response highlights a fundamental failure of the Trusted Partner system.

Case management and decision making

Trusted Partners interviewed for this report expressed a range of questions around how Meta receives, assess, and prioritizes Trusted Partner reports, including questioning if AI or automation were used as part of the process. Partners also wish to understand the decision-making processes in these cases, including how it may differ for different categories of report, and who is ultimately accountable for the decisions made.

Internews submitted the following questions to Meta about this process in June 2022:

1. *Can Meta explain how reports are processed, prioritized, and directed to internal teams once they have been submitted?*

2. *At what stages is automation used in assessing Trusted Partner reports?*

3. *When reports are submitted or involve content in languages other than English how are these dealt with?*

4. *What role does the report's geographic location have on prioritizations (e.g. are some countries prioritized over others)?*

Meta provided the following answer in February 2023:

> Automation is not used in assessing Trusted Partner reports. Reports submitted to the Trusted Partner Channel (TPC) are received directly by our 24/7 Global Escalations team. The escalations team will first assess the content for priority level, to ensure that anything that might result in imminent harm goes to the absolute top of the queue. We take into account a range of factors including known violence or crisis in the country or region, whether there is an ongoing or near-term election, and the policy area implicated.
>
> Depending on the language, the type of violation of the reported content, and other particulars of the report, our escalations team may loop in other teams in order to help assess the content. For example, if the content is not in English, we will loop in a native speaker. If the content is harmful misinformation, we will have an expert on our misinformation policies help assess. If we need further information if the content is an edge case – not clearly a violation of our policies – we may loop in our Content Policy team, who writes the policies.
>
> After cross-functional teams align on the appropriate next steps, Meta will resolve the escalation, which could mean removing or restoring content or accounts, or taking additional action such as disabling a hashtag, or alerting other internal teams to a concerning trend. Meta will then inform partners of the action taken.

Many Trusted Partners expressed frustration at the opaqueness of this process, questioning who at Meta is looped into the report submission and exactly at what stage. Overwhelmingly, partners want to ensure that they are including the most relevant information in their reports and ensuring that they are being directed to the right people at Meta to make a speedy and informed decision. A number of partners have thought deeply about this processes.

"We are not only raising content for review. More and more we are escalating accounts and trends or campaigns that we see picking up... We are trying to understand Meta's needs when it comes to different categories of report. We need to collectively think through the types of issues that are being escalated and work through the routing."
– Meta Trusted Partner

Several partners highlighted that using the existing system they did not know how to best frame the importance or urgency of a specific issue, to ensure that it was prioritized accurately and routed to the relevant technical or policy channels.

"If there was a template we could use to rate the urgency and select the category of issue that would make it simpler."
– Meta Trusted Partner

Partners also expressed concerns about who at Meta can access the reports and see who was flagging issues to the Trusted Partner program. In many countries state actors or state-aligned actors are responsible for causing the issues that Trusted Partners are reporting. In some instances, Meta's public policy staff may be responsible for liaising with these same governments or entities. There is a concern that if these staff are involved in decision making it may create a conflict of interest, or potentially a security risk for Trusted Partners operating in those countries.

When it comes to decisions Meta takes based on Trusted Partner input, partners feel there is very little transparency and a great deal of confusion. Decisions are supposed to be based on Meta's policies, but in practice it is unclear who within Meta is the final arbiter. This is especially confusing in cases where Meta may apply their 'escalation only' policies such as that around misinformation and harm, which theoretically rely on experts such as Trusted Partners to evaluate both the veracity and harmfulness of content.

"They did tell us about this policy [misinformation and harm] but I have never known them to use it."
– Meta Trusted Partner

"In the case of REDACTED I reported the posts that falsely said he was part of ISIS – in one post they even photoshopped his picture. I said in my report this is fake and it is dangerous, but they said it didn't violate the standards. Then when REDACTED was killed I wrote back and I shared the link to the news article about his murder. Only after that they took down the posts I had reported after he was killed."
– Meta Trusted Partner

Cases where activists, human rights defenders or journalists are falsely accused of crimes or association with terrorist groups were brought up by partners in multiple countries as examples where Meta has failed to act despite Trusted Partners assessing the content as false, maliciously intended, and likely to result in physical harm to target. Meta's policies in these cases are unclear, and the decision-making process is equally opaque. In some cases, Trusted Partners received a response saying that no action would be taken on their report, then sent the report to a personal contact at Meta who was able to have the initial decision reversed.

The logic of the Trusted Partner program is that Trusted Partners bring contextual and linguistic expertise, and are able to help Meta understand the nuances of harmful content in different countries. Yet many Trusted Partners interviewed for this report felt that this very expertise was being ignored, sometimes putting lives at risk.

"There are posts that will result in harm if they are shared. Like simply posting a picture of a person with their car and number plate and the name of the person, and a caption saying something like 'you know what you have to do.' Which means go kill them, within the dynamics of the context…. Or they use language like 'here is this person, go kiss him.' And 'kiss him' means kill him in this context, this is very clear. We know what those terminologies mean because we understand the context. But Meta were not interested in that, even when we tried to explain the meaning to them."
– Meta Trusted Partner

"Most of the cases are clear cut, there is nothing very borderline. But sometimes there are cases that require context knowledge or linguistic understanding. The use of language may not be direct [i.e. not a defined slur or explicit threat], but the meaning is clear to those who understand the language and context."
– Meta Trusted Partner

Content that may not violate the letter of Meta's policies can still cause real harm – and the Trusted Partner program is supposed to help prevent that. A coded threat may have a low potential for harm in some contexts, but if it comes from a known supporter of a violent armed group and is made against an identifiable and credible target then that threat is very real. Trusted Partners are Meta's acknowledged 'subject matter experts' in this context. In some of these cases Meta appears to act based on Trusted Partner advice, whilst in very similar cases a pro forma response is sent saying that the content does not violate Meta's policies. In either case, Trusted partners have little insight into who is making the decision and on what basis it is being made.

Meta's response to the draft version of this report included the following comments:

> Partner reporting behavior, including compliance with reporting protocols and partners' familiarity with our Community Standards, also significantly affects response times. We provide training to facilitate clear reporting and are grateful for the time and efforts groups have dedicated to this work. We have created extensive online educational materials, supplemented by direct engagement, to strengthen Partner reporting and improve the speed and efficiency with which we ingest content reports.

These are two important points. The quality and detail of Trusted Partner reports clearly has an impact on Meta's ability to respond in a timely fashion. Poor quality or confusing reports will inevitably slow down the process for everyone. Meta has provided training materials to assist with this process, and these are an important resource.

However, these training materials do not address many of the concerns raised above, and even Trusted Partners highly engaged in the program continue to be confused by Meta's processes and decision making in many cases.

If Trusted Partners had a clear understanding of how different categories of report were triaged and managed throughout the decision-making process this would improve the quality of report, as well as build trust and confidence in the system. When this transparency is absent many partners assume that good processes are not in place or are not being followed.

Volume

In order to evaluate the performance of Meta's Trusted Partner program it is essential to understand the scope and volume of the program.

In answers provided by Meta to questions posed for this report (see ANNEX 1), Meta confirmed that as of February 2023:

> There are **465 organizations** enrolled in Meta's Trusted Partners program, covering 122 countries. Meta has at least one Trusted Partner in each of the 122 countries. In addition to local partners, we work with regional organizations that cover multiple countries. These groups are counted once in the network of partners.

This leaves 73 countries without any country level representation in the Trusted Partners program. Internews also asked specific questions about the geographic breakdown of Meta's Trusted Partner program, requesting information on how many partners were included from each major geographic region (see ANNEX 1). Meta did not provide any information in response to these questions.

Internews posed the following questions to Meta regarding the number of reports that Trusted Partners submit through the channel:

How many Trusted Partner reports does Meta receive?

a. How many Trusted Partner reports were submitted in the most recent month?

b. How many Trusted Partner reports were submitted in the last 12 months?

In response to these questions Meta provided the following answer:

> The Trusted Partner Channel receives about 1000 escalations per month.

Meta has only provided a rough figure, but it provides some scale and allows us to make some equally rough calculations. 1000 Trusted Partner reports a month works out at **under 33 reports submitted to the channel a day**. With 465 partner organizations this means that **on average each partner is submitting around one report every two weeks**.

These figures are broadly in line with the estimates and records provided by Trusted Partners interviewed for this report. Notably the number of reports submitted by individual partners varied greatly, with some treating Trusted Partner reporting as a key element of their workflow, and others barely using the service at all. Several partners shared that they submit reports to Meta's Trusted Partner channel daily, while other partners submit reports only once or twice per year, and several partners said that they no longer use the reporting channel at all, either out of frustration or due to getting a better response via informal channels. It is not clear if cases escalated via Signal or other informal channels are documented or count towards the 1000 cases a month figure provided by Meta.

Based on the sample size involved in this review (around 5% of Meta's global Trusted Partners) **a significant proportion of the monthly escalations are submitted by a relatively small number of Trusted Partners.**

Staffing levels

How many staff are required to adequately assess and respond to 33 Trusted Partner reports a day, and meet the needs of 465 global partners? Internews understands that the Trusted Partner program was impacted by Meta layoffs in 2022 and is likely to have further cuts in 2023. **Given the issues with the program outlined above, cost-saving cuts to the program would further jeopardize the program and worsen the Trusted Partners experience.** We asked Meta the following questions in relation to staffing levels:

How many staff does Meta have working full time on the Trusted Partner program?

a. How was this resourcing level calculated?

Meta provided this response:

> The Trusted Partner Program is jointly managed by Meta's Content Policy and Global Operations teams, working in close collaboration with regional Public Policy teams who are responsible for overall relationship management with local partners. The Content Policy team leads in the definition of program strategy, develops training materials, and coordinates outreach with Trusted Partner organizations. The Operations team receives and actions Trusted Partner reports. While we can't share specific numbers, there are more than 50 people across Content Policy and Operations who work on the Trusted Partner program, and many more regional policy leads who hold relationships with NGOs in their region.

The figure of 50 people provided by Meta is confusing – and arguably deliberately obfuscating - as Meta has not given any indication of the percentage of their time each of those people can devote to the Trusted Partner program, and how much is given to other functions.

Currently Meta tells us they resolve an individual escalation "after cross-functional teams align on the appropriate next steps." From an external perspective it seems that **more dedicated staff and a clearer and more transparent decision-making process may address some of the issues of timeliness.**

Elsewhere Meta tells us that: "reports submitted to the Trusted Partner Channel (TPC) are received directly by our 24/7 Global Escalations team", but we do not know how many full-time employees (if any?) make up this team that receives and initially evaluates the reports. We also do not know what other escalations the "24/7 Global Escalations team" may be dealing with, beyond reports received through the Trusted Partner channel.

Meta did not directly address staffing levels in their response to the draft version of this report, which came shortly before another round of layoffs at the company.

Without more information it is impossible to evaluate how many staff are required to adequately respond to Trusted Partner needs. However, we can again look to Ukraine as a positive example. If Meta is able to respond to Ukraine Trusted Partner reports within a 24-to-72-hour target, as they appear to be doing, then with adequate resourcing they could do the same elsewhere.

Onboarding and email whitelisting

While Trusted Partners are selected as organizations, functionally each organization must designate a limited number of individuals to use the Trusted Partner channel. These individuals are expected to take part in a remote onboarding process. For those wishing to use the in-app reporting mechanism they must use their own personal platform accounts – which many refuse to do or do so with extreme reluctance, citing privacy and safety concerns. Organizations must also nominate specific email addresses to be added to the Trusted Partner whitelist, with only whitelisted emails received by the channel.

This process has caused significant problems for many partners, in this case especially impacting larger organizations. The limitation on whitelisted email addresses has led to an increased workload for those individual staff who have access to the channel, and forced organizations to create their own workflows and mechanisms to meet with the restrictions imposed by Meta's design. This also leads to problems when staff move on or change roles, with changes then required to Meta's whitelist.

"The staff member who was on the email whitelist left before I even joined REDACTED, so we actually had no one on the system who could report. I didn't know how it worked or if my email was on the list. I had to do the onboarding, but then by the time I knew about it, it was right before the election and of course I didn't have time to do it properly."
– Meta Trusted Partner

Internews submits reports relating to many different countries and regions, but has been told by Meta that a maximum of six whitelisted email addresses is allowed per organization. No reason for this restriction was provided. Internews has attempted to address this issue using shared mailboxes allowing multiple people to submit reports, however this system has presented its own drawbacks, especially given the sensitivity of the issues potentially reported. Other multi-regional partners interviewed for this review have stated that the requirement that a limited and specific individual submit reports has created extra workload for that person beyond their scope-of-work, and meant that those who have the most direct knowledge and contextual expertise may be prevented from directly accessing the program.

One of the simplest issues raised during this review relates to the current lack of an automated 'bounce-back' notification to Trusted Partner reports submitted by non-whitelisted email addresses. If you submit a report to the Trusted Partner email address from a non-whitelisted email address Meta tells us the email is never received by anyone at Meta and the report is lost. As it stands, the sender in this case does not receive any email notification informing them that their email has not been received.

A failure to send automated email bounce-backs has almost certainly resulted in Trusted Reports being lost, and could potentially be a contributing factor to the program's poor response rates. Trusted Partners may send a report from a non-whitelisted email address in error, either due to using a different email account than usual, or because of changes made to Meta's whitelist.

While this issue may be a factor when it comes to poor response rates it should not be overstated. Each report submitted by a whitelisted email address does receive an automated reply acknowledging receipt, and partners interviewed for this review say that it is largely these reports that are not being responded to. To further complicate things, Meta has previously had acknowledged issues failing to send automated receipt emails, although this issue appears to be largely resolved.

Consultation on policies, products, and practices

While the escalation channel was by the far most common interaction that most Trusted Partners had with the program, Meta's own description of the program's intentions focus heavily on consultation on policies, as well as enforcement. Meta publicly recognizes that Trusted Partners have 'subject matter and regional expertise', which 'help strengthen our policies by enabling us to consider a range of perspectives to inform our content moderation efforts.' Such consultation should be applauded, and has no doubt improved Meta's content policies and enforcement guidelines to a great degree. Despite this, in practice many partners feel that these consultations are severely lacking in intent, process and follow-through.

"They invite us to meetings, or 'consult' us on policies, but then nothing changes. They tell us how useful it was speaking with us and make us feel important, and then we never hear back from them. Eventually I realized, they're gaslighting us! It's just gaslighting. It is all to make us feel special and keep us quiet."
– Meta Trusted Partner

"We told them that the most important time was after the election, we spoke with them at length about it. But they ignored that and disappeared as soon as it was done. And of course there was a huge amount of misinformation and disinformation in the weeks after the election and that caused real problems. So they totally ignored us, but then the burden was still on us to report. We still had to do the work. I mean?!"
– Meta Trusted Partner

Many of the Trusted Partners interviewed for this review shared that they had participated in a range of consultations with Meta, both as a direct result of their participation in the Trusted Partner program and prior to their joining. Overall partners report that Meta is comparatively more proactive in its consultation with partners than other platform companies, such as Google or Twitter. While this engagement is desperately needed, without follow-up or meaningful accountability these consultations often seem performative.

In the feedback provided by Meta to the draft version of this report Meta raised several specific instances of consultations with Trusted Partners as positive examples of this interaction:

> The Trusted Partner program rests on deep consultation with Partners. For example, when Trusted Partners in certain countries told us that reporting content by email posed safety risks, we created a secure in-app reporting mechanism for them to use. The grants program was likewise designed in consultation with civil society organizations, to respect organizations' preference for independence, while at the same time providing essential resources to groups operating in resource constrained environments. There are many other such examples.
>
> We also support networking and exchange between Trusted Partners while respecting partners' requests to remain anonymous. In October 2022, for example, we organized a Middle East Community Summit that included Trusted Partners from across the region, and we have hosted similar events for Sub-Saharan Africa and for the Asia Pacific region.
>
> Finally, we engage frequently and productively with Trusted Partners on content policy issues. For example, the policy under which Meta removes misinformation "where it is likely to directly contribute to the risk of imminent physical harm" was designed and is enforced with the input of many Trusted Partners, and Meta deeply appreciates this engagement. Our commitment to listen to Partners does not mean, of course, that we will revise our policies in response to each individual piece of feedback we receive. Similarly, we do not remove all content reported to us as violating by Partners; our policies dictate what is and isn't removed.

These specific examples are notable as in each case Internews or Trusted Partners who were interviewed for this report – including those quoted above in this section on consultation - participated in the consultations and meetings that Meta has flagged. **While Meta regards these examples as positive models of consultation, the partners who have been involved have deep reservations, and differing acounts of the experience than that provided by Meta above.**

One of the key country level Trusted Partners who raised safety issues with email reporting to Meta and suggested the in-app reporting mechanism Meta describes in their feedback provided this perspective on the consultation process with the company:

"The in-app reporting mechanism was meant to address security concerns we had, back in 2018. The problem is Meta refused to do a proper user requirement exercise and only consulted us once it had a working prototype. We had provided them with a consolidated overview of our needs - but they didn't engage us over them.

5 years on, we end up in a situation where despite having the tool built for us, we barely use it, and continue to rely on third party messaging apps. It's a shame - because this should have been an opportunity to co-design something which could have saved us all a lot of time. I recently came across the list of those early requirements. They still hold and we would love to engage Meta over them again - and reflect on our collective learnings when it comes to how to improve co-design."

– Meta Trusted Partner

Internews engaged heavily with Meta regarding the design of its policy around 'misinformation "where it is likely to directly contribute to the risk of imminent physical harm"', which Meta also raised in its feedback as a positive model. Throughout the consultation process Internews made it clear that the proposed policy lacked clarity and transparency, and would lead to significant confusion when it came to practical enforcement. Other Trusted Partners raised similar issues. These concerns were not reflected in the final policy.

It is worth diving into this example a little deeper, as this policy is critical to keeping Meta's platform users safe, and Meta relies heavily on Trusted Partners for it's operationalization:

> We remove misinformation or unverifiable rumours that **expert partners** have determined are likely to directly contribute to a risk of imminent violence or physical harm to people. We define misinformation as content with a claim that is determined to be false by an authoritative third party. We define an unverifiable rumour as a claim whose source expert partners confirm is extremely hard or impossible to trace, for which authoritative sources are absent, where there is not enough specificity for the claim to be debunked, or where the claim is too incredulous or too irrational to be believed.
>
> We know that sometimes misinformation that might appear benign could, in a specific context, contribute to a risk of offline harm, including threats of violence that could contribute to a heightened risk of death, serious injury or other physical harm. We work with **a global network of non-governmental organisations (NGOs), not-for-profit organisations, humanitarian organisations and international organisations that have expertise in these local dynamics.**

Many of the greatest frustrations expressed by Trusted Partners interviewed for this report related directly to confusion around this policy. Meta says that they remove misinformation that expert partners – i.e. its Trusted Partners – have determined are likely to directly contribute to a risk of imminent violence or physical harm. But in many cases they do not remove this content, even when Trusted Partners flag it to them and explain the risks. A particular trend, observed in multiple countries, and by multiple partners, is when individuals such as journalists or activists are falsely accused of crimes or of membership of opposition or terrorist groups. In some cases these figures have been killed without Meta taking action on Trusted Partner reports.

There are many questions about the operationalizing of this policy around misinformation and physical harm – all of which were raised with Meta during the initial consultation on the policy itself, and none of which have been addressed to the Trusted Partners who are relied on to implement the policy. What is the burden of proof expected of the expert partner when they say that something is 'likely to directly contribute to a risk of imminent violence or physical harm'? Meta has set no bar other than the assessment of the expert partner – but Meta also reserves the right not to listen to a partner when they advise that harm is likely. Who at Meta makes that decision, and on what criteria is it made? What is a Trusted Partner supposed to do when they disagree with Meta's decision? Who are Meta accountable to when they get it wrong? These are grave questions which Meta has still not addressed, four years after the consultation process was conducted and the policy decided.

Overall Meta's feedback to this section further emphasises that clear, mutually agreed guidelines and expectations around consultations need to be set. Without such guidelines Meta can say they have relied on consultation, even as those they have consulted feel that their concerns have been ignored and their input is being publicly misrepresented.

"There needs to be notes taken of any consultation and then both parties should have a chance to review and make edits before it is finalized. And then we need to have a copy of that for our records. Otherwise we speak to them and they tick a box saying they have done a consultation, tick!, and they just do what they were going to do anyway! When they tell people that they consulted us we need to be able to say, 'but that's not what we told them'."
– Meta Trusted Partner

Burden on Trusted Partners

Many Trusted Partners who participated in this review framed their membership in the program explicitly as a service they were providing, either to Meta or to their community, or to both. The partners spend their limited time, energy, funding, and other organizational resources to participate in Meta's Trusted Partner program, with no compensation. Trusted Partners see this investment on their part as providing a clear benefit to Meta – by taking the time to report harmful content and participate in consultations, the Trusted Partners are improving Meta's products and helping keep Meta's platform users safe.

Publicly Meta also frames the Trusted Partner program as a program that provides clear benefit to the company:

> "We are grateful for the partnerships that we have with expert civil society organisations that help us to better understand local context, trends in speech and signals of imminent harm."

The benefit that the Trusted Partners receive from their participation in the Trusted Partner program is more ambiguous. Meta acting on Trusted Partner input to protect users of Meta's products is seen by some partners as a benefit to the partner, whilst for others it is simply action that the company should be taking in any case.

"It is their [i.e. Meta's] responsibility to enforce their rules. Should I be grateful when they take down content I report? No. They should be grateful that we helped them when they couldn't find it themselves. We are helping them to do their jobs, they are not helping us. It is not our job to clean up their mess."
– Meta Trusted Partner

Providing this service to Meta has resulted in an increased workload and sometimes a heavy emotional burden, as described by many Trusted Partners. The types of content partners describe includes horrific violence, sexual violence, threats, and toxic harassment - often targeting members of their own community. Arguably even harder to deal with, Trusted Partners regularly find themselves in the position of having to intervene to protect someone's safety. This can range from cases where someone has been arrested and their account needs to be shut down, to examples where threats have been made, or someone's personal information has been published along with incitement to violence. This responsibility for others can place extreme stress on Trusted Partners, and is exacerbated when partners cannot be confident of a quick or positive response from Meta.

"It is so discouraging. I hope this review helps them understand that we are trying to help them. We are trying this for free, voluntarily, with the objective of making their platform safe. Because they don't seem to understand."
– Meta Trusted Partner

In some countries and contexts Trusted Partners face threats to their own safety. This includes organizations operating in conflict areas or representing oppressed minority groups. In some cases, monitoring and reporting against state actors or violent groups can open Trusted Partners to risks of persecution or reprisals.

For the most part these risks were already in place before organization's joined the Trusted Partner program, however there is a significant chance that if their role in raising certain issues or having certain content removed were to become known, this may increase their organizational and personal risk profile. For this reason, operational security is a key concern for many Trusted Partners.

Trusted Partners also speak of a wider sense of responsibility to their community as they are often one of the only conduits for people wishing to contact Meta. When other members of civil society, such as activists or journalists, have issues with their accounts, or believe themselves at risk, they often turn to a Trusted Partner as people in a position to help them. In most countries that have a Trusted Partner there are only one or two organizations enrolled in the program, and there are never more than a handful. Some Trusted Partners have embraced this position, actively engaging with other civil society organizations to identify issues and provide a service to their peers. For other partners this position provides more pressure and strain on limited resources.

Meta does not compensate Trusted Partners for their contribution to the Trusted Partner program, although some partners do receive funding from Meta. **Meta's public comments on the program contain a misleading framing of this relationship:**

> "Meta provides trusted partners with funding to support our shared goals of keeping harmful content off our platforms and helping to prevent risk offline."

For those Trusted Partners – including Internews – that do receive funds from Meta, it is either in the form of donations or as contracts for services rendered, however it is understood that any such financial relationship is separate from membership of the Trusted Partner program. No partners who participated in this review receive funding that is directly tied to the Trusted Partner program.

When asked, some Trusted Partners felt that Meta should provide financial compensation to participating organizations, others felt that accepting Meta funding would compromise their integrity and make it harder to criticise the company. In response to a draft version of this report Meta provided the following feedback in relation to financial assistance provided to Trusted Partners:

> The [DRAFT] Report does not accurately reflect the nature or scope of the Trusted Partner grants program. These grants are not intended to provide payment for a service. Rather, the program seeks to support the organizational sustainability of civil society partners operating in challenging environments with limited resources. We consider this an ecosystem level investment to support a more resilient and independent civil society sector.
>
> Since 2021, Meta has provided donations to Trusted Partners based outside North America and Western Europe who demonstrate consistent engagement with the channel. The recipients constitute roughly half of our partners located in Latin America, Asia, Africa and the Middle East.

None of the Trusted Partners interviewed for this report had received any funding through the 'Trusted Partner grants program' at the time of interview. Meta's assertion that 'roughly half' of its Trusted Partners outside of the US and Europe have received funding of this kind is not disputed, however it cannot be independently verified based on the information provided. It is also unclear how many Trusted Partners are eligible (i.e. 'roughly half' of how many?) because, as discussed above, Meta has declined to answer questions about the geographic breakdown of its Trusted Partners, including how many partners are located outside of North America and Western Europe.

Most Trusted Partners are not-for-profit organizations of some sort, which means they are heavily reliant on funders or donations (although some also conduct contract work). In practice this means in many cases government and institutional funders are directly financing the staff time that Trusted Partner organizations contribute to Meta's Trusted Partner program.

The partners who most regularly reported content had projects that were devoted to monitoring and reporting harmful online content. These programs are ultimately paid for by funding bodies including USAID, the US State Department, the UNDP, or the European Commission. If you accept that Meta derives a benefit from the Trusted Partner program – a position Meta publicly holds – then that benefit is effectively financed by US and European taxpayer funds that have been allocated by governments to support international development and humanitarian assistance. Across 465 organizations this represents millions of dollars of public aid funding being used to support Trusted Partners reporting and providing consulting services to Meta.

"Every minute that we spend complaining to Meta is a minute we could be spending doing something else that is more useful."
– Meta Trusted Partner

Perhaps surprisingly, the Trusted Partners who participated in this review were largely prepared to accept these burdens and funding arrangements, provided Meta does what is required to improve the program, keep people safe and understand the diverse needs of its global product users. For the Trusted Partner program this means **understanding Trusted Partner needs, making clear commitments to specific processes and targets, and investing sufficient resources to meet those commitments.** Overwhelmingly, Trusted Partners see this as the bare minimum requirement for a company of Meta's footprint and global impact.

"[REDACTED] from the Human Rights team told us that the Trusted Partner program does not make money for Meta so it is unreasonable for us to expect too much. I'm sorry but fuck that, quit your job in shame. You know what else doesn't make money? Fire escapes! But if you're building a building you have to have them. It is about safety! They are spending how much on their stupid headset world? Don't you dare cry poor. Don't you dare."
– Meta Trusted Partner

"I just don't understand… they're a multi-billion dollar company and they can't have a proper trusted partner system in place?"
– Meta Trusted Partner

Recommendations

The following recommendations combine proposals and requests shared directly by Trusted Partners, and a synthesis of needs based on the review findings.

1 Program overhaul and co-design

Meta should commit to an overhaul of the entire Trusted Partner program, starting from first principles. What is the goal of the Trusted Partner program? What objectives is it trying to achieve? What does Meta expect from the program? And what should Meta's partners' expectations be? Currently the answers to these questions are not clear – or certainly not to Meta's Trusted Partners.

For it to be effective there must be a genuine and transparent co-design process between Meta and the partners who participate in the program. If the partners are not involved with the design of the program, then the program will not meet partner needs, and the problems outlined in this review will continue.

A complete program overhaul is obviously a significant undertaking and will take time to deliver. Meta should begin by making a public commitment and sharing a plan and timeline with current Trusted Partners.

2 Decision making and enforcement

Meta should outline a clear decision-making process and ensure it is well understood by Trusted Partners. This process should make it clear who is empowered to make decisions relating to Trusted Partner reports and how they are accountable. Policies and decision-making processes around **enforcement actions that rely on input from expert partners – such as those around misinformation and harm - must be urgently clarified** in a way that is transparent and understandable to any partners who may be involved in this process.

3 Program resourcing and staffing

Meta must resource the Trusted Program adequately. Team structure, KPIs and incentives should be aligned with responsiveness and partner needs. The number of staff engaged on the Trusted Partner project must be sufficient to meet performance targets. Meta should regularly evaluate its performance and ensure its resource allocation is in line with requirements.

4 Targets and Transparency

Meta should set clear targets for response times to Trusted Partner reports. These targets should be developed in consultation with partners to ensure they align with requirements, and could potentially include different targets for different categories of report. Targets should be public. They should include a target for average response time, as well as a target for maximum response time.

Meta should measure its performance against these targets and report back to partners on a regular basis. Meta should share with each partner the figures for their overall performance, as well as their average and maximum response times for the individual partner's reports submitted during the reporting period.

In feedback provided by Meta to a draft version of this report (see ANNEX 2) Meta stated that "We are indeed working to develop new methods of sharing information about the overall impact and performance of the Trusted Partner program, consistent with security, confidentiality preferences, and data protection of the many hundreds of organizations who participate." If Meta are already working on these new methods they should be working with current Trusted Partners to ensure these methods meet partner needs. We urge them to engage Internews and other partners who have contributed to this report in this process.

Recommendations

5 Emergency response

To adequately respond to the most urgent cases there must be a straightforward and transparent way to prioritize urgent cases. Whatever this process is, it must empower Trusted Partners to determine which of their cases are most urgent. It is Trusted Partners who are the acknowledged context subject matter experts, and it is they who best understand the potential impact of any given issue (regardless of the relevant Meta policy). This could be a traffic light system or even a separate dedicated channel. The design of this process should be undertaken in partnership with the partners who will use the channel to ensure that it meets their needs.

6 Counterparts and focal points

Each Trusted Partner organization should have designated and specific counterparts within Meta's Trusted Partner team who are looped into all their reports. This system should be transparent and well understood by the partners. Trusted Partners should know who their counterparts are and be able to contact them for any questions or needs connected to the Trusted Partner program. Obviously one Meta staff person can serve as counterpart to multiple Trusted Partners. Contacting a counterpart should not replace using the regular Trusted Partner channel. Counterparts should have familiarity with the country or context in which their Trusted Partner counterpart is operating and should be empowered to action or prioritize cases.

7 Partner consultation

Meta must work with Trusted Partners to establish clear and transparent, mutually agreed guidelines around partner consultation. These should set clear expectations for when consultation will take place, as well as include provisions around notetaking and sharing of minutes for partners' approval, to ensure partner input is documented and fully understood by both parties.

8 Dashboards and reporting templates

Multiple partners interviewed for this report suggested that a Trusted Partner dashboard would make the program more effective and help resolve some of the other issues raised above, including that of security. Partners envision a dashboard that would allow them to see all the reports they had submitted to Meta and track their status (e.g. 'open', 'resolved', etc.). This dashboard could also be used to view analytics, including Meta's response times and how often Meta acted on a partner's reports. This would both make Meta more accountable and help partners improve the quality of their reports. Training materials and other resources could also be accessed through such a dashboard.

Similarly several partners stated that they would prefer to submit their Trusted Partner reports through a standardized template, to ensure that they were sharing all the information that Meta needed to make the correct decision. Other partners rejected this idea and preferred the simplicity of reporting through the in-app mechanism or via email. While flexibility is important, adding the option of reporting via a template could help partners who are unsure of what to include, and improve the quality of reports. Having an alternative to email that does not involve using their personal platform accounts may also improve the digital security of partners particularly at risk of reprisals for their contributions to the Trusted Partners program.

No new product or process changes of this kind should be introduced by Meta without consulting widely with participating Trusted Partners and conducting co-design and product testing as the company would with a commercial product.

Recommendations

9 Partner coordination

Participants in Meta's Trusted Partner program would benefit from the ability to coordinate, share their experiences and support one another's work. The partners interviewed for this review expressed a strong desire to meet other partners and explore areas of collaboration and mutual aid. This should be supported and facilitated wherever possible.

A major barrier to partner coordination is that currently only Meta has a list of all the Trusted Partner program participants. There are valid privacy and security reasons why Meta may not share this list and we do not recommend the list be made public. However, Meta should facilitate coordination efforts by notifying its Trusted Partners of future coordination initiatives and providing them with details of how to participate and self-organize should they so choose.

10 Email whitelisting and bounce-backs

Meta should allow Trusted Partner organizations to add as many email addresses to the Trusted Partner whitelist as they need.

Meta should immediately add an automated bounce-back response to any emails sent to the Trusted Partner reports email from non-whitelisted email addresses. This bounce-back email should make it clear that their report has not been received by Meta.

Internews remains available to discuss the findings of this review and the Trusted Partners' feedback and assist Meta in any of these matters.

Annex 1. Meta's Response to Review Questions

Part One: Questions put to Meta

The following questions were put to Meta in June 2022 in the below format. The answers provided by Meta in February 2023 can be found below.

According to Meta's transparency centre: "[Meta's] network of trusted partners includes over 400 non-governmental organisations, humanitarian agencies, human rights defenders and researchers from 113 countries around the globe."

1. Can you confirm the current number of trusted partners enrolled in the program?
2. Can you confirm the current number of countries covered by the Trusted Partner program?
3. How is the geographic distribution of trusted partners calculated? Does Meta have at least one dedicated partner representing each of the 113 countries included? Or are some countries included due to coverage by an organization with an international scope, such as Internews, that is considered to represent multiple countries?

Sharing the exact number of trusted partners in specific countries may in some cases be sensitive, however we cannot identify any potential privacy or safety concerns that could arise from sharing regional totals. With that in mind,

4. How many Trusted Partners do you have in Africa?
    a. How many separate countries do these represent?
5. How many Trusted Partners do you have in Asia and the Pacific?
    a. How many separate countries do these represent?
6. How many Trusted Partners do you have in LATAM?
    a. How many separate countries do these represent?
7. How many Trusted Partners do you have in MENA?
    a. How many separate countries do these represent?
8. How many Trusted Partners do you have in Europe and Eurasia?
    a. How many separate countries do these represent?
9. How many Trusted Partners do you have in North America?
    a. How many of these North American partners are not from the USA?
10. Does Meta have defined criteria for partner participation in the Trusted Partner program? If so, can you share these criteria? If they can't be shared can you explain why?
11. Is there a vetting process for participation in the Trusted Partner program?
12. When were these criteria and/or processes adopted? Do all current partners meet the current criteria and vetting requirements?
13. Was there any external consultation process before drafting participant criteria? If so, who was consulted?
14. How many Trusted Partner reports does Meta receive?
    a. How many Trusted Partner reports were submitted in the most recent month?
    b. How many Trusted Partner reports were submitted in the last 12 months?
15. Can Meta explain how reports are processed, prioritized, and directed to internal teams once they have been submitted?
16. At what stages is automation used in assessing Trusted Partner reports?
17. When reports are submitted or involve content in languages other than English how are these dealt with?
18. What role does the report's geographic location have on prioritizations (e.g. are some countries prioritized over others)?
19. Trusted Partners have shared with Internews that that many reports they submit never receive any response from Meta. Is there any internal policy which outlines circumstances when a response is not required or where not providing a response is acceptable?

   a. If so, can these guidelines be shared with partners?
20. What percentage of Trusted Partner reports receive any response from Meta (not including an automated response to say that the report has been received).
   a. In the most recent month?
   b. In the last 12 months?
21. For those reports that do receive a response, what is the average time between the moment that the report is submitted by the Trusted Partner to the time at which they receive a response from Meta (not including an automated response to say that the report has been received)?
   a. In the most recent month?
   b. In the last 12 months?
22. Does Meta have internal targets for response rates and times to Trusted Partner reports?
   a. If so, can these targets be shared with partners?
   b. If targets cannot be shared, why not?
23. What percentage of Trusted Partner reports result in any content moderation action taken by Meta?
   a. In the most recent month?
   b. In the last 12 months?
24. How many staff does Meta have working full time on the Trusted Partner program?
   a. How was this resourcing level calculated?

Part Two: Responses Provided by Meta

The following responses were provided by Meta in February 2023. The time between original submission of questions and eventual response was 200 days. Meta reformatted the questions as copied below. A number of questions were not answered, particularly questions relating to response times and response rates.

Can you confirm the current number of trusted partners enrolled in the program? Can you confirm the current number of countries covered by the Trusted Partner program? How is the geographic distribution of trusted partners calculated? Does Meta have at least one dedicated partner representing each of the 113 countries included? Or are some countries included due to coverage by an organization with an international scope, such as Internews, that is considered to represent multiple countries?

There are **465 organizations** enrolled in Meta's Trusted Partners program, covering **122 countries**. Meta has at least one Trusted Partner in each of the 122 countries. In addition to local partners, we work with regional organizations that cover multiple countries. These groups are counted once in the network of partners.

Does Meta have defined criteria for partner participation in the Trusted Partner program? If so, can you share these criteria? If they can't be shared can you explain why? Is there a vetting process for participation in the Trusted Partner program? When were these criteria and/or processes adopted? Do all current partners meet the current criteria and vetting requirements? Was there any external consultation process before drafting participant criteria? If so, who was consulted?

In selecting our Trusted Partners, we seek organizations that have experience in social media monitoring, an interest in learning about our content policies, demonstrate a commitment to keeping online communities safe, and represent marginalized groups who are disproportionately affected by harmful content.

We require that Trusted Partners be independent from direct government control or influence and non profit organizations.

- We acknowledge that government control or influence may or may not include influence as a result of government funding or being a statutory body set up by governments. These organizations tend to have an independent mandate or are NGOs that are dependent on government resources yet maintain independence from government influence. For example, we have partnered with international organization (e.g. UNHCR) which are overseen by member states, yet maintain a level of separation from direct government influence.
- With respect to organizations that are profit generating, we recognize that in certain political contexts, the NGO sector is so severely restricted that for-profit entities are the only types of registrations permitted. In these cases, and with guidance from our internal due diligence teams, human rights function, and regional public policy, we may make exceptions and partner with for-profit entities.

These criteria were informed by consultations with civil society organizations and academics who provided feedback on Meta's approach to inclusivity, community engagement, and transparency.

All candidate organizations are screened by our due diligence team. Our team uses public records, open-source information as well as proprietary databases to assess potential risks. The methodology includes scoping the jurisdiction and verifying corporate/official registration information; conducting strategic media research in English- and local-language(s) to identify derogatory media reporting and noteworthy information; screening against global sanctions and compliance watchlists; and utilizing open source and proprietary databases to identify lawsuits and regulatory records associated with the organization.

These processes were formalized in 2019. In 2022, following the re-opening of the channel after an 18-month onboarding freeze, we introduced a trial period for all new Trusted Partners. This process entails a partnership review, 3-6 months after new partners have been onboarded. We also introduced a Trusted Partner Code of Conduct which outlines roles and responsibilities of Meta and Trusted partners to ensure mutual accountability and establish guardrails for protecting the integrity of the program.

How many Trusted Partner reports does Meta receive?

The Trusted Partner Channel receives about 1000 escalations per month.

Can Meta explain how reports are processed, prioritized, and directed to internal teams once they have been submitted?

Automation is not used in assessing Trusted Partner reports. Reports submitted to the Trusted Partner Channel (TPC) are received directly by our 24/7 Global Escalations team. The escalations team will first assess the content for priority level, to ensure that anything that might result in imminent harm goes to the absolute top of the queue. We take into account a range of factors including known violence or crisis in the country or region, whether there is an ongoing or near-term election, and the policy area implicated.

Depending on the language, the type of violation of the reported content, and other particulars of the report, our escalations team may loop in other teams in order to help assess the content. For example, if the content is not in English, we will loop in a native speaker. If the content is harmful misinformation, we will have an expert on our misinformation policies help assess. If we need further information if the content is an edge case – not clearly a violation of our policies – we may loop in our Content Policy team, who writes the policies.

After cross-functional teams align on the appropriate next steps, Meta will resolve the escalation, which could mean removing or restoring content or accounts, or taking additional action such as disabling a hashtag, or alerting other internal teams to a concerning trend. Meta will then inform partners of the action taken.

Trusted Partners have shared with Internews that many reports they submit never receive any response from Meta. Is there any internal policy which outlines circumstances when a response is not required or where not providing a response is acceptable? If so, can these guidelines be shared with partners?

What percentage of Trusted Partner reports receive any response from Meta (not including an automated response to say that the report has been received). In the most recent month? In the last 12 months?

For those reports that do receive a response, what is the average time between the moment that the report is submitted by the Trusted Partner to the time at which they receive a response from Meta (not including an automated response to say that the report has been received)? In the most recent month? In the last 12 months?

Does Meta have internal targets for response rates and times to Trusted Partner reports?

We recognize that the Covid pandemic severely impacted our operations and resulted in poor reporting experiences for our partners from 2019 - 2021. During this period our content review teams operated at limited capacity and were unable to respond as quickly to trusted partner channel reports as we would like and as they have done in the past. Under these difficult circumstances, we prioritized the most harmful content for our teams to review, such as risk of imminent physical harm or violence.

In 2022, we were able to improve our overall operational resources for content review teams and are steadily increasing our ability to respond to Trusted Partner reports in a timely manner.

We generally expect reports to be reviewed and actioned within 1 to 5 days, though especially complex cases may take longer. All Trusted Partner reports receive an automated response acknowledging receipt, though we acknowledge there have been issues with this in the past.

The goal of the Trusted Partner Program is to provide a channel for expert civil society organizations to report content and accounts which require deeper contextual understanding to be reviewed effectively. Trusted Partner reports often include edge cases or point to broader trends that can require complex investigations and the contribution of multiple Meta teams to evaluate. As a result, Trusted Partner reports can sometimes take longer to be investigated and reviewed.

What percentage of Trusted Partner reports result in any content moderation action taken by Meta?

We don’t specifically break down any actions we take as a result of reports through this program. However, we publish quarterly reports on the content we take down on Facebook and Instagram which you can find here: https://transparency.fb.com/data/community-standards-enforcement/

How many staff does Meta have working full time on the Trusted Partner program?

The Trusted Partner Program is jointly managed by Meta’s Content Policy and Global Operations teams, working in close collaboration with regional Public Policy teams who are responsible for overall relationship management with local partners. The Content Policy team leads in the definition of program strategy, develops training materials, and coordinates outreach with Trusted Partner organizations. The Operations team receives and actions Trusted Partner reports. While we can’t share specific numbers, there are more than 50 people across Content Policy and Operations who work on the Trusted Partner program, and many more regional policy leads who hold relationships with NGOs in their region.

Annex 2. Meta's Feedback to Draft Report

Internews shared a draft version of this report with Meta for their comments and feedback in April 2023. At this time the draft report was also shared with contributing Trusted Partners for their input. In May 2023 Meta provided the below response. Elements of this feedback were incorporated into the final draft.

We'd like to take you up on your offer to provide feedback on the Report. For purposes of these comments, we'll be focusing on areas where the Report incorrectly frames issues or omits information we think is important.

We welcome constructive dialogue with any of our Trusted Partners to help us improve and keep people safe on our platforms. To inform these conversations we'd appreciate knowing whether there is any existing program that has some or all of the elements you're seeking -- that would be very helpful for our learning.

Putting Trusted Partner in Context

In several places, the Report presents Trusted Partner as a singular mechanism for emergency response. It's important to bear in mind that Trusted Partner is one of several reporting channels. As you know, Meta also operates a human rights defender response channel, in collaboration with digital security organizations, to address security risks faced by human rights defenders around the globe. We also work with Access Now's Digital Security Helpline to ensure all human rights defenders have access to these digital security resources, whether or not they are Trusted Partners.

The Trusted Partner program is part of Meta's broader enforcement ecosystem, including the use of technology to proactively identify and remove harmful content, and content review teams applying Meta policies to the millions of pieces of content reported every day.

Support for Trusted Partners

The Report does not accurately reflect the nature or scope of the Trusted Partner grants program. These grants are not intended to provide payment for a service. Rather, the program seeks to support the organizational sustainability of civil society partners operating in challenging environments with limited resources. We consider this an ecosystem level investment to support a more resilient and independent civil society sector.

Since 2021, Meta has provided donations to Trusted Partners based outside North America and Western Europe who demonstrate consistent engagement with the channel. The recipients constitute roughly half of our partners located in Latin America, Asia, Africa and the Middle East.

Many recipients have told us that the grants have a significant positive impact on their work – e.g., by funding fact-checking initiatives, staff training, and overall operational costs. We appreciate that you've been clear about the limits of the methodological design in your Report and the small percentage of Trusted Partners your research has been able to access.

Consultation

The Trusted Partner program rests on deep consultation with Partners. For example, when Trusted Partners in certain countries told us that reporting content by email posed safety risks, we created a secure in-app reporting mechanism for them to use. The grants program was likewise designed in consultation with civil society organizations, to respect organizations' preference for independence, while at the same time providing essential resources to groups operating in resource constrained environments. There are many other such examples.

We also support networking and exchange between Trusted Partners while respecting partners' requests to remain anonymous. In October 2022, for example, we organized a Middle East

Community Summit that included Trusted Partners from across the region, and we have hosted similar events for Sub-Saharan Africa and for the Asia Pacific region.

Finally, we engage frequently and productively with Trusted Partners on content policy issues. For example, the policy under which Meta removes misinformation "where it is likely to directly contribute to the risk of imminent physical harm" was designed and is enforced with the input of many Trusted Partners, and Meta deeply appreciates this engagement. Our commitment to listen to Partners does not mean, of course, that we will revise our policies in response to each individual piece of feedback we receive. Similarly, we do not remove all content reported to us as violating by Partners; our policies dictate what is and isn't removed.

Prioritization and Response Time

We prioritize the most harmful content for our teams to review, such as risk of physical harm or violence. As we noted in our initial response to Internews' research questions, we use a variety of indicators to establish the priority of a report. These include whether there is known violence or crisis in the country or region, whether there is an ongoing or near-term election, the policy area implicated, and the severity of the potential content-related risks.

Partner reporting behavior, including compliance with reporting protocols and partners' familiarity with our Community Standards, also significantly affects response times. We provide training to facilitate clear reporting and are grateful for the time and efforts groups have dedicated to this work. We have created extensive online educational materials, supplemented by direct engagement, to strengthen Partner reporting and improve the speed and efficiency with which we ingest content reports.

We acknowledge the variety of Partner experiences documented in the Report, and we are committed to continue improving training resources and ingestion systems to address these outliers and strengthen the program. However, the reporting issues of the small sample of Trusted Partners who contributed to the Report do not, in our view, represent a full or accurate picture of the program.

Transparency

We recognize the value of increased transparency, both with respect to clarity on shared goals and performance, and to recognize the significant impact of Trusted Partner reporting. We note your Report also recognizes the need for strong operational security protocols; this accounts for some of our program design and structures. While data protection laws prevent us from sharing information about actions taken with regard to other users, we strive to provide high-level feedback to our partners through group consultations and one-on-one debriefs.

We are indeed working to develop new methods of sharing information about the overall impact and performance of the Trusted Partner program, consistent with security, confidentiality preferences, and data protection of the many hundreds of organizations who participate.

Reporting Templates & Dashboards

We appreciate the need for clear reporting guidelines and tracking mechanisms for Trusted Partner reports. The cross-functional teams that support the Trusted Partner program have developed tools to address these needs, including our online learning platform launched in 2022 that provides self-paced training on Meta content policies and reporting best practices. We are in the process of developing standard reporting templates, tailored for different harmful content types, such as hate speech, violence and incitement, and misinformation and harm, to further facilitate reporting from partners.

Our recent debrief with Internews Kyrgyzstan on how to report Coordinated Inauthentic Behavior reflects best practice in this regard and we are committed to providing more of these opportunities for constructive dialogue in the future.

With respect to dashboards, these are available to Trusted Partners on Facebook and Instagram through the in-app reporting function. We are also exploring the feasibility of implementing the recommendation on a reporting bounce-back email for non allow-listed users seeking to access the reporting channel.

We welcome the opportunity to consult with partners on these new reporting tools to ensure they are as effective as possible.