
How Can Platform Engagement with Academics and Civil Society Representatives Inform the Development of Content Policies?

A look at Meta’s COVID-19 Misinformation Policies

Peter Stern, Sarah Shirazyan, and Abby Fanlo

1 Introduction

On January 30, 2020, the WHO designated the novel coronavirus a “public health emergency of international concern.”¹ The same day, Facebook announced it would begin removing “content with false claims or conspiracy theories that have been flagged by leading global health organizations and local health authorities that could cause harm to people who believe them.”² These efforts focused on removing claims designed to discourage treatment for, or taking appropriate precautions related to, COVID-19. This type of content included claims related to false cures or prevention methods or claims that created confusion about available health resources.³ On March 25, 2020, Facebook announced it would remove additional false claims, including that physical distancing doesn’t help prevent the spread of COVID-19.⁴ In December 2020, the company offered a further update, announcing that it would remove additional claims, including false claims that COVID-19 vaccines contain microchips or that the vaccines are tested on specific populations without their consent.⁵ In February 2021, Facebook announced that it would remove widely debunked false claims about vaccines, including that vaccines cause autism or cause the disease they are meant to protect against.⁶

On what basis did Facebook⁷ arrive at these policy decisions? 2020 and 2021 were times of considerable uncertainty: people questioned what was true and disputed who could be trusted. At the same time, scientific understanding about COVID-19 was evolving, raising questions about how to verify claims regarding the virus and associated treatments. Existing scholarship on crisis informatics suggested that people would turn to informal networks, including family and friends, for information amid this ambiguity and panic. This literature also suggested people would seek information about the virus on social media. So how did Facebook determine how to address

1. CDC 2022.

2. See Jin 2020, update that was originally published on January 30, 2020.

3. Ibid.

4. Clegg 2020.

5. See Jin 2020, update that was originally published on December 3, 2020, and titled “Removing False Claims About Covid-19 Vaccines.”

6. Rosen 2020, update that was originally published on February 8, 2021 and titled “Removing More False Claims About COVID-19 and Vaccines.”

7. This paper refers to Facebook when discussing any policies and engagement surrounding those policies that occurred prior to the name change to Meta.

false claims about the coronavirus? As the pandemic evolved, how did Facebook apply its existing misinformation policies to what users were saying about COVID-19 on its platforms?

Before COVID-19 emerged, content policy teams within Facebook had already begun thinking about how existing misinformation policies could apply to harmful health misinformation. However, the outbreak of COVID-19 focused attention more closely on misinformation related to the virus. More specifically, the policy team wanted to understand whether experts believed misleading health claims could contribute to physical harm offline, in the context of a global pandemic, and whether the company should *remove* posts containing false COVID-19 claims; rely solely on *labeling* such content and *reducing* its distribution on the platform; or *take no action* at all. Further, the policy team needed guidance on the specific false claims associated with the pandemic that should be subject to policy enforcement.

A few things made this question particularly difficult. First, given that scientific understanding evolves rapidly during a crisis, it was unclear how to determine the veracity of claims about the disease at any particular moment—or whether such certainty was even possible. Second, claims about the virus could impact users in one location differently than in another, due to geographic variation in the pandemic’s progression and in access to healthcare and other resources.

Protecting public safety while respecting expression is a core challenge of content policy development at Meta, including in formulating the company’s approach to COVID-19 misinformation. Striking a balance between free speech and its legitimate curtailment is complex and nuanced; moreover, given the company’s global user base, Meta’s policies will impact people in ways that vary drastically around the world.

Often, the effort to deal with these challenges in the policy development process at Meta involves gathering input from academics, civil society representatives, community leaders, and users around the world who have relevant subject matter expertise or lived experience. A dedicated team within Meta’s Content Policy organization carries out this stakeholder engagement function.⁸

The work of the Content Policy Stakeholder Engagement team, as it has developed at Meta, entails a rigorous methodology that seeks to bring a diverse and well-informed set of viewpoints into the policy development process. The primary goal of this article is to shed light on that process, by describing how the Content Policy Stakeholder Engagement team solicited input on Meta’s approach to COVID-19 misinformation, how that feedback was shared internally, and ultimately how that feedback helped fashion Meta’s misinformation policies during the pandemic. However, it is important to emphasize that stakeholder engagement is only one piece of the content policy development process at Meta—there are many additional inputs, including company priorities and values—that inform our policy decisions. Another objective of the piece is to illuminate the challenges associated with stakeholder engagement work, including identifying a diverse set of stakeholders to consult, reconciling variation in stakeholder input, and providing transparency about the people and organizations the team engages with. Finally, this article seeks to share learnings about addressing misinformation that came from engagements on the topic, in order to show the value of this methodology for surfacing feedback and ideas for content and product policies. Overall, by describing Meta’s approach to stakeholder engagement in the case of COVID-19 misinformation, the article hopes to contribute to the rapidly developing field of stakeholder engagement.

8. Facebook Transparency Center 2022a.

2 Content Policy Stakeholder Engagement at Meta

Meta’s Content Policy Stakeholder⁹ Engagement team is dedicated to identifying relevant external stakeholders on various policy topics, collecting their perspectives, and disseminating those views internally. The team brings external feedback into the development of Meta’s Community Standards, which govern the content that users are permitted to share on Facebook and Instagram. The team also contributes to the formulation of policies aimed at addressing low-quality or problematic content that does not meet the threshold for removal from Meta’s platforms. For example, stakeholder feedback helped shape the development of policies that govern Meta’s ranking¹⁰ and recommender systems¹¹ and that address overt or covert influence operations.¹²

In this emerging field, there are important questions to be asked about how stakeholders are selected for consultation and about the fairness and efficacy of relying on representative groups as a means to scale engagement with a worldwide user base.¹³ Some commentators have suggested sweeping changes to tech companies’ engagement practices, including public notice and comment-type mechanisms, or juries and referenda bodies composed of lay people tasked with providing input on content policies.¹⁴ Meta’s Content Policy Stakeholder Engagement team is actively considering ways to improve, in dialogue with scholars and internal teams at Meta whose work touches on governance and stakeholder engagement.

Under the team’s current approach, identifying specific stakeholders to engage with is a process guided by three core principles: expertise, inclusivity, and transparency. These principles, which are described in more detail below, facilitate engagement with stakeholders that have a wide spectrum of backgrounds and viewpoints. This breadth enables the team to weigh the impact of Meta’s policies and understand consequences that the team might not otherwise have considered.

2.1 Expertise

Crafting platform rules to govern online speech is a highly complex process. While Meta has invested in hiring people with relevant knowledge in a wide range of policy areas, it is not possible for Meta to have expertise on every emerging policy challenge. Moreover, the range of views among experts can vary significantly—so relying on internal expertise alone might limit the perspectives considered in the development of a policy. The team therefore engages with academics, NGOs, and international institutions with expertise

9. The term “stakeholder” is fluid and potentially ambiguous. For Meta’s Content Policy Stakeholder Engagement team, the word refers to groups and individuals outside of Meta, including civil society organizations, activist groups, thought leaders, and academics, that can provide feedback on the development of Meta’s content policies.

10. Facebook Transparency Center 2022c.

11. Facebook Transparency Center, n.d.

12. Gleicher 2020.

13. Brenda Dvoskin at Harvard has written compellingly about these challenges and limitations. See, e.g., Dvoskin, n.d., at 14–15, and passim: “At Facebook, consultation with civil society is a key step in the development of all changes to content moderation rules. Dedicated staff consult with a broad group of people around the world who bring diverse perspectives. Unlike a notice-and-comment process, participation is not open to the public. Rather, input is solicited to specific stakeholders. This means that access is heavily controlled by the company ... Facebook does not promise to balance or account for all the interests expressed by civil society. No specific guidance to weigh in stakeholders’ ‘comments’ exists. However, because it is a standardized phase of policy development, and the outreach in the United States is broad and inclusive of diverse viewpoints, it is the closest model encouraging institutional participation of civil society in rulemaking processes.” For a conceptual overview, see Gorwa 2022.

14. See Zittrain 2022, which encourages consideration of “community governance” mechanisms to help draw difficult content policy lines and to “do so in a way that confers legitimacy on the participants’ decisions.”

in particular policy areas, such as digital and civil rights, and international human rights law. The team takes a broad view of “expertise” and frequently engages with individual stakeholders or advocacy groups who can share lived experience relevant to the policy under consideration, including groups that may be particularly affected by the type of speech a policy is intended to address.

2.2 Inclusivity

A focus on building an inclusive stakeholder base is also embedded in Meta’s engagement process. Stakeholder breadth and diversity—in terms of geography, personal background, ideology, viewpoint, and other factors—is critical for the development of Meta’s policies. How online content is perceived varies significantly based on local context, including recent political events or medical crises. Moreover, views on content policies are often informed by local levels of institutional trust or press freedom. For example, allowing certain forms of adult nudity may seem unacceptable to individuals in the Middle East, while removing such content might seem unacceptable to individuals in Scandinavia. Stakeholder engagement seeks to flesh out these differing perspectives, and the team regularly consults with organizations whose policy positions not only fall on a broad spectrum but also are contradictory, to ensure that policies are based on multiple inputs.

Ensuring regional diversity in engagement is an important part of this work. While Meta’s policies are global, each part of the world wants the company to understand local content issues at a deep level. Engagement provides a means to meet that demand—but it poses challenges. A content or product policy issue may gain significant attention in one region but remain out of the spotlight elsewhere; if public commentary or academic research on a policy issue is underdeveloped in a given region, it can be difficult to identify local experts. In authoritarian regimes, or countries that are undergoing a democratic transition, independent civil society may be nascent. In those countries, the team may sometimes find a more limited pool of stakeholders, given the goal is to engage with independent and representative voices. Whenever possible, the team tries to mitigate this challenge by engaging with academics who have expertise in that particular country, even if they are not physically based in the region.

The team also engages with stakeholders representing diverse identities (gender, sexuality, ethnicity, religion etc.). Including these communities in engagements carries challenges of both identification and accessibility. While some marginalized or historically underrepresented groups may have coalitions that represent their interests, other groups may face barriers—local laws, cultural practices, or socioeconomic status, for example—to organizing and speaking up. Moreover, in an age when engagement is often virtual, stakeholders generally must have access to a computer or telephone to connect with Meta. Mindful of such structural and societal barriers to engagement, there is now a dedicated function within the team that focuses on developing inclusive engagement strategies to ensure that such communities have an opportunity to share their perspectives with Meta. This group has developed practical ways to meet stakeholder knowledge and accessibility needs, including the provision of prepaid data cards, sign language support, and translations.

Finally, viewpoint diversity is also a core part of stakeholder engagement. The team deliberately seeks out advocates with differing perspectives in any policy development. In most cases, identifying stakeholders with views on multiple sides of a policy issue is possible. Regularly consulting with organizations explicitly dedicated to advocating for values across the policy spectrum also helps to ensure that Meta’s policies are built on a broad set of viewpoints.

2.3 Transparency

Transparency is another central element of this team’s work. First, any engagements provide stakeholders with an overview of the policy development process at Meta and include a discussion of the stakeholder engagement function. Second, the team works to follow up with stakeholders after Meta reaches a final policy decision, to explain the policy choice in light of stakeholder feedback. Sometimes, this follow up takes a long time, and stakeholders are eager to hear back more quickly. To address this issue, the team is working on ways to provide interim updates to stakeholders.

Increasingly, external stakeholders also want to know the names of the individuals and organizations that Meta consults on a particular policy. They argue that this kind of transparency is essential for auditing Meta’s engagement practices and for helping stakeholders determine “what other stakeholders they are up against” in advocating for their views with Meta.¹⁵

Leaving aside the “zero sum” assumption built into this question, which may be debatable, this is a question the team continually reflects on. However, the current practice is not to share the names of people or organizations who privately provide feedback on Meta’s policies, without prior consent. There are several reasons for this approach. First, many stakeholders seek confidentiality in speaking with Meta. Often the issue is safety: the need for protection is especially acute for members of vulnerable communities, who may be the victims of harmful speech. Second, Meta’s policies will be better if stakeholders are completely frank, without concern that their decision to offer feedback to Meta will give rise to public criticism. As a practical matter, if the team were to decide that transparency should trump such concerns, the pool of stakeholders willing to consult with Meta would almost certainly shrink.

The team’s work is guided by those three principles—but what does that look like in practice? Is there a methodology for applying these principles to stakeholder engagement? The next section answers these questions in the setting of COVID-19.

3 Stakeholder Engagement in Practice—the Example of COVID-19 Misinformation

Prior to the COVID-19 pandemic, Meta adopted a three-pronged policy approach to misinformation. The company would *remove* false claims if they could directly contribute to imminent physical harm, while otherwise *reducing* the distribution of content determined by third-party fact-checkers to be misinformation and *informing* users about such falsity. Meta’s strategy in this area was shaped by the advice of experts, against a backdrop of existing company policies, values, and experience. This section describes several important pieces of expert feedback that contributed to Meta’s approach.

First, misinformation presents significant conceptual and definitional challenges. Experts representing international institutions, human rights defenders, and free expression advocates have consistently emphasized the difficulty of defining misinformation. People have different levels of information about the world around them and may believe that something is true when it is not. Many daily statements accepted as true are not rigorously verifiable. Moreover, it can be difficult to separate “misinformation” from opinion. For example, try contrasting the statement “2+2=4” with “Bill Gates

15. Dvoskin, n.d.

plans to implant microchips in COVID-19 vaccines.” Anyone can independently verify the truthfulness of the first statement according to the basic—and widely accepted—rules of arithmetic, but there is no such accepted process—or set of rules—for independently verifying the latter one. To navigate whether personal opinions, speculation, and everyday commentary are true or false, people have typically looked to societal institutions like universities, government authorities, newspapers, broadcast journalists, courts, and international organizations; however, trust in such institutions has been declining.¹⁶ Altogether, these realities make defining and identifying misinformation extremely challenging for anyone—including tech platforms.

Second, under international human rights law, everyone has the right to freedom of expression, which includes the freedom to hold opinions, as well as the right to receive and impart information and ideas. International human rights law protects speech that “offends, shocks or disturbs”¹⁷—“irrespective of the truth or falsehood of the content.”¹⁸ This does not mean that false or misleading information can never be restricted, but, as the UN Special Rapporteur on Freedom of Expression, Irene Khan, has stated, any restriction on such information must establish a close and concrete connection to the protection of one of the legitimate aims under international human rights treaties.¹⁹ Special Rapporteur Kahn has also made clear that prohibition of false information is not in itself a legitimate aim under international human rights law.²⁰ She has explained that the directness of the causal relationship between the speech and the harm, and the severity and immediacy of the harm, are key considerations in assessing whether the restriction is necessary.²¹ Similarly, false information may be prohibited if it rises to the level of incitement to violence, but the relevant treaty provisions do not make explicit reference to untruthful information.²² Freedom of opinion can never be restricted. In fact, many experts—particularly freedom of expression advocates—argue that private companies should not determine truth. Jacob Mchangama (Justicia), for example, believes it would be “dangerous” for private companies to determine truth versus falsehood and impossible for them to solve the trust deficit in societal institutions. Mchangama argues that Meta should be franker about this limitation: “You should say, ‘Unless we see misinformation and disinformation being used to create clear harms, like Myanmar, we don’t see it as our mission to try and adjudicate the truth.’”²³

Lastly, researchers studying political instability, experts in crisis informatics and rumoring, and civil society groups focused on genocide prevention have underscored that any link between misinformation and harm will be highly context specific. A false claim might incite violence given the historical context of a particular place, or recent intergroup conflict there, but a similar claim might be much less likely to contribute

16. Perry 2021; Brenan 2021; Edelman 2022.

17. *Handyside v. The United Kingdom* 1976.

18. See A/HRC/47/25: *Disinformation and freedom of opinion and expression - Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 2021, para. 38. See also *Joint declaration on freedom of expression and “fake news”, disinformation and propaganda* 2017 and Shirazyan et al. 2020.

19. A/HRC/47/25: *Disinformation and freedom of opinion and expression - Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 2021, para. 40.

20. para. 40.

21. para. 41.

22. Relevant treaties and instruments of International Human Rights Law include *Universal declaration of human rights* 1948, 217 A (III); *International Covenant on Civil and Political Rights* 1966, p. 171, Art. 19 and 20; *International Convention on the Elimination of All Forms of Racial Discrimination* 1965; *European Convention for the Protection of Human Rights and Fundamental Freedoms* 1950, as amended by Protocols Nos. 11 and 14, Art. 10, 11, and 17; *American Convention on Human Rights* 1969; *African Charter on Human and Peoples’ Rights (“Banjul Charter”)* 1981; *ASEAN Human Rights Declaration* 2012; *General comment no. 34 on freedom of opinion and expression, Article 19*; *Rabat Plan of Action* 2013.

23. Sarah Shirazyan Interview with Jacob Mchangama 2022.

to harm in a different setting. Moreover, experts have suggested that predicting when misinformation might contribute to harm would be very difficult on a global level—but somewhat less so on a regional or country specific basis. Thus, they encouraged Meta to incorporate local signals into any effort to determine when misinformation could lead to imminent physical harm.

For these reasons (and others lying outside the scope of this article), Meta adopted a well-publicized policy of “remove,” “reduce,” and “inform,” set out in public statements,²⁴ in conjunction with third-party fact-checking partnerships. Perhaps most importantly, the pre-COVID-19 policy framework explicitly contained a provision under which Meta would take down misinformation when local partners with relevant expertise told us a particular piece of content (like a specific post on Facebook) could contribute to a risk of imminent physical harm.²⁵ This policy had not been applied for the removal of entire categories of false claims on a worldwide scale, however. During the pandemic, the question for Meta was how to apply its approach to misinformation in this new context.

3.1 Using Engagement to Build Meta’s COVID-19 Misinformation Policies

Throughout 2020 and 2021, stakeholder engagement played an important role in the development of Meta’s approach to COVID-19 misinformation. Stakeholder feedback informed the launch of the initial policies in early 2020, and it also contributed to policies on COVID-19 vaccine misinformation, which launched later that year and in 2021. This section describes the process of engagement throughout these years.

As COVID-19 spread across the world in early 2020, Meta’s policy team began considering how the company’s approach to misinformation that could contribute to imminent physical harm would apply to false claims about the virus. The application of the existing policy to this new issue area raised several important questions. Specifically, the policy team wanted to understand whether experts perceived a link between misleading health claims online and physical harm offline; whether public health authorities believed there was any potential harm from such claims; and whether certain circumstances—such as epidemics or pandemics—made health misinformation particularly likely to contribute to imminent physical harm. For example, would a post claiming “garlic cures COVID-19,” in the midst of the pandemic, pose a risk of harm sufficient to warrant removal? As the pandemic evolved throughout 2020 and 2021, new questions arose, such as how to address misinformation about COVID-19 vaccines, e.g., claims that Bill Gates implanted a microchip in them.

To get feedback on those questions, Meta’s Content Policy Stakeholder Engagement team undertook a four-step process that characterizes the team’s approach to stakeholder engagement: (1) development of an engagement strategy, (2) execution of the strategy, (3) internal communication and consideration of stakeholder feedback, and (4) follow-up with stakeholders about Meta’s final policy decision.

3.1.1 Strategy Setting

Strategy setting is perhaps the most critical part of the engagement process. It involves developing an understanding of the existing literature on a particular topic, framing questions for stakeholders, identifying relevant experts on a topic, and incorporating stakeholder diversity into the engagement process.

Surveying the Literature

24. Facebook Transparency Center 2022b.

25. Meta publicly announced this policy in July 2018. See Newton 2018.

External engagement for Meta’s content policy development is similar to what social scientists describe as “interviewing.” As a data collection method, interviewing is only successful if the interviewer has a baseline knowledge of the topic.²⁶ The first step of engagement is therefore familiarizing oneself with key notions, themes, and actors in the relevant policy area. However, in early 2020, when Meta first began developing a COVID-19 misinformation policy in response to the pandemic, the virus itself was novel and relatively unknown. The team therefore relied on scholarship focused on crisis informatics and health misinformation more broadly.²⁷

Existing work in social science suggested that individuals rely on their social networks for information during crises. When disaster strikes, people are confused and afraid—and consequently engage in “collective sensemaking,” trying to make sense of uncertain and devastating circumstances.²⁸ At first, people turn to official channels for information, such as the mainstream media or government authorities. However, the information from these sources may not be satisfactory; mainstream media may often highlight the most sensational aspects of an ongoing problem²⁹ and government officials may offer limited, contradictory, or delayed information.³⁰ In fact, early on in the pandemic, authorities offered competing guidance on mask usage. The WHO and the US Centers for Disease Control & Prevention (CDC) recommended that only people with COVID-19 symptoms, or those caring for individuals with the virus, wear masks, while other public health experts in China, the Czech Republic, and the United States argued that everyone should wear masks in public spaces to prevent the virus’s spread.³¹ According to academic research, this lack of clarity from authorities—especially during crises—can lead individuals to seek information from the media or from informal networks, including colleagues, friends, or family contacts, rather than official sources.³² For example, during the MERS outbreak in South Korea, a lack of trust in the information disseminated by public institutions increased reliance on informal networks for building understanding.³³ These findings implied people would want to connect socially as the pandemic spread, suggesting social media would be a likely avenue for information dissemination about COVID-19.

Crisis events also render populations more vulnerable to misinformation. Research on rumoring behavior broadly, and wartime communication more specifically, reveals that the urgency and uncertainty surrounding crises facilitate the spread of unverified information. In such situations, people feel that information sharing is critical—due to the importance of the topic—yet the ambiguity of the situation makes verifying facts and sources difficult.³⁴ During public health crises, the relevant science evolves,

26. Hensler 2011.

27. However, as the pandemic continued, specific research on COVID-19 misinformation emerged, and the team stayed up to date on this new scholarship.

28. Starbird, Spiro, and Koltai 2020.

29. See Oh, Agrawal, and Rao 2013, 409. On page 409, the authors state: “The literature maintains that institutional mainstream media have a tendency to repeatedly zoom in on the sensational aspects of a disaster from a single onlooker’s perspective.” See also Tuman 2009, 196. Tuman states “news media are often drawn (like entertainment media in books or movies) to stories that suggest conflict and the potential for what is shocking and sensational.” See also Nacos, Bloch-Elkon, and Shapiro 2007, 112. The authors found that media coverage of terrorism post 9/11 focused on “shocking, sensational, and disconcerting news,” such as new national terrorism alerts. See also Covello, Winterfeldt, and Slovic 1987. The abstract states “the media often play the role of transmitter and translator of information about health and environmental risks, but have been criticized for exaggerating risks and emphasizing drama over scientific facts.”

30. See Jang and Baek 2019, 991. See also Covello, Winterfeldt, and Slovic 1987; Wright 1987; Oh, Agrawal, and Rao 2013; Starbird, Spiro, and Koltai 2020.

31. An article in *Science* outlined the competing expert recommendations regarding wearing masks to prevent the spread of COVID-19. See Servick 2020.

32. See Jang and Baek 2019; Oh, Agrawal, and Rao 2013, 409–10; Mileti and Darlington 1997, 89, and Erickson et al. 1978.

33. See Jang and Baek 2019 and Starbird, Spiro, and Koltai 2020.

34. See Oh, Agrawal, and Rao 2013; Allport and Postman 1947, and Starbird, Spiro, and Koltai 2020.

creating uncertainty about the facts even among experts. However, communicating this uncertainty to the public can be challenging,³⁵ further increasing popular reliance on informal networks for knowledge gathering. Thus, in addition to increasing the likelihood that individuals would turn to social media for information, COVID-19, the existing literature suggests, created an environment in which rumors and misinformation about the virus might spread. Altogether, these findings helped Meta's Content Policy Stakeholder Engagement team better understand the relationship between crisis situations and information dissemination, and that understanding informed the questions that the team posed to stakeholders.

Framing Questions

The next step in engagement is developing questions for stakeholders. The goals of any policy development, as framed internally, are often Meta-centric in their approach and jargon. The Content Policy Stakeholder Engagement team's role is to write questions that are open and accessible to an external audience, yet sufficiently address the core goals of the policy development.

Context and scope:

- How would you define health misinformation?
- Under what conditions does health misinformation spread?
- Can health misinformation lead to physical harm? If so, when?
- What criteria are used to evaluate and then declare a public health emergency?
- What effect, if any, does a country's public health infrastructure have on the consequences of health misinformation?
- What effect, if any, does trust in public health authorities have on the consequences of health misinformation?

Harm:

- How should Meta assess the link (if any) between false health claims and harmful consequences? Should Meta rely on outside organizations to do this? If so, which ones?
- How should Meta consider the potential for mortality vs. other severe health harms?
- Who is an appropriate authority to identify and debunk health misinformation?
- When discussing the severity of risk to individuals or to public health, what criteria do medical professionals and health organizations use?

Action:

- What actions should Meta take in response to health misinformation on the company's platforms (e.g., remove the content, reduce the distribution of the content, add labels, or do nothing)?
- When is labeling or reducing the distribution of health misinformation more proportionate to the potential harm than removing it entirely?
- When is taking no action on health misinformation the proportionate response?
- What interventions can increase trust in public health measures or disease treatments?

35. Holmes et al. 2009.

- What are the unintended consequences of each approach?
- Given that public health guidance on COVID-19 is likely to shift, how can Meta continue to monitor and update its policies and their enforcement? How should the company prioritize updating its policies, especially as Meta has a finite number of human moderators?

Identifying Stakeholders

After framing questions, the team begins identifying categories of stakeholders to speak with, based on the principles of expertise and inclusiveness described earlier in this commentary. To get feedback on how to address COVID-19 misinformation during the pandemic, the team decided to seek out stakeholders with expertise in several different categories: researchers studying misinformation and disinformation; experts in national security and public safety; professional fact-checkers; experts in public health, infectious diseases, vaccine behavior, and health communication; experts in freedom of expression law; and advocates for free speech and digital rights. Additionally, the team sought out stakeholders with expertise in socioeconomic, racial, and regional disparities in health experiences and outcomes. For example, there are communities of people across the world that distrust public health authorities, due to their ongoing lived experience or the historical experience of their community. The team therefore sought to engage with members of such communities to ensure Meta's COVID-19 misinformation policies accounted for those experiences.

The team also reached out to experts with different regional backgrounds. As described earlier in this paper, inclusivity is a core principle of the team's approach to stakeholder engagement. Moreover, prior to the pandemic, public health experts emphasized that the unique political, historical, and medical setting in a country could impact health misinformation and potential harms. For example, when it became public that the CIA used a fake hepatitis vaccination campaign in Abbottabad, Pakistan to help locate Osama Bin Laden, trust in vaccines suffered—especially in Pakistan.³⁶ Campaigns to eradicate polio in Pakistan were hampered by accusations that healthcare workers, who were providing the vaccine, were US spies, and the WHO declared that Pakistan was one of three countries in the world where polio remained endemic.³⁷ Several workers even faced physical attacks. Levels of trust in polio vaccines throughout Pakistan plummeted further in 2019, when a misleading video, which claimed the polio shot left children ill, went viral.³⁸ Rana Safdar, the coordinator of Pakistan's National Emergency Operations Centre for Polio Eradication, said that millions of children were not immunized against polio due to the surge in vaccine hesitancy. Thus, the team expected that stakeholders with different regional backgrounds would have different views on the harm associated with COVID-19 misinformation—and the best way to address it. In fact, during the pandemic, stakeholder feedback on the best approach to COVID-19 misinformation varied by region, driven by differing experiences with vaccine uptake and the penetration of anti-vaccine movements. (The section of this paper on communication and implementation of stakeholder perspectives covers this variation in more detail.) Overall, during the pandemic, the team spoke with stakeholders from every major region of the world, including North America, South America, Africa, the Middle East, the Asia-Pacific, and Europe, about Meta's approach to COVID-19 misinformation.

Execution

After setting the engagement strategy, the team begins reaching out to stakeholders,

36. "Polio Eradication: the CIA and their unintended victims" 2014.

37. Ibid.

38. Hadid 2021.

with an eye toward being transparent about the policy Meta is considering, the questions for discussion, and why the particular stakeholder's expertise or lived experience is relevant. While stakeholders do not always agree to engage, most people are open to having a conversation with Meta. Stakeholders have a variety of reasons for engaging with the company. For Marius Dragomir (Central European University), offering feedback to Meta provided an opportunity for decades-long work on state media and media capture to "finally have practical application and lead to policy interventions."³⁹ While he had hoped for many years that work in his field would lead to changes by governments and supranational organizations, he said that engaging with Meta "was the first time that we've seen our work used meaningfully in practice."⁴⁰ Others, like Jacob Mchangama, are driven by a desire to "make things less bad versus revolutionize Meta's approach to content moderation."⁴¹ Mchangama noted, "we see a lot of movement and pressure to remove content and we want to bring the perspective to advocate for the opposite direction."⁴²

The Content Policy Stakeholder Engagement team typically conducts engagement in a variety of formats: 1:1 discussions, roundtables, and recurring expert circles (which meet regularly to discuss different topics). External engagement in a 1:1 setting is akin to conducting semi-structured interviews in a social science setting, in which the interviewer knows what they want to learn—and thus, has a set of questions prepared—but the conversation is able to vary and change substantially between participants.⁴³ This approach facilitates comparing answers to the same questions across different stakeholder backgrounds and regional contexts, but also allows questions to adapt within and across conversations based on the stakeholders' feedback.

In roundtables and recurring expert circles, the team usually poses a set of questions to the group and opens up a broader discussion. This format allows Meta to convene stakeholders with different subject matter expertise and regional backgrounds in one space, making possible a useful cross-pollination of opinions. Stakeholders can learn from each other, and they often realize that their recommended approach to a policy challenge might not be effective in a different context. Sometimes, these interdisciplinary conversations yield ideas for policies that will work across regions and issue areas. Other times, they reveal more disagreement than consensus—highlighting the challenges of writing policies for a global user base.

Discussions with stakeholders on COVID-19 misinformation during the pandemic occurred in many of these formats, and overall, the team engaged with around 180 stakeholders.

3.1.2 Internal Communication & Implementation of Stakeholder Perspectives

Throughout the engagement process, the Content Policy Stakeholder Engagement team documents and communicates interim findings to colleagues within and outside Meta's Policy organization. Moreover, the colleagues who are writing the relevant policies—or developing relevant products—often join discussions with stakeholders so they can hear directly from them. Once engagement is complete, the team shares a formal analysis of

39. Sarah Shirazyan *Interview with Marius Dragomir 2022*.

40. Ibid.

41. Sarah Shirazyan *Interview with Jacob Mchangama 2022*.

42. Ibid.

43. The team has found the following resources on semi-structured interviewing particularly useful: Leech 2002, which provides techniques for asking effective questions and building rapport in semi-structured interviews; Stofer 2019, which helps place semi-structured interviews in the context of other one-on-one interview methods, and Brinkmann 2014.

all stakeholder perspectives with the colleagues who will ultimately write the policy—as well as other teams that will weigh in on it. The analysis highlights key learnings from stakeholder engagement and identifies trends in the types of stakeholders who offered particular points of feedback. This analysis often informs Meta’s ultimate policy decision.

Several themes emerged in the team’s engagements on COVID-19 misinformation during the pandemic. First, medical and public health experts emphasized that health misinformation could be particularly harmful if it fell into the following categories: false cures, false information designed to discourage treatment (such as false claims about hospital practices or safety), false claims about prevention (such as false claims that masks are not effective in preventing the spread of airborne diseases), false information about availability of access to health resources, or false information about the location or severity of a disease outbreak. Experts argued that such claims could contribute to imminent physical harm by increasing the risk that individuals will ignore the guidance of health authorities, which in turn could increase the risk of individuals contracting the virus and spreading it to others. Given this potential harm, many health experts encouraged Meta to remove harmful health misinformation.

In contrast, many freedom of expression (FoE) advocates opposed removing COVID-19 misinformation, unless it incited violence against particular groups. They were skeptical about the link between false claims and imminent physical harm—especially in cases where the harm was less direct (e.g., reducing people’s willingness to social distance). Moreover, FoE advocates worried that an overly aggressive policy might encourage governments to seek removal of a much broader spectrum of content than necessary to avoid imminent harm. Additionally, they argued that people need space to debate and discuss the virus and associated treatments, to find their own way toward an accurate understanding of the evolving science.

Instead of removal, many FoE advocates suggested that Meta reduce the distribution of COVID-19 misinformation and provide users with authoritative information to counteract false claims. Many health experts also favored surfacing authoritative information, though they believed that for certain claims, doing so should occur alongside removals—not in lieu of them. Overall, during engagements throughout the pandemic, many experts emphasized that promoting authoritative information would help address the ambiguity inherent in crisis situations. Because individuals respond to such uncertainty by seeking information, experts argued, mitigating the spread of false claims about COVID-19 would require providing credible answers to people’s unresolved questions about the pandemic.

On the other hand, experts in science communication warned about the risks of constantly redirecting users to authoritative sources. They explained that users can become frustrated—and perceive that their concerns are not being taken seriously—if they are continually referred to external websites where information is difficult to find and the language is opaque. Others warned that relying on public health institutions, such as the WHO or CDC, to provide authoritative information could make those organizations vulnerable to additional criticism and misinformation. Science communication experts also recommended conveying uncertainty about the virus using Meta’s existing misinformation labels, to mitigate the risk of distrust in the future if scientific understanding changed. Nevertheless, these experts recognized that communicating uncertainty via a short label on a post would be very challenging. On the whole, subject matter experts varied in their recommendations for the best way to address COVID-19 misinformation.

Throughout the pandemic, there was also significant regional variation in stakeholder

feedback on this policy development effort. For example, many stakeholders in the Asia-Pacific region and Sub-Saharan Africa called for the removal of misinformation about the COVID-19 vaccine. These experts argued that such false claims would negatively impact public confidence in vaccines, which was already plummeting in their regions, and further stall vaccination campaigns. However, many public health experts in the US and Europe urged Meta to avoid removing misinformation about the COVID-19 vaccine, expressing concern that removals could fuel conspiracy theories about the vaccine and its origins. Instead, they recommended that Meta promote high quality information about the vaccine, reduce the distribution of false claims, and take stronger measures against anti-vaccine Groups. There was even regional variation in feedback from freedom of expression advocates. While advocates in Europe were worried about an overly broad interpretation of “imminent harm,” advocates from other regions expressed less concern—arguing that the pandemic made the harm associated with false health claims more direct.

How did Meta balance those clear differences in stakeholder perspectives? As expressed in its statement of values, the company prioritizes user voice but will restrict speech if it threatens the safety or privacy of users. In the case of COVID-19 misinformation, Meta tried to balance these considerations in the following way.

First, the company determined that the risks associated with certain misinformation about COVID-19 were significant enough at the time, given the ongoing public health emergency, to warrant removal under Meta’s existing policy banning misinformation that could contribute to imminent physical harm. As of June 2022, Meta removes 80 distinct claims regarding the virus and associated vaccines.⁴⁴ However, Meta does not determine which claims to remove itself—but rather relies on public health experts, including the WHO and CDC, to assess the falsity of COVID-19-related claims and determine which claims could contribute to imminent physical harm. The decision to remove harmful COVID-19 misinformation was informed by feedback from a wide range of stakeholders, who told the Content Policy Stakeholder Engagement team, throughout the pandemic, that certain types of health misinformation, including false cures and false information about preventing or treating the virus, could encourage individuals not to follow critical health advice in the context of public health emergency (e.g., COVID-19 social distancing guidelines). Because COVID-19 is such a highly transmissible disease, experts explained, individuals who chose not to follow health protocols not only put themselves at risk of contracting the disease but could also endanger the health of the community.

To understand the effect of this policy change in practice, consider the example of a piece of content claiming that “COVID-19 can be transmitted by 5G technology.” Under the policies described above, Meta removes health misinformation that could contribute to imminent physical harm in the context of the COVID-19 pandemic, including this 5G claim, on a global scale.⁴⁵ Prior to the pandemic, Meta did not remove entire categories of misinformation on a worldwide scale; instead, the company removed misinformation when local partners with relevant expertise told us a particular piece of content (such as a specific post on Facebook) could contribute to a risk of imminent physical harm.⁴⁶ The COVID-19 misinformation policies, which were informed by stakeholder engagement, therefore changed the way Meta addresses health-related misinformation, like the 5G claim, in the context of a pandemic.

Second, Meta created a COVID-19 Information Center, which provides users in 189 countries with a central place to get the latest news and information on the pandemic,

44. Facebook Help Center,

45. Ibid.

46. Clegg 2022.

as well as resources and tips to stay healthy.⁴⁷ This decision reflected the advice of both medical experts and freedom of expression advocates at the time, who consistently emphasized the importance of uplifting authoritative information about the virus. Through this effort, and related educational pop-ups on Facebook and Instagram, Meta has provided authoritative information about COVID-19 and associated treatments to over two billion people.

3.1.3 Informing Stakeholders about the Policy Decision

Once the company makes a policy decision, the Content Policy Stakeholder Engagement team typically follows up with stakeholders to let them know what was decided and how their feedback was considered. Given the feedback that certain stakeholders provided, the team knew that many free speech experts—and even public health experts in the US and Europe—might be disappointed that the company had decided to remove COVID-19 misinformation during the pandemic, while other experts would likely welcome the decision. Stakeholders have previously told the team, however, that they do not expect Meta to adopt all their recommendations—though they do expect to be told what motivated Meta’s ultimate decision. Thus, in any follow-ups, the team tries to communicate how the company balanced differing stakeholder perspectives on a topic and/or why Meta did not take a particular approach. As Marius Dragomir explains, “It is important that we have clear expectations and open dialogue...But it is stupid to expect that you (Meta) will implement everything we suggest.”⁴⁸ He adds that “he never felt there was a lack of communication.”⁴⁹

Nevertheless, stakeholders clearly desire additional transparency about the impact of their suggestions on policy decisions. Jacob Mchangama says, “It is difficult to measure the impact [of our engagements with Meta], mostly because we don’t have visibility to your discussions internally. I think one way to do it...is to formally tell us how you came to the conclusions...Otherwise, it’s endless talk and you don’t know if anything came out of it.”⁵⁰ Marius Dragomir suggests the team share publicly how Meta arrives at specific policy decisions. “You share the process, but you are not sharing what models (research/frameworks) you are looking at and how you are arriving at the conclusions you do.”⁵¹

The follow-up is not typically the end of the conversation, however. The team encourages stakeholders to provide ongoing feedback as research evolves or local context changes. Researchers will often reach out after publishing new findings related to a policy issue, and those individuals who participate in the team’s recurring expert circles provide feedback on similar policies on a regular basis. When Meta revisits policies, the team reaches back out to the stakeholders who provided feedback the first time around. Moreover, revisions to any of the Community Standards are tracked by a change log on Meta’s Transparency Center, so that external stakeholders can clearly see how policies have been updated.

4 Conclusion

This paper outlines the principles that guide Meta’s approach to stakeholder engagement. The paper also describes what stakeholder engagement looks like in practice, us-

47. *Coronavirus (COVID-19) Information Center*,

48. *Sarah Shirazyan Interview with Marius Dragomir 2022*.

49. *Ibid*.

50. *Sarah Shirazyan Interview with Jacob Mchangama 2022*.

51. *Sarah Shirazyan Interview with Marius Dragomir 2022*.

ing the development of Meta's COVID-19 misinformation policies during the pandemic as an example. However, the piece is not intended to be a final statement of Meta's approach to misinformation in general or COVID-19 misinformation specifically.

First, the “remove, reduce, and inform” approach comes with challenges, which Meta's policy teams continue to consider as they refine the company's policies. According to some experts, including Marius Dragomir, removing misinformation can “backfire and create more distrust in institutions.”⁵² Other experts emphasize that removing misinformation on one platform can simply push it to unmoderated spaces, which doesn't necessarily solve the problem. The scholarly debate on misinformation has also started to shift, as academics place more emphasis on “inoculating users” against misinformation—and less on removing it altogether.⁵³ Reducing the distribution of false claims raises questions about the balance between freedom of speech and freedom of reach. Giving users authoritative information requires the company to identify “authoritative, local sources to uplift,” which can be especially challenging in places that lack an independent civil society or free press—or where government institutions lack capacity. Stakeholders have also warned that the “reduce and inform” treatments can sometimes fuel conspiracy theories, because such treatments may be viewed as evidence that Meta is conspiring to reduce the spread of important information.

Second, on July 26, 2022, Meta formally asked the Oversight Board for a Policy Advisory Opinion on the company's approach to COVID-19 misinformation. As many, though not all, countries around the world seek to return to normal life, Meta is requesting the Oversight Board to advise on whether the policies introduced at the beginning of this global crisis will serve as the right approach in the months and years ahead.⁵⁴ More specifically, the Board's guidance will address whether Meta should continue removing harmful COVID-19 misinformation—or instead rely on reducing its distribution and labeling it.

This piece therefore sheds light on the role of stakeholder engagement in developing Meta's misinformation policies up to this point in time. The goal of sharing this experience is to contribute to the growing academic and civil society conversation surrounding stakeholder engagement as a practice, to bring transparency to this important element of policy development at Meta, and to get feedback from the Trust & Safety community on the approach of the Content Policy Stakeholder Engagement team at Meta.

52. Sarah Shirazyan *Interview with Marius Dragomir* 2022. See also *The online information environment* 2022.

53. See Traberg, Roozenbeek, and Linden 2022 and Lewandowsky and Van Der Linden 2021.

54. See Clegg 2022. The full request is available at “Oversight Board announces new cases and review of Meta's COVID-19 misinformation policies,” July 2022, <https://www.oversightboard.com/news/385467560358270-oversight-board-announces-new-cases-and-review-of-meta-s-covid-19-misinformation-policies/>.

References

- A/HRC/47/25: *Disinformation and freedom of opinion and expression - Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 2021. United Nations High Commissioner on Human Rights, April. <https://www.ohchr.org/en/documents/thematic-reports/ahrc4725-disinformation-and-freedom-opinion-and-expression-report>.
- African Charter on Human and Peoples' Rights ("Banjul Charter")*. 1981. Organization of African Unity, June.
- Allport, Gordon W, and Leo Postman. 1947. *The psychology of rumor*. Henry Holt.
- American Convention on Human Rights*. 1969. Organization of American States, November.
- ASEAN Human Rights Declaration*. 2012. Association of Southeast Asian Nations.
- Brenan, Megan. 2021. *Americans' Trust in Media Dips to Second Lowest on Record*. Section: Politics, October. Accessed August 25, 2022. <https://news.gallup.com/poll/355526/americans-trust-media-dips-second-lowest-record.aspx>.
- Brinkmann, Svend. 2014. "Unstructured and semi-structured interviewing." *The Oxford handbook of qualitative research*, 277–99.
- CDC. 2022. *CDC Museum COVID-19 Timeline*, August. Accessed August 25, 2022. <https://www.cdc.gov/museum/timeline/covid19.html>.
- Clegg, Nick. 2020. "Combating COVID-19 Misinformation Across Our Apps." *Meta* (March). Accessed August 25, 2022. <https://about.fb.com/news/2020/03/combating-covid-19-misinformation/>.
- . 2022. "Meta Asks Oversight Board to Advise on COVID-19 Misinformation Policies." *Meta* (July). Accessed August 25, 2022. <https://about.fb.com/news/2022/07/oversight-board-advise-covid-19-misinformation-measures/>.
- Coronavirus (COVID-19) Information Center*. Meta. Accessed August 25, 2022. https://www.facebook.com/coronavirus_info.
- Covello, Vincent T, Detlof von Winterfeldt, and Paul Slovic. 1987. "Communicating scientific information about health and environmental risks: Problems and opportunities from a social and behavioral perspective." In *Uncertainty in risk assessment, risk management, and decision making*, 221–39. Springer.
- Dvoskin, Brenda. n.d. "Representation without Elections: Civil Society Participation as a Remedy for the Democratic Deficits of Online Speech Governance." *Villanova Law Review*.
- Edelman. 2022. *2022 Edelman Trust Barometer*. Accessed August 25, 2022. <https://www.edelman.com/trust/2022-trust-barometer>.
- Erickson, Bonnie H, TA Nosanchuk, Liviana Mostacci, and Christina Ford Dalrymple. 1978. "The flow of crisis information as a probe of work relations." *Canadian Journal of Sociology/Cahiers canadiens de sociologie*, 71–87.
- European Convention for the Protection of Human Rights and Fundamental Freedoms*. 1950. Council of Europe, November.
- Facebook Help Center. *COVID-19 and Vaccine Policy Updates & Protections*. Accessed August 25, 2022. <https://www.facebook.com/help/230764881494641>.

- Facebook Transparency Center. 2022a. "How stakeholder engagement helps us develop the Facebook Community Standards." *Meta* (January 19, 2022). Accessed August 25, 2022. <https://transparency.fb.com/policies/improving/stakeholders-help-us-develop-community-standards/>.
- . 2022b. "Our approach to misinformation." *Meta* (February 28, 2022). Accessed August 25, 2022. <https://transparency.fb.com/features/approach-to-misinformation>.
- . 2022c. "Types of content we demote." *Meta* (June 17, 2022). Accessed August 25, 2022. <https://transparency.fb.com/features/approach-to-ranking/types-of-content-we-demote/>.
- . n.d. "What are recommendations on Facebook?" *Meta*, accessed August 25, 2022. <https://www.facebook.com/help/1257205004624246/>.
- General comment no. 34 on freedom of opinion and expression, Article 19*. United Nations High Commissioner for Human Rights.
- Gleicher, Nathaniel. 2020. *Labeling State-Controlled Media On Facebook*, June. Accessed August 25, 2022. <https://about.fb.com/news/2020/06/labeling-state-controlled-media/>.
- Gorwa, Robert. 2022. *Stakeholders*. Yale Information Society Project, August.
- Hadid, Daa. 2021. "Pakistan's Polio Playbook Has Lessons For Its COVID-19 Vaccine Rollout." *NPR* (February). Accessed August 25, 2022. <https://www.npr.org/sections/goatsandsoda/2021/02/24/968730972/pakistans-polio-playbook-has-lessons-for-its-covid-19-vaccine-rollout>.
- Handyside v. The United Kingdom*. 1976. European Court of Human Rights, December. <https://hudoc.echr.coe.int/app/conversion/pdf/?library=ECHR&id=001-57499&filename=001->
- Hensler, Deborah R. 2011. "Designing empirical legal research: A primer for lawyers."
- Holmes, Bev J., Natalie Henrich, Sara Hancock, and Valia Lestou. 2009. "Communicating with the public during health crises: experts' experiences and opinions." *Journal of risk research* 12 (6): 793–807.
- International Convention on the Elimination of All Forms of Racial Discrimination*. 1965. Office of the United Nations High Commissioner for Human Rights.
- International Covenant on Civil and Political Rights*. 1966. United Nations General Assembly, December.
- Jang, Kyungeun, and Young Min Baek. 2019. "When information from public health officials is untrustworthy: The use of online news, interpersonal networks, and social media during the MERS outbreak in South Korea." *Health communication* 34 (9): 991–98.
- Jin, Kang-Xing. 2020. "Keeping People Safe and Informed About the Coronavirus." *Meta* (December). Accessed August 25, 2022. <https://about.fb.com/news/2020/12/coronavirus/>.
- Joint declaration on freedom of expression and "fake news", disinformation and propaganda*. 2017. Organization for Security and Cooperation in Europe, March. <https://www.osce.org/fom/302796>.
- Leech, Beth L. 2002. "Asking questions: Techniques for semistructured interviews." *PS: Political Science & Politics* 35 (4): 665–68.

- Lewandowsky, Stephan, and Sander Van Der Linden. 2021. "Countering misinformation and fake news through inoculation and prebunking." *European Review of Social Psychology* 32 (2): 348–84.
- Mileti, Dennis S, and JoAnne DeRouen Darlington. 1997. "The role of searching in shaping reactions to earthquake risk information." *Social Problems* 44 (1): 89–103.
- Nacos, Brigitte L, Yaeli Bloch-Elkon, and Robert Y Shapiro. 2007. "Post-9/11 terrorism threats, news coverage, and public perceptions in the United States." *International Journal of Conflict and Violence (IJCV)* 1 (2): 105–26.
- Newton, Casey. 2018. "Facebook says it will begin removing misinformation that leads to violence." *The Verge* (July). Accessed August 25, 2022. <https://www.theverge.com/2018/7/18/17588624/facebook-misinformation-removal-violence-hate-speech-sri-lanka>.
- Oh, Onook, Manish Agrawal, and H Raghav Rao. 2013. "Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises." *MIS quarterly*, 407–26.
- Perry, Jonathan. 2021. "Trust in public institutions: Trends and implications for economic security | DISD." *United Nations, Department of Economic and Social Affairs* (July 20, 2021). Accessed August 25, 2022. <https://www.un.org/development/desa/dspd/2021/07/trust-public-institutions/>.
- Rabat Plan of Action*. 2013. A/HRC/22/17/Add.4. Annual Report of the United Nations High Commissioner for Human Rights, January.
- Rosen, Guy. 2020. "An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19." *Meta* (April). Accessed August 25, 2022. <https://about.fb.com/news/2020/04/covid-19-misinfo-update/>.
- Sarah Shirazyan Interview with Jacob Mchangama*. 2022, July 15, 2022.
- Sarah Shirazyan Interview with Marius Dragomir*. 2022, July 15, 2022.
- Servick, Kelly. 2020. "Would everyone wearing face masks help us slow the pandemic? | Science | AAAS." *Science* (March). Accessed August 25, 2022. <https://www.science.org/content/article/would-everyone-wearing-face-masks-help-us-slow-pandemic>.
- Shirazyan, Sarah, Allen S. Weiner, Yvonne Lee, Madeline Magnuson, and et. al et. 2020. *How to Reconcile International Human Rights Law and Criminalization of Online Speech: Violent Extremism, Misinformation, Defamation, and Cyberharassment*. Stanford Law School Law and Policy Lab, September. Accessed August 25, 2022. <https://law.stanford.edu/publications/how-to-reconcile-international-human-rights-law-and-criminalization-of-online-speech-violent-extremism-misinformation-defamation-and-cyberharassment/>.
- Starbird, Kate, Emma S. Spiro, and Kolina Koltai. 2020. "Misinformation, Crisis, and Public Health—Reviewing the Literature – Version V1.0." *Social Science Research Council - MediaWell*, Live Research Review (June). Accessed August 25, 2022. <https://mediawell.ssrc.org/literature-reviews/misinformation-crisis-and-public-health/versions/v1-0/>.
- Stofer, Kathryn A. 2019. "Preparing for One-on-One Qualitative Interviews: Designing and Conducting the Interview: AEC675/WC338, 8/2019." *EDIS* 2019, no. 4 (August): 4. <https://doi.org/10.32473/edis-wc338-2019>.

“Polio Eradication: the CIA and their unintended victims.” 2014. *The Lancet* 383 (May). <https://www.thelancet.com/action/showPdf?pii=S0140-6736%2814%2960900-4>.

The online information environment. 2022. The Royal Society, January. <https://royalsociety.org/topics-policy/projects/online-information-environment/>.

Traberg, Cecilie S, Jon Roozenbeek, and Sander van der Linden. 2022. “Psychological inoculation against misinformation: Current evidence and future directions.” *The ANNALS of the American Academy of Political and Social Science* 700 (1): 136–51.

Tuman, Joseph S. 2009. *Communicating terror: The rhetorical dimensions of terrorism*. Sage Publications.

Universal declaration of human rights. 1948. UN General Assembly.

Wright, J Ward. 1987. “The bureaucratic dimension to risk analysis: The ultimate uncertainty.” In *Uncertainty in Risk Assessment, Risk Management, and Decision Making*, 135–43. Springer.

Zittrain, Jonathan. 2022. “How to Fix Twitter and Facebook.” Section: Ideas, *The Atlantic* (June). Accessed August 25, 2022. <https://www.theatlantic.com/ideas/archive/2022/06/elon-musk-twitter-takeover-mark-zuckerberg/661219/>.

Authors

Peter Stern is Director, Content Policy Stakeholder Engagement at Meta. He holds a B.A. from Amherst College and an M.A. (History) and J.D. from the University of California, Berkeley.

Sarah Shirazyan shirazyan[at]fb.com, is Content Policy Manager | Stakeholder Engagement Lead for Misinformation Policy at Meta. She has a J.S.D. from Stanford Law School, where she is also a Lecturer in Law.

Abby Fanlo is Content Policy Manager, Stakeholder Engagement at Meta. She is a PhD Candidate in Political Science at Stanford University (expected January 2023). She also holds an M.A. and a B.A. in Political Science from Stanford.

Acknowledgements

We thank Jacob Mchangama and Marius Dragomir for speaking with us about the practice of stakeholder engagement and allowing us to share their feedback in this piece. We also thank our colleagues on the Content Policy Stakeholder Engagement team, Yvonne Lee and Nicola Aitken, for their support throughout the process of writing this paper. Finally, we thank all of our colleagues across multiple teams at Meta who reviewed this piece and offered thoughtful feedback.