FACEBOOK, INC.
COMMUNITY STANDARDS ENFORCEMENT REPORT, Q1 2021
May 19, 2021
11:30 a.m. ET

Operator:	Hello, and welcome to today's Press Call. There will be prepared remarks and a Q&A to follow. To ask a question after the prepared remarks conclude, please press the "1", followed by the "4". As a reminder, this conference is being recorded today, Wednesday, May 19, 2021.
	Now, I'd like to turn the call over to Sabrina, who will kick this off. Thank you.
Sabrina Siddiqui:	Hi everyone, thank you for joining us. You should have received embargoed materials, including our data snapshots ahead of this call. We are on the record and this call is embargoed until 10 a.m. Pacific.
	Today, you will hear opening remarks from Vice President of Integrity, Guy Rosen; Vice President of Content Policy, Monika Bickert; and CTO, Mike Schroepfer.
	We will then open up the call for questions. In addition to releasing our Community Standards Enforcement Report this quarter, we are also sharing additional integrity and transparency items. This includes our biannual transparency report, intellectual property takedowns report, and new updates on data scraping.
	With that, I will kick it over to Guy.
Guy Rosen:	Thank you, Sabrina, and thanks for joining, everyone, today. We are publishing our Community Standards Enforcement Report for the first quarter of 2021. This report includes metrics on how we enforce our policies from January through March of this year across 12 policies on Facebook and 10 on Instagram.
	We view this report as a quarterly touchstone, and a way for us to give updates on our enforcement results, on policy changes, on technology updates. Now as part of our ongoing transparency efforts, today, we are also launching a new

Transparency Center website on transparency.fb.com as a single destination for all of our various integrity and transparency efforts.

This Transparency Center will explain how we remove harmful content, how we reduce the spread of problematic content, and how we inform people with additional context. The quarterly enforcement report that we're discussing today will also live there going forward, and we're making it easier to view these enforcement reports, to view trends, and to connect the metrics to their specific policy areas.

This website is a starting point and we plan to continue to add more information and build it out in the months and the years to come.

Now back to the numbers in this report. First of all, we believe we shouldn't be checking our own homework, and last year, we committed to undergoing a third party, independent audit to ensure that the metrics we published here are accurate.

We selected EY to undertake this engagement, and we look forward to working with them on this audit. In this quarter's report, we're adding prevalence metrics on Instagram for two areas, adult nudity and violent and graphic content. Prevalence is the percentage of views on our platform that are of violating content.

It's the main metric that we use to evaluate effectiveness of our enforcement, and it matters because it captures not what we took down, but what we missed and was ultimately seen by people. In the first quarter of 2021, prevalence of adult nudity on both Facebook and Instagram was between .03 to .04 percent. Violent and graphic content decreased to between .03 to .04 percent on Facebook; and on Instagram, it was between .01 and .02 percent.

We're committed to sharing more data and matching Instagram metrics with what we share on Facebook. And that's why we're releasing these additional prevalence numbers for the first time here. And we'll continue to share them every quarter going forward as well.

Now hate speech prevalence on Facebook continues to decrease, as a downward trend three quarters in a row. When we first reported this number in Q3 of last year, it was between .10 to .11 percent. In Q4, it fell to between .07 to .08 percent and in this most recent quarter, Q1, it is now between .05 to .06 percent. This improvement and prevalence on Facebook is due to changes we continue to make to reduce problematic content in newsfeed.

On another area, the amount of content we actioned under our regulated goods drugs policy was lower in Q1 as we made some adjustments to our automation

to ensure that we're upholding accuracy. This caused a temporary drop and we've since improved our accuracy and ramped automation back up.

And COVID-19 continues to be a major public health issue around the world and we're committed to helping communities access authoritative information including about vaccination. Since the beginning of the pandemic, we've removed 18 million pieces of content for violating our COVID-19 misinformation policies across Facebook and Instagram.

And we know from public health research that people are more likely to get vaccinated if they see others in their community that are doing so. In countries where vaccines are available to most of the population, places like the U.S. and the U.K., we ramped up our efforts to show people when their friends and neighbors share their support for vaccines through profile frames and stickers. More than 5 million people globally have used these profile frames and more than 50 percent of people in the U.S. on Facebook have already seen someone use the vaccine profile frames, which we developed in collaboration with the Department of Health and Human Services and the CDC.

On Instagram, we also introduced stickers for people to share that they got vaccinated or share their support for vaccines. Since April these have been used more than 7 million times and more than 25 percent of accounts on Instagram in the U.S. have seen someone use a COVID-19 vaccine sticker.

Now finally in India, COVID-19 is affecting nearly every citizen and that includes our content reviewers as well. In the interest of their safety, we have temporarily suspended in-office work and shifted content review to our global network as necessary.

Now with that, I'd like to turn it over to Monika.

Monika Bickert: Thanks, Guy. And hi, everyone. I lead the team that writes our policies on what content is allowed on our services. We continually review these policies to keep up with shifting dynamics around the world and we work regularly with hundreds of organizations and experts outside the company to do this.

It will come as no surprise that the months reflected in this CSER were incredibly busy for my team. We managed the policies for the elections in the U.S. and Myanmar. We released a corporate human rights policy, we began responding to the Oversight Board's recommendations, and we continued to address COVID related content while the virus surged in some parts of the world and the vaccine rollout began in others.

On COVID, Guy touched on some of our efforts but I want to share other steps that we're taking to support the public health community's response to

the pandemic. Since April of 2020, we've been collaborating with Carnegie Mellon University and the University of Maryland on a global survey to gather insights about COVID-19 symptoms, testing, vaccination rates and more.

This is the largest survey of its kind, with over 60 million total responses and more than 170,000 responses daily across more than 200 countries and territories. This effort generates localized insights for researchers, public health officials and policymakers who are working to end the pandemic.

Adults who are hesitant about getting a COVID-19 vaccine indicate that they are most hesitant due to concerns about side effects. And most respondents who are hesitant due to concerns about side effects indicate that they would be most persuaded by those within their community, either health care providers they know or friends and family.

We're continuing to share these insights with public officials, non-profits and others to help inform more effective vaccination efforts.

We continue to remove harmful misinformation about the virus and prohibit ads that try to exploit the situation for financial gain. To debunk claims that are false but would not – if believed true – create a safety risk, we continue to work with over 80 fact-checking organizations that review and rate content in more than 60 languages around the world.

We've also introduced notification screens to give people more context about COVID-19-related links or photos that they're preparing to share. Our goal here is to help people understand what they plan to share before they actually share it.

Moving to the Oversight Board, as I mentioned at the top of my remarks, this past quarter my team has begun to respond to the Oversight Board's recommendations. The board has selected 13 cases since they began their operations last year and they've issued about 40 recommendations.

We've responded to each of those recommendations within 30 days and we've committed to action on the majority of them. Beyond the case content impacted by the board's decisions, which is binding, it's these recommendations that have lasting impact on our content moderation across the platform.

Also, as I'm sure you know, the board issued its decision earlier this month regarding Facebook's initial decision to suspend President Trump from our platforms. Now, the board upheld our decision but they did not specify the appropriate duration of the penalty.

So as well as determining a clear and proportionate response on former President Trump's suspension, Nick Clegg and my team are working through the details and the implications of the board's many recommendations. This is important work, we want to get it right and Nick will set out a response and detail – in detail in due course.

We look forward to continuing to receive the board's decisions in the years to come and believe that their work is helping us to improve our service.

On regulation, we believe the Community Standards Enforcement Report and other transparency reports like ours should be the norm and we think they can help inform ongoing debates about regulation.

We support an approach where the rules of the Internet are updated with reforms to federal law that protect freedom of expression while still allowing platforms like ours to remove content that threatens people's safety.

Specifically, we support updating Section 230 to, first of all, require all platforms to be more transparent about their standards, processes and actions; second, establish periodic reporting requirements; and third, establish a safe harbor approach with exceptions for smaller companies, where larger platforms are required to meet industry best practices for addressing illegal content.

We hope with the new launch of our transparency center we keep making progress where we can and we can also help inform the ongoing regulation conversation.

And with all of that, I'll turn it over to you, Schroep.

Mike Schroepfer: Thanks, Monika. As you've heard today, defining our community standards and how to enforce them is a deeply human challenge. And once those human decisions have been made, our A.I. technology enables them to be implemented at a scale for billions of users.

Today, I want to make – to take a minute talking about one of the A.I. systems we put into production last year and how it's helping our human review teams make a much larger impact.

Thanks to a string of technological breakthroughs made by our researchers in recent years, this new system does something that no automated system could do before. It can look at the entirety of a piece of content, including images, video, text, and comments in multiple languages and evaluate it against multiple types of policy violations.

It sees the bigger picture, which is something that A.I. has been much better at across the board in the last few years.

In the past our A.I. system for enforcing our policies could each only do one narrow thing. One could look at a photo and tell you if it contained nudity, for example. Another could scan the text of a post for violent language. Overtime, we've built thousands of these narrow systems and have it run them on every post, stitching together the output. As you can imagine, that approach has real limitations.

This new system can evaluate the content holistically and consider whether it complies with multiple policies. It's helping us catch more violations and make fewer mistakes. It also does a much better job of identifying priority cases needed to be sent for human review.

In 2020, we tested this new system to see how it might help prioritize which pieces of content you need a pair of human eyes evaluating them. These tests led to significant improvements and the effectiveness of our prioritization system. Meaning our teams of our human reviewers were able to do more valuable work on daily basis. This new system is also capable of going beyond individual posts because it can view content holistically. It can consider the entire pages or groups over time and evaluate their overall health instead of being limited to checking whether an individual comment or image breaks our rules, it can view them as part of a larger body of content that may be trending in a negative direction.

What's interesting is that this evolution of A.I. isn't just happening in content moderation. In so many different spaces, we're seeing the same pattern of systems developing a deeper picture of the information they're working with.

For example, our Computer Vision tools have gotten better at seeing the full content of an image or video and our language translation systems are making leaps in their ability to comprehend multiple languages at once. And most importantly, this is not slowing down.

Research breakthroughs made over the last year and even in the last few months suggested extraordinary imperative improvement in A.I. is still ahead of us. The numbers we've reported in today's Community Standards Enforcement Report are a reflection of progress made in recent years. And the good news is that progress shows no sign of stopping anytime soon.

And with that, I'll turn it over to the operator for questions.

Operator: Thank you. We will now open the line for questions. To ask a question press the "1" followed by the "4."

Your first question comes from the line of Shannon Bond with National Public Radio. Please go ahead.

Shannon Bond: Hi, thanks for taking the question. I'm wondering, Guy, can you talk at all about enforcement or actions that you're taking currently on the misinformation and calls for violence that's been reported to be spreading on the platforms related to Gaza and Israel right now. I mean this is obviously an active conflict. I imagine your teams are monitoring it closely. Can you give us any update on how you're approaching the situation?

Monika Bickert: Hi, yes, this is Monika, I can take that one. And I first want to say the situation on the ground is obviously terrible and our hearts go out to everybody who's affected by the horrific ongoing violence. Of course, safety and freedom of expression have always been important in our service – on our services, and when we look at a situation like this, we want to make sure that we are able to monitor real-time what's going on.

So we've actually set up a special operations center that has 24-hour capabilities with native speakers of Arabic and Hebrew so that we can stay on top of trends, make sure that we are identifying content that violates our policies and remove that quickly, and also identify any errors we've made (and faulty) – in removing content that doesn't violate it and make sure that we're restoring that.

Operator: Thank you. Our next question comes from the line of Issie Lapowsky with Protocol. Please go ahead.

Issie Lapowsky: Hi, thanks so much for doing this. Two questions about some data points in here. To what do you attribute the decline in CNCEI enforcement that we're seeing over the last year? And then separately, but similarly, I noticed that organized hate saw more enforcement in Q1 than it looks like ever before; I'm wondering if you can spell out what you attribute that shift to?

Guy Rosen: Hey, thanks for the question on two important areas. So for child nudity and sexual exploitation areas, there were actually a couple of issues. In the fourth quarter of 2020, we made some changes to our media matching systems as part of updates to our systems in response to the European Privacy and Electronic Communications Directive, the ePrivacy Directive, and we later discovered there was a technical issue in the implementation of those changes that did also impact countries outside of the EU as well.

So when we discovered that error we fixed it and we have been retroactively removing content that was – that was previously missed. Now in the first quarter of 2021, we encountered a separated second technical issue that was also related however to compliance with the ePrivacy Directive. We're in process of addressing that, and again, going back to retroactively remove and

           take any action required on content within the confines of what is permitted under the ePrivacy Directive.

           On your second question on hate organizations, these are improvements in our technology that continue to improve how proactive we are and detect more violating content.

Operator: Thank you. Our next question comes from the line of Jamie Bellassandro from La Republica. Please proceed with your question.

Jamie Bellassandro: Hi, thank you. A couple of things. First of all when you talk – when you talk about hate speech, (do you) also conspiracy theory when the conspiracy theory include of course violence in some sense, or do you recognize conspiracy theory in the system or the system is able to, to recognize it?

           And the – and the second question is about the technology this year, or maybe it was last year, I don't remember, but you talk about the new A.I. called (XLMs) was able to understand different language at the same time. I was thinking that the new system is the evolution of that, and how is the way to success in different languages? Thank you.

Monika Bickert: Sure, maybe I'll start and then – and then hand it over to Schrep for the A.I. question.

           So first on conspiracy theories – this is Monika – let me – let me just sort of take a step back and say that there's a couple ways that we tackle this. First, we remove any accounts that are part of a violence-inducing conspiracy network. And so, you've probably seen some of our announcement around QAnon and the actions we've taken there to identify and remove QAnon accounts and those primarily dedicated to spreading QAnon conspiracy theories.

           Then there's also the content level, meaning that we look at specific posts, and if we have a safety organization telling us that something is false and that there is an imminent risk of safety – an imminent risk to safety, then we'll remove that content.

           Now beyond that, if we have content that is false but it does not carry with it a safety risk, that's when we work with our fact checking partners, and there's more than 80 of them around the world, they are all certified by the International Fact Checking Network. If they label something false, then people see that label when they see it, when they go to share it, when they have previously shared it, and we go ahead and also reduce the distribution of that content and point people to accurate information.

So all of that – our approach will, of course, continue to evolve, but all of that helps us address conspiracy theory content. Schroep?

Mike Schroepfer: Yes, thank you for asking the question. I believe it was about one of the technologies, which is what we call XLMR, and this is a state of the art, very large, self-supervised language model. That's a word soup of a lot of technical acronyms, but the thing about this that's important that you're highlighting is a lot of existing A.I. systems are limited by how robust the training data is you can use to train them, and we have generally speaking in the world a lot more examples of things in English than in other languages, and this is true for hate speech and others.

And so, based on that existing state of the art A.I. systems would perform better in English than in other languages where the sort of training the amount of training information we have for them is lower.

The breakthrough in technologies like XLMR is we're training one system that covers multiple languages rather than individual systems for each language, and the biggest advantage of that is it actually learns things from one language and can transfer them to another. So when we have large data sets in English, it will actually improve the performance of this classifier versus a specific classifier in a language, and those improvements generally increase for languages that are less well-served on the internet. So the smaller amount of data you have on that language, the greater these technologies will work.

This is an explicit area of focus for our research lab on how to do large-scale translation and large-scale multilingual models. This was a technology that didn't work maybe three or four years ago. And so, this is very modern, recent stuff and is one of the reasons why we can improve our coverage across the world in enforcing these policies.

Operator: Thank you. As a reminder to register for a question, press the "1" followed by the "4".

Our next question comes from the line of Cristina Criddle with BBC. Please proceed with your question.

Cristina Criddle: Hi. Thank you. And you mentioned a little bit that your algorithms are getting better at protecting things and taking stuff down, but obviously you have your human moderators, too. And I just wondered at what stage do you envision your sort of technical mechanisms taking over that job, if at all, and what areas do you think you'll be able to sort of trust that tech more than others?

Guy Rosen: Hey, this is Guy, thanks for that question. This will always be a combination of people and technology working together, and if you sit back, technology can help us in a few ways. It can help proactively detect things, some of which can be removed automatically, and some of it will be reviewed by our review teams in order to make the final determination.

It can help amplify and automate – or amplify the impact of any review decision, so for example once a decision has been made about a piece of content by a human reviewer, systems can automatically identify identical contents or similar contents and ensure we're actually taking action on those as well and accelerating that impact.

And third, we also use technology to help prioritize what the human reviewers are looking at so that we're – so that we're looking at the things that are getting the most distribution, which means they're most contributing to prevalence, and that are the most severe. And so through those areas, we see this will continue to be a collaboration between people and technology together for years to come.

Operator: Thank you. We now have a question from the line of Shannon Vavra with Cyber Soup. Please proceed with your question.

Shannon Vavra: Hi there. Thank you for taking the call with all of us. I wonder if you could return to some of the conversation about misinformation in the situation in Gaza and Israel. Could you tell us a little bit about this 24-hour operation center you stood up, like how many people are working there, when you stood it up, are these new hires, are they employees redirected from other projects and work assignments, and how long do you have – plan to have it up further, obviously indications from Israeli officials that this may be going on for some time? Thank you.

Monika Bickert: Sure, happy to. So first, this kind of special operation center is something that we've done in the past when we have a situation where there's a safety threat on the ground or there's an election, and we know that the landscape will be changing rapidly.

So what we do, these are existing employees, what we do is we take people, for instance on our content review team, we have people who are native Hebrew speakers, native Arabic speakers on our content policy team. We have people who are focused on things like hate speech, threats of violence, incitement, graphic violence. We have all of those people regularly, but what we do when we have a special operation center is we take those people from across the company who work on these things and we put them all together so that as the situation is evolving, the communication is very fast.

And any trends that are spotted, we're able to make – to assess those trends, figure out how that – what the interplay is between those trends and our content policies, get updated guidance to our reviewers, and make sure that we're implementing that guidance very quickly.

And of course, we work with safety partners around the world. We also work with fact checking organizations. And so I talked in my response to the last question about our general push for misinformation, and our fact checking network, there's more than 80 of them that we work with, they do include fact checking partners in both Palestine and Israel, and in both Arabic and Hebrew languages.

And in addition, we've activated a feature so that the fact checkers can use keyword detection to group related content in one place. And what this does is it just makes it easier for them to find content that they think they should fact check. This is a feature that we've used in the past for situations like COVID-19, the wildfires in California and Australia, elections in Brazil and in the U.S., other events like that. We'll keep this operation center going, and as long as necessary, and we'll be looking to assess the situation on the ground to inform our decisions on that.

Operator: Thank you. Just as a reminder, to register for a question press the "1" followed by the "4", thank you.

Our next question comes from the line of Mike Swiss with MLex, please go ahead.

Mike Swiss: Hi, thanks for taking my question. I'm asking about scraping. Could you provide a little bit more commentary there in terms of whether Facebook is a larger target than other platforms for scraping exploits? And specifically, the phone number enumeration that you went into more detail here. Is that what was happening with the situation that came out in early April with the 533 million phone records that were published online?

Sabrina Siddiqui: Thank you for that question. We actually have Mike Clark, Director of Project Management on the line and he can take that question.

Mike Clark: Thank you for the question, Mike. I think as they talk about in "The Post" today, it is a industry-wide issue. I think even beyond Facebook and other social media platforms, we saw just even in the April time frame similar things at LinkedIn and Clubhouse and refer to that in the Newsroom Post that we did earlier.

It is a challenge and as you asked the question around phone number enumeration, that phone number enumeration is exactly the description of what we were identifying as the source of the data from what – from what

your question was with regards to April. And I hope that answers your question.

Operator: Thank you. Our next question comes from the line of Julia Arciga with Politico. Please proceed with your question.

Julia Arciga: Hello, thanks for taking my question. I've got two quick ones, actually, regarding the 18 million pieces of COVID-19 related misinformation. I was wondering if you could speak to a country breakdown, whether that is going to be available or if you have the information off the top of your head. Kind of seeing where that information is coming from?

And my second question was in regards to the content restriction based on local law. I was wondering whether you knew if those kind of requests from countries like the UK, Turkey, and Brazil or enforcements on current local laws or were they new local laws that those countries put in place recently? Thanks.

Monika Bickert: Let me start with the first one. When it – when it comes to the content that we have removed for violating our COVID misinformation policies, and I just want to give a little bit of background here. We have policies of removing misinformation around COVID-19 or COVID-19 vaccines when it contradicts public health guidance. We looked at sources like WHO and the CDC. And where if people believe that content, it could lead to a safety risk, them getting the virus, them spreading the virus, and so forth.

So, yes, we have removed more than 18 million pieces of COVID misinformation starting from, I think, we started measuring that in March. We don't have a country breakdown for that but we'll continue to look at ways to add transparency in the future of course.

Your second question was about our restrictions on speech from – at the request of governments. I don't have an answer for when they've passed specific laws. I imagine that's something that would probably vary by which specific request it is and which country it is. But I'll tell you what our process is.

Basically, if a country comes to us, they have to submit an official request telling us that content is illegal. The first thing we do is we actually just look and see if the content violates our policies, because sometimes it does and if it does we just remove it. So, if they – if they flag for us content that violates our policies, we treat that just as we would if any person had reported it to us.

If it doesn't that's when we actually do the legal assessments. We look at the law, we look at whether it's consistent with international norms and human rights principles and if the content in question actually falls under that law.

And then if we restrict it, we restrict that content from visibility in that country only and the result of that is what you see in our government request report. I hope that answers the question.

Operator: Thank you. And up next we have a question from the line of Rachel Gilmore with GlobalNews. Please go ahead.

Rachel Gilmore: Hey there. I have a question about some of the COVID-19 misinformation. So, I understand that you removed 18 million pieces, but I was hoping you might be able to contextualize that a little bit.

So, I'm curious about how many reports you generally get versus how many you take down, as well as how many posts go up on Facebook and Instagram on an average? So, I'm trying to get a – trying to get a sense of what portion of those posts are proving to be misinformation or hate speech. Thank you.

Monika Bickert: Thanks for the question. I don't really have – I don't have any numbers to give you on that. I can just tell you that there's a lot that we're doing here. I mean, there's of course what we're removing that's the more than 18 million number of post number. There's the – about 167 million pieces of content that have been rated false by fact-checkers that we work with and that's where misinformation is about COVID, but if it were believed true, it would not contribute or cause any kind of safety risk.

And then, of course there is, and we've talked a little bit about this in past, but we have other policies where we have specific provisions about COVID-19. For instance, if you say that people in a certain group are responsible for deliberately spreading the virus or that they created the virus, that sort of content we would remove under our hate speech policies.

Beyond that, this is something that in addition to making sure that we're removing that sort of content or labeling misinformation that our fact-checkers rate false, we're also doing so much, Guy mentioned a little bit of this, but we're doing so much to proactively get the message out there about accurate information around the virus and also about the vaccines.

So Guy talked a little bit about things like the sticker pack and the profile frames, but there's also directing people through our labels and through other measures to make sure that they have access to accurate information about the vaccines. And this is all part of our goal to help people get one step closer to vaccination.

We think we have, because of our reach, we have – we have the opportunity to really make a difference here. We're partnering with health authorities and those conversations are very frequent, where we are finding out what's the

	best way to get people the information that they need to understand the virus, to understand the vaccines and to make informed, responsible choices.
Operator:	Thank you. Up next, we have a question from the line of from the line of Victor Silva with G1. Please go ahead sir.
Victor Silva:	Hi, thanks for having my question. I would like to ask about actions against violations committed by politicians. In April, Brazilian President Bolsonaro, for example, had a video flagged for violating our policy against misinformation about COVID-19. This wasn't the first time that he violated the rules on Facebook and Instagram. I would like to know how are you dealing with these issues, specifically, related to politicians. Thanks.
Monika Bickert:	Thank you. So our COVID-19 misinformation policies do apply to everybody around the world that includes heads of state and presidents. And we have removed content from leaders around the world for violating those policies and that does include in Brazil.
	I do want to be careful and say that we allow discussions about the impact of policy measures like lockdowns or developments in scientific research. So, what we're focused on removing here, and again, this is not just heads of state but it includes head of state, it is everybody on our service. What we are focused on removing is information that directly contradicts the prevailing health guidance and could also create a health risk.
	So an example there is if you're saying that hydroxychloroquine is a proven cure or if you are saying that certain groups of people are immune from ever getting the virus, that's the sort of content that we have removed from the early days of the pandemic through our partnership with health officials who help us identify the categories of content that we should remove.
Operator:	Thank you. And our next question now will come from the line of Musadiq Bidar with CBS News. Please go ahead sir.
Sabrina Siddiqui:	And this will be our last question.
Musadiq Bidar:	Hey, good morning. Thank you for the opportunity.
Sabrina Siddiqui:	This will be our last question. Sorry, go ahead.
Musadiq Bidar:	Got it. No worries. I have two questions, first, about the special operations center. What date does that get setup and go live and do you have any data to share from that team related to misinformation coming from Palestine and Israel? Like do we have data to see which side is pushing out more false content?

Second, can you talk about the new A.I. system that can look at text and video and history of pages? What area of content do you imagine this new tool will have the most impact on?

Monika Bickert: So maybe, Schroep, maybe I can start with the question and maybe I'll hand it over to you to address the second question.

So on the first question, I don't have a specific date on when we stood up the operation center. I mean I can tell you that from the very beginning of the violence on the ground, that's something that our teams were in touch around the clock on. I mean I mentioned before that we do have native Arabic speakers, native Hebrew speakers on our content review teams. And then also on my team, we have people who are responsible for liaising the safety group, understanding what the issues are on the ground and giving updated guidance to reviewers.

So even before we stood up the special operations center, that work was ongoing. What we did was we've now formalized that into a operation center where all of these people are real-time communicating with one another so that they can more quickly spot and respond to the changing landscape.

Mike Schroepfer: On the technology, the exciting thing about this is how broadly applicable it is. Because the way I would sort of describe the advancements is if you think of five years ago, state of the art might be not much better keyword matching. And a lot of the challenges we deal with today are much more subtle particularly when you're talking about hate speech, you get things like memes where I have a photo and the text and the photo and the text completely change the meaning.

So if you imagine a photo of a skunk and the overlay in the text is you smell great today. You smell great today is a compliment. The skunk completely changes the meaning. You have to understand the entire context there in order to get what it is. That was sort of out of the reach of our systems just a short while ago.

And when you think about misinformation and viral hoaxes, this is an adversarial force that is trying to evade our classifier. So they may take a screenshot or blur it a little bit or change the words around a little bit so it's not exact matches but you get the gist of it. And these systems as they ingest sort of and learn on their own and just to go technical for a second that the real breakthrough here is we have in production systems that are using techniques called self-supervised learning.

Which means instead of being spoon fed data by a trained person, they basically crawl large data sources and learn on their own, which is the way humans do. This is the way you learn language and physics and a whole

bunch of other things. And these systems because they can ingest a whole lot more data can learn a lot more subtle context and they can also put pieces together on these things.

And so these systems are able to get more context and they're able to evaluate across multiple violation types. So you could say maybe this isn't hate speech, but it's bullying because it's mean. And the same system can sort of output that.

So what you kind of see, a sort of undercurrent in these CICL reports over the last nine reports is that we're generally getting better at most things. And that is a combination of immense operational work from the team you're talking to here. We're optimizing and training the human workforce and a series of tools that will get better literally day over day, week over week, month over month and that's the A.I. technology. And as those two things sort of combined together, we're getting more effective at enforcing against this content.

Operator: Thank you. We'll turn it now to Ms. Sabrina. Thank you.

Monika Bickert: And Sabrina, before you take it back, in response to the last caller, I did just check and it looks like it was last week that we formalized the special operations center. So, again, we've had those – we've had all of those personnel working on these issues and in regular contact, but we formalized that into this special operations center last week.

Sabrina Siddiqui: Thanks for that, Monika. And thank you everyone for joining the call. As a reminder, everything on this call and the materials that were shared are embargoed until 10:00 a.m. Pacific. If you have any other questions, please follow-up with me. Thanks and have a great day.

Operator: Thank you. This concludes the Press Call. Thank you for joining. You may now disconnect your line.

END