S. Hrg. 117–681

SOCIAL MEDIA'S IMPACT ON HOMELAND SECURITY

HEARING

BEFORE THE

COMMITTEE ON HOMELAND SECURITY AND GOVERNMENTAL AFFAIRS UNITED STATES SENATE

ONE HUNDRED SEVENTEENTH CONGRESS

SECOND SESSION

SEPTEMBER 14, 2022

Available via the World Wide Web: http://govinfo.gov

Printed for the use of the
Committee on Homeland Security and Governmental Affairs

SOCIAL MEDIA'S IMPACT ON HOMELAND SECURITY

S. Hrg. 117–681

SOCIAL MEDIA'S IMPACT ON HOMELAND SECURITY

HEARING

BEFORE THE

COMMITTEE ON HOMELAND SECURITY AND GOVERNMENTAL AFFAIRS UNITED STATES SENATE

ONE HUNDRED SEVENTEENTH CONGRESS

SECOND SESSION

SEPTEMBER 14, 2022

Available via the World Wide Web: http://www.govinfo.gov

Printed for the use of the
Committee on Homeland Security and Governmental Affairs

U.S. GOVERNMENT PUBLISHING OFFICE
51–542 PDF WASHINGTON : 2023

COMMITTEE ON HOMELAND SECURITY AND GOVERNMENTAL AFFAIRS

GARY C. PETERS, Michigan, *Chairman*

THOMAS R. CARPER, Delaware
MAGGIE HASSAN, New Hampshire
KYRSTEN SINEMA, Arizona
JACKY ROSEN, Nevada
ALEX PADILLA, California
JON OSSOFF, Georgia

ROB PORTMAN, Ohio
RON JOHNSON, Wisconsin
RAND PAUL, Kentucky
JAMES LANKFORD, Oklahoma
MITT ROMNEY, Utah
RICK SCOTT, Florida
JOSH HAWLEY, Missouri

DAVID M. WEINBERG, *Staff Director*
ZACHARY I. SCHRAM, *Chief Counsel*
CHRISTOPHER J. MULKINS, *Director of Homeland Security*
ALAN KAHN, *Chief Investigative Counsel*
MORAN BANAI, *Senior Professional Staff Member*
PAMELA THIESSEN, *Minority Staff Director*
SAM J. MULOPULOS, *Minority Deputy Staff Director*
CLYDE E. HICKS, JR., *Minority Director of Homeland Security*
MARGARET E. FRANKEL, *Minority Professional Staff Member*
LAURA W. KILBRIDE, *Chief Clerk*
ASHLEY A. GONZALEZ, *Hearing Clerk*

CONTENTS

WITNESSES

WEDNESDAY, SEPTEMBER 14, 2022

ALPHABETICAL LIST OF WITNESSES

APPENDIX

SOCIAL MEDIA'S IMPACT ON HOMELAND SECURITY

WEDNESDAY, SEPTEMBER 14, 2022

U.S. SENATE,
COMMITTEE ON HOMELAND SECURITY
AND GOVERNMENTAL AFFAIRS,
Washington, DC.

The Committee met, pursuant to notice, at 10 a.m., in room SD–342, Dirksen Senate Office Building, Hon. Gary Peters, Chairman of the Committee, presiding.

Present: Senators Peters, Hassan, Rosen, Ossoff, Portman, Johnson, Lankford, Romney, Scott, and Hawley.

OPENING STATEMENT OF CHAIRMAN PETERS[1]

Chairman PETERS. The Committee will come to order.

In recent years, domestic terrorism, and specifically white supremacist, conspiracy related, and anti-government violence, has become one of our nation's greatest homeland security threats.

Last October, the Committee held a hearing to examine the role social media platforms play in the amplification of domestic extremist content and how that content can translate into real-world violence. We heard from expert witnesses who discussed how recommendation algorithms, ad targeting, and other amplification tools end up pushing increasingly extreme content to users because that type of content is what keeps people active on the platforms.

Unfortunately, because these platforms are designed to push the most engaging posts to more users, they end up amplifying extremist, dangerous and radicalizing content. This includes QAnon, Stop the Steal, and other conspiracy theories, as well as white supremacist and Anti-Semitic rhetoric.

In some cases, this content may not necessarily violate a company's community guidelines. In other cases, even content that is in clear violation of company policies remains on the platforms, and is often only removed after public pressure. In both cases, this content does significant harm to our society and stokes real-world violence.

We have seen this happen time and time again. From the 2017 neo-Nazi "Unite the Right" rally in Charlottesville, Virginia that was organized using a Facebook event page, to the violent January 6, 2021, attack on the U.S. Capitol spurred to action in part by Stop the Steal content that repeatedly surfaced online, to the shooter who livestreamed as he massacred Black shoppers at a Buffalo

[1] The prepared statement of Senator Peters appears in the Appendix on page 95.

supermarket, there is a clear connection between online content and offline violence.

Over the years, we have heard many explanations from social media companies about their content moderation policies, efforts to boost trust and safety, and actions taken to remove harmful accounts.

There is no question that those efforts are certainly important, but there is a question of whether those actions are enough to effectively address the spread of dangerous content online and the resulting threats it poses to our homeland security.

The central question is not just what content the platforms can take down once it is posted, but how they design their products in a way that boosts this content in the first place, and whether they build those products with safety in mind to effectively address how harmful content spreads.

That is the focus of today's hearing where we will have the opportunity to hear from two panels of witnesses, outside experts, including former Facebook and Twitter executives, as well as current senior executives from Meta, YouTube, TikTok, and Twitter, who are charged with designing social media products used by billions of people around the world.

The overwhelming majority of social media users have very little information about why they see certain recommended content in their feed, and there is very limited transparency into how social media companies balance their business decisions with the need for online safety, including what resources they invest into limiting the spread of harmful content.

Our goal is to better understand how company business models and incentive structures, including revenue generation, growth, and employee compensation, determine how social media products are built and the extent to which current incentives contribute to the amplification of content that threatens homeland security.

For nearly a year, I have been pressing Meta, YouTube, TikTok, and Twitter for more information on their policies to monitor and remove extremist and conspiracy content that advocates violence, as well as the relationship between their recommendation algorithms and targeted advertising tools that generate much of the companies' revenues, and the amplification of extremist content.

The companies' response to those inquiries have been incomplete and insufficient so far.

This morning, we will hear from two former executives and a technology journalist with social media expertise about the internal product development process and the business decisions these companies make, including tradeoffs between revenues and growth and their trust and safety efforts, as well as how they interact with foreign governments.

Later this afternoon we will hear directly from the Chief Product Officers (CPO) of Meta, YouTube, and Twitter and the Chief Operating Officer (COO) of TikTok, the Executives who are charged with making these business decisions and driving the strategic vision of the companies.

I certainly look forward to a productive discussion with both panelists. Welcome to this Committee here today. We look forward to your testimony.

Ranking Member Portman, you are now recognized for your opening comments.

OPENING STATEMENT OF SENATOR PORTMAN[1]

Senator PORTMAN. Thank you, Mr. Chairman, and I thank the experts for being here. We look forward to hearing from you. This is going to be an interesting hearing.

This past Sunday observed the 21st anniversary of the tragic September 11, 2001 (9/11) terrorist attacks, and over these past couple of decades our country has adapted to combat the most pressing threats to our nation's security, and that is good. But the advent of social media has added a new dimension to the ever-evolving threat landscape and created new considerations for combating terrorism, human trafficking, and many other threats.

During last October's hearing on how algorithms promote harmful content I focused on how social media business models contribute to the amplification of terrorism and other dangerous activities. Since then, the Committee has identified ways in which social media companies' product development processes tend to conflict with user safety. Whistleblower testimony has revealed that in numerous occasions the leaders at social media companies were aware that certain platform features increased threats to user safety and chose not to mitigate such concerns. We will hear about that today.

It is unfortunate that the American public must wait for whistleblower disclosures to find out about ways in which platforms are knowingly and unknowingly harming their users. The lack of transparency in the product development process, the obscurity of algorithms, and misleading content moderation statistics create an asymmetric information environment in which the platforms know all, yet the users and policymakers and the public actually know very little.

One consequence of this lack of transparency is related to China. I have serious concerns about the opportunities that the Chinese Communist Party (CCP) has to access TikTok's data on American users. There are now over 100 million Americans, including 40 million under the age of 19 who use TikTok. This TikTok data remains vulnerable to the Communist Party of China, both as the CCP tries to exploit its access to U.S. data and exert influence over the content that U.S. users see.

For example, despite moving U.S. data servers to the United States, TikTok and ByteDance employees in China retain the ability to access this data. If that is not true we would like to hear about that today.

Also we learned yesterday, from Senator Grassley's opening statement in a Senate Judiciary Committee hearing with the Twitter whistleblower that Twitter failed to prevent Americans' data from being accessed by foreign governments. In fact, Senator Grassley spoke about how several Twitter employees were actually foreign agents of India, China, and Saudi Arabia, which is concerning and speaks to why Congress needs more information from platforms on how they secure user data.

---

[1] The prepared statement of Senator Portman appears in the Appendix on page 99.

Another consequence of poor transparency relates to content moderation. While I recognize that content moderation is a key component to creating safe platforms for users, it cannot be the only thing. Transparency reports released by companies often detail the amount of content that has been removed for violating company policy. However, these reports do not account for violating content that is left up on the platform and yet goes undetected.

It also does not account for content that is incorrectly censored, as we often see with many conservative voices on social media. I, like many of my colleagues, have been critical of the political biases held by big tech platforms, which have resulted in systematic takedowns of accounts that hold ideologies with which the left and liberal media disagree.

We will hear about that today, but these takedowns are often done under the guise of combating misinformation which, in fact, they are just combating conservative viewpoints that conflict with their own. Any steps taken to address the impact media on homeland security must account for First Amendment protections, of course, and safeguard free speech.

For us to have a responsible conversation about the impact of harmful content on American users and homeland security we need to talk about how current transparency efforts have worked or not worked. Congress must enact legislation that will require tech companies to share necessary data so that research can be done to evaluate the true extent of how harms from social media impact Americans.

As some of you know, I have been working on legislation along those lines with Senator Coons to establish bipartisan legislation to do just that. The Platform Accountability and Transparency Act (PATA) would require the largest tech platforms to share data with vetted independent researchers and other investigators so that we can all increase our understanding of the inner workings of social media companies and regulate the industry based on good information that we simply do not have now, that we can learn through this process.

Again, I thank the witnesses for being here and I look forward to having your expertise help to guide us in these complicated issues, and thank you, Mr. Chairman, for holding this hearing.

Chairman PETERS. Thank you, Ranking Member Portman.

It is the practice of this Homeland Security and Governmental Affairs Committee (HSGAC) to swear in witnesses, so if each of you would please stand and raise your right hands.

Do you swear that the testimony you will give before this Committee will be the truth, the whole truth, and nothing but the truth, so help you, God?

Mr. ROETTER. I do.

Mr. BOLAND. I do.

Mr. CAIN. I do.

Chairman PETERS. You may be seated.

Today's first witness is Alex Roetter, the former Senior Vice President of Engineering at Twitter. In his previous role, Mr. Roetter helped grow Twitter's monthly active users to over 300 million and build the ad network from near zero revenue to $2.5 billion a year.

Mr. Roetter also spent six years at Google on a variety of projects including building the world's largest computational advertising platform. He was in the room for major decisions about products at Twitter and is familiar with the priorities that were weighed as products were created, as well as how those products are then built.

Mr. Roetter, welcome to the Committee. You may proceed with your opening remarks.

TESTIMONY OF ALEX ROETTER,[1] FORMER SENIOR VICE PRESIDENT FOR ENGINEERING (2014–2016), TWITTER

Mr. ROETTER. Good morning, Mr. Chairman and members of the Committee. Thank you for inviting me here today.

We live in a world where an unprecedented number of people consume information from social networks. Viral content and misinformation can propagate on these platforms on a scale that is unseen in human history. Regulators must understand companies' incentives, culture, and processes to appreciate how unlikely voluntary reform is.

In over 20 years of working in Silicon Valley as an engineer and as an Executive, I have seen firsthand how several of these companies work. Today I will talk about how these companies operate and actionable ways to demand transparency.

The product development lifecycle works as follows. First, teams of product managers, engineers, and designers are assigned specific metrics to maximize. These metrics carefully track user engagement and growth as well as revenue and financial indicators. Other metrics, such as user safety, are either not present or much less important.

Second, teams use an experimental system to launch changes to small percentages of users. The effect of every experiment on key metrics is measured extremely accurately. Absent are detailed metrics tracking impacts on user safety. For example, I never once saw a measurement such as did a given experiment increase or decrease the spread of content later identified as hate speech.

Third, executives review these experimental dashboards regularly and make decisions on which experiments to launch. These reviews are run by product and engineering. Other functions like legal or trust and safety are absent or do not play a substantial role.

Culturally, these companies are informal hierarchies with the "builders," by which I mean engineers, product managers, and designers, held in the highest regard. Other functions are viewed much more skeptically. The strong bias is to make sure that corporate bureaucracy does not slow down product development.

These companies conduct regular performance evaluations and promotions, and these drive peer recognition, career advancement, and cash and stock awards. The main data collect is what impact an individual's work has on key metric families. Only a minority of builders get promoted based on impact to trust and safety metrics, as those impacts are not valued as highly.

[1] The prepared statement of Mr. Roetter appears in the Appendix on page 104.

What data has been shared selectively to date is mostly non-illuminating statistics designed to create the appearance that they are taking the problem seriously. When one of the largest companies in the world says it is spending what seems like a large, absolute number, that number must be put in context and compared to the size of other initiatives, for example, product efforts or how much they spent on stock buybacks. Large investments amounts are not sufficient. We must demand transparency based on measuring actual results.

Similarly, when a company points to how much content it has taken down, that has to be understood in terms of its reach in the network. Removing a billboard in Wyoming is very different than removing a billboard in Times Square.

For real transparency I recommend assembling an independent group of researchers and data scientists. Task them with enumerating the right questions to ask and the set of data they need to answer them. Fund them to continually do this work and refine their questions and data requests.

The government is able to demand transparency in technically demanding fields. For example, third-party auditors of public company financial statements are able to balance the public's need for reliable financial statements with a company's need to keep information confidential.

Until such transparency exists, every assurance by any of these companies has to be taken on faith. Transparency is necessary but not sufficient. Until we change the fact that user attention and profits are what companies care about above all else, all the data-sharing in the world will not address the problem.

Policy and legal experts have previously testified before the Committee on ways that incentives could be changed. Incentives matter. Companies behave differently when they care about the quality of content. For example, having inappropriate ads could materially harm financial performance, so most advertising systems place ad copy removal as a step that has to occur before the new ad ever makes its way to users. On the other hand, user-generated content is allowed to go live instantly.

Incentives also shape companies' recommendation algorithms. For example, TikTok and ByteDance feed young people in China a diet of educational science and math content via their recommendation algorithms. The Chinese version of the app even enforces a daily usage limit. Contrast this to how U.S. companies target content to young Americans, optimizing their engagement of revenue at any cost.

Any suggestion for more useful transparency will be met with many objections. The status quo is simply too lucrative. Do not underestimate these companies' ability to fight requests for information. After all, the legal team at Google alone has the same number of lawyers as all the employees of the Federal Trade Commission (FTC).

Given what we know about companies' incentives, processes, and culture, we should not expect meaningful progress voluntarily, and we should view their commitments extremely skeptically, however, with the proper transparency and regulatory environment I believe

we can change their incentives and start to see real, measurable progress against these problems. Thank you.

Chairman PETERS. Thank you.

Our next witness is Brian Boland, a former Vice President of Partnerships Product Marketing, Partner Engineering, Marketing, Strategic Operations, and Analytics at Facebook. Mr. Boland worked at Facebook for 11 years. He worked in several roles including leading a 500-person multifunction team focused on product strategy, market strategy, partner engineering, operations, analytics, and marketing. These high-impact teams worked across Facebook products and features including watch, video, news, group admins, developers, payments, and audience network. Before joining Facebook, he worked at Microsoft and other tech companies.

Mr. Boland, welcome to the Committee. You may proceed with your opening remarks.

TESTIMONY OF BRIAN BOLAND,[1] FORMER VICE PRESIDENT (2018–2020), PARTNERSHIPS PRODUCT MARKETING, PARTNER ENGINEERING, MARKETING, STRATEGIC OPERATIONS, & ANALYTICS, FACEBOOK

Mr. BOLAND. Good morning, Mr. Chairman, and Members of the Committee. Thank you for holding these hearings that cover such important issues for our nation and the world, and thank you for inviting me here today to provide testimony on my experiences as senior executive at Facebook, now known as Meta.

For the last few years I have grown increasingly concerned about the roles that Facebook, Instagram, YouTube, Twitter, and TikTok play in driving the growth of misinformation, extremism, and generally harmful content. I worked at Facebook for 11 years in a variety of leadership roles, helping to shape product and market strategies for a broad array of products, including advertising news, video media, and more. During my tenure at the company I worked for the most senior executive and was deeply embedded in the product development process.

In my last two years of my time at the company, the CrowdTangle team and product was a part of my organization. CrowdTangle is a tool that provides limited, albeit industry-leading transparency in the public news feed content on Facebook. What finally convinced me that it was time to leave was that despite growing evidence that the news feed may be causing harm globally, the focus on and investments in safety remained small and siloed.

The documents released by Frances Haugen, the Facebook whistleblower who last fall testified here, highlight issues around polarization globally and the power of Facebook to lead people down a path to more extreme beliefs. These papers demonstrate thoughtful, well-researched documentation of the harms that concerned me. The research was done by highly skilled Facebook employees who are experts in their field, and was extensive.

Rather than address the serious issues raised by its own research, Meta leadership chooses growing the company over keeping people safe. While the company has made investments in safety, these investments are small are routinely abandoned if they do not

[1] The prepared statement of Mr. Boland appears in the Appendix on page 110.

impact company growth. My experience at Facebook was that rather than seeking to research and discover issues on the platform before others found them, they would rather reactively work to mitigate the public relations (PR) damage for issues that came to light.

I have come to believe that several circumstances have put Americans at risk from the content on these platforms. The first is the growth over safety incentive structure that leads to products that are designed and built without a primary focus on safety. The next is the unprecedented lack of transparency available from these platforms so that we can analyze content and understand the impact from these tools. Finally, the lack of clear oversight for the business practices of these companies.

We have faced challenges like this before with new technologies. In the 1960s, Congress addressed the dramatic rise in fatalities caused by the rapid increase in automobile use in the United States. That industry experienced explosive growth in the companies focused on growth in sales, and it turns out safety did not sell. The creation of the National Highway Traffic Safety Administration (NHTSA), at the time the National Highway Safety Bureau (NHSB), empowered an agency to study the available data, and in partnership with researchers and other agencies to take steps to make driving in America rapidly and significantly safer. Today, automobile manufacturers portray safety as a selling point such that they welcome verification of these efforts.

The problem with these social media platforms today is that we lack public data to understand the current issues, and there is extremely limited ability to research these platforms and almost no ability to protect our future and creation a version of crash testing the car. Imagine if, in the 1960s, we had no way of knowing the deaths that were happening from cars, and we had no way of knowing it was increasing so rapidly. That lack of data is where we are today with social media platforms.

The reality is that for all the debate about whether social media is predominantly good or bad, the truth is that we do not really know. If anyone tells you they know, they do not know. I believe that we have a right to know. The good news is that with the right incentives in place and rules around transparency we can develop a better understanding of these issues and take steps to mitigate the harms.

If we take these steps we can do now what we did with the automobile. We can empower agencies and researchers to deeply understand the issues, and through changes in incentives, public education, and better development, build a path to a future where we still get the amazing benefits from these products while mitigating the harms that we barely understand today.

Today I hope to shed light on product development process, internal and external incentive structures for these organizations, and the critical importance of transparency. I appreciate your work to better understand these issues and deliver real-world solutions to the American people. Thank you.

Chairman PETERS. Thank you.

Our final witness of our first panel is Geoffrey Cain, Senior Fellow for Critical Emerging Technologies at Lincoln Network. Mr. Cain is an award-winning foreign correspondent and author. His

work has taken him to the world's most authoritarian and remote places, from inside North Korea to the Trans-Siberian Railway across Russia, from investigations into genocide in Cambodia to experiments in technological surveillance in China.

Mr. Cain has served as a tech Congress fellow with the House Foreign Affairs Committee minority and supported a range of issues, including China, tech sanctions, and investigative work.

Mr. Cain, welcome to our Committee. You may proceed with your opening comments.

TESTIMONY OF GEOFFREY CAIN,[1] SENIOR FELLOW FOR CRITICAL EMERGING TECHNOLOGIES, LINCOLN NETWORK

Mr. CAIN. Good morning, Chairman Peters, Ranking Member Portman, and Members of the Committee. It is an honor to be invited to testify here today on social media's impact on national security. Today I will talk about one of the greatest technological threats facing our homeland security and democracy: TikTok, the social media app that reports to a nefarious Chinese company called ByteDance.

As an investigative journalist in China and East Asia for 13 years, I have been detained, harassed, and threatened for my reporting on Chinese technology companies. Today I will show you how TikTok has orchestrated a campaign of distraction and deflection to mask the alarming truth.

Americans face the grave unprecedented threat of software in our pockets that contains powerful surveillance and data-gathering capabilities, owned by private companies that must comply with the dictates of a foreign authoritarian government ruled under the Chinese Communist Party.

The CCP had signaled its ambitions to assert global jurisdiction over private companies everywhere as a condition for doing business in China. TikTok, therefore, is a disaster waiting to happen for our security and the privacy of our citizens.

We will have TikTok executive here today later. According to their internal public relations guidelines, leaked to the media, they are required to, "Downplay the parent company, ByteDance, downplay the China association, and downplay artificial intelligence (AI)."

The public relations guideline states that if you ask them about the influence of the Chinese company, ByteDance, and its influence over its American product, TikTok, which is used by many Generation Z teenagers, its executives must deceptively tell you that ByteDance is a separate company in China and that you should talk to ByteDance instead.

They will attempt to confuse you, claiming that TikTok takes a localized approach, hiring local moderators, implementing local policies, and showing local content. They will not tell you about an individual who is unnamed so far, called the "Master Admin" in Beijing—this has been leaked to the media, to Buzzfeed—who has had access to all Americans' data. They also will not tell you that they, at TikTok, report to ByteDance executives in China, and ByteDance reports to the Chinese Communist Party.

[1] The prepared statement of Mr. Cain appears in the Appendix on page 117.

TikTok's fast expansion into the American market was only possible because China has rigged the market. The Chinese government offered ByteDance vast market protection in China, all while banning competing American social media apps, Facebook, Instagram, Twitter, and Google.

Like all Chinese companies, ByteDance runs an in-house Communist Party Committee that enforces the political loyalty of its employees. In 2018, ByteDance and TikTok's founder and previous Chief Executive Officer (CEO), a man named Zhang Yiming, wrote a public letter promising Chinese regulators that his company would follow "core socialist values," would introduce these "correct values into technology and products" and would ensure his products promoted the Chinese Communist Party's agenda.

These values, he wrote, included "strengthening the work of the party construction," "deepening cooperation with official party media," and strengthening "content review in line with these party values."

ByteDance's public statement in China should be cause for alarm, considering American government employees, military personnel, and workers in sensitive and strategic industries use TikTok.

When TikTok began growing its present in the United States in 2016 and 2017, I was an investigative journalist in China's western region of Xinjiang, where I was writing my second book, The Perfect Police State, which is an investigation into the Chinese surveillance dystopia.

I learned that ByteDance and TikTok were expanding into America, and I knew that this was ominous because I had been speaking to a former worker for the Ministry of State Security, a major intelligence and extremely powerful intelligence body, who had told me that he had worked with numerous companies, including ByteDance, to expose the data of ethnic minorities in China. It was not hard. It simply happened.

ByteDance has also had an active role in suppressing news about the atrocities, which included physical and psychological torture, internment in concentration camps, forced sterilizations, and the wholesale destructions of mosques and other cultural artifacts. So this is very serious.

I am aware of time so I do have much more in the written testimony if you would like to ask. But thank you for your time today, and I look forward to answering your questions.

Chairman PETERS. Thank you, Mr. Cain.

Extremist groups, including QAnon followers, Islamic State of Iraq and Syria (ISIS), and white supremacists certainly have expanded their ranks by recruiting individuals on major social media platforms. The Christchurch shooter, who killed 51 people and inspired the Poway and El Paso shooters was radicalized on YouTube and livestreamed his attacks on Facebook to rally others to his cause.

Three years later, a shooter in Buffalo, New York, streamed his attack on Twitch, which acted quickly to take it down but the video was soon circulating widely on Facebook.

Mr. Roetter, would you tell this Committee why do these platforms' recommendation algorithms spread this extremist content just so rapidly?

Mr. ROETTER. Thank you for the question. The way to understand these recommendations, they do not have intentionality about specific types of content. But the way they work is they assemble a massive amount of information, they model everything about your usage, your interests, your geography, who you are connected to, what you have engaged with historically—and then they model the content and they try to match those optimally.

What makes this so dangerous is there is a positive feedback loop, and if you pick something not controversial, just pick a hobby—knitting, for example—if I think that you are somewhat into knitting I might recommend some knitting content to you because I believe you will engage with it. You do engage with it, and that both makes you more interested in knitting because you are doing the hobby more, but also it feeds back to the algorithm, which then has signaled that you do like more of this content.

The next day, or the next session, it is more confident that you will engage in this content and you will go further down the rabbit hole. Obviously, with knitting that is fine, but this is true for all sorts of content. Because of this feedback loop, if you have some proclivity or some interest in some topic, you will be fed more of that. That generally feeds your interest, and you are fed more and more.

That is why we see people that start off with more things in common than differences sort of splitting and fracturing as they each go into their worlds that are more and more different and have less in common with other people. This is all an inevitable consequence of this optimization of driving engagement.

Chairman PETERS. In terms of content that takes you down the rabbit hole, are companies able to change some of those algorithms to prevent that from occurring, at least with that kind of content, and how would that work?

Mr. ROETTER. You certainly could, in theory. It will never happen, given the current incentive structure. These are for-profit companies. They are incented to maximize profit, and before they have realizable profits they are incented to show massive user growth to convince investors that they will be massively profitable in the future.

The way they do that is getting people to come back to their platform over and over. The way they do that is for optimizing engagement. As long as the algorithms are optimizing for showing you things that you will engage with we will always have this positive feedback loop property. I show you something you are interested in, you get more interested it, you are more likely to keep engaging.

You can build an AI to train for anything, but you pick an incentive based on the overarching incentive of the environment you find yourself in, in this case a public company that is reporting to shareholders. Until those incentives change we should not expect the AI to optimize for anything other than engagement and profit maximization.

Chairman PETERS. We are going to hear later this afternoon from chief product officers at some of these major companies. Is it pos-

sible for them to set different priorities for product development to address the spread of extremist content? Is that within their purview, and is that something they should be able to talk about?

Mr. ROETTER. In practice it is not possible, and the reason is these are just individuals. This is not a matter of a few bad eggs running companies. This is a system that these people find themselves in. They are in a system where they have to report user growth, engagement, increasing attention from the users, and profit.

I should add, this attention game is a race to the bottom. If I build a product that less addictive than a competitor's product, by definition user eyeballs and attention will go over to the competitor. Then I have to, in turn, make my product more addictive to pull people back or I will quickly be abandoned by investors.

Given that structure there is no way that a product leader or any other executive at a company could optimize for anything other than those core metrics, engagement and revenue, because that is the system they find themselves in.

Chairman PETERS. They cannot do it by themselves. It has to be broader than that. But they are at the front end of that, or at the beginning of that, to understand exactly how that incentive structure, how those priorities shape the work that they do? They can talk about that, I suspect.

Mr. ROETTER. They could talk about that. They are doing exactly what you would expect them to do, given the environment they find themselves in. As long as the incentives of those companies are what they are, they will continue to behave the way they are behaving, and if they did not, it would hurt the trajectory of the company.

Chairman PETERS. Very good. Mr. Boland, why are the actions taken by trust and security teams at these platforms just not enough to deal with this problem? We are going to hear a lot about these teams, I think, this afternoon. Why is that not enough?

Mr. BOLAND. Yes, I imagine that you will hear that there are significant investments from the companies in trust and safety, and it is true they make some investments.

The important thing to think about with trust and safety effort is that if it is siloed from the rest of the process, if it is a last-check safeguard or a group that is small and not a core part of the way that people build products, it will always be an afterthought, and it will always be a team that has to fight in the battle of tradeoffs between their ways that they would like to improve trust and safety and impacts to growth. Impacts to growth translate to impacts to revenue.

That dual tension of not receiving enough resources and being at odds with the product development teams and the product development process makes it so that team has to fight for any sort of interventions they want to put in place.

You can change the way that product teams could work with those organizations. A good example would be the efforts that Facebook is now putting in place around privacy. For a number of years we know that Facebook and Meta had not been at the top of its game on privacy. After the last sort of issues with the FTC, the company has invested significantly in efforts around privacy,

and has made that something that the product teams and the product managers actually have to care about.

As long as that team in trust and safety is off to the side, if there are not the incentives in place that say to the company, "You really have to make sure that when people are making day-to-day decisions they are prioritizing these efforts," that team is fighting a losing battle, both in resources, because they are battling against a number of people, and incentives at the company because they have to justify every single change that they want to make. It is not a core part of how the teams think.

Chairman PETERS. At best, there is going to be follow-up. The products are going to get launched. They are going to potentially cause problems that team may have even recognized but they were not able to interject that effectively during the product development phase. Then later they may be engaged but at that point the genie is out of the bottle, so to speak, before they can get engaged. Is that correct?

Mr. BOLAND. That is correct. You can kind of understand this if you think about where these companies started and the incredibly short period of time that they have grown to be as successful as they are. They still feel like startups, in the way that the leaders think, even though these are some of the biggest companies in the world.

At the beginning of their lifecycle it was about trying to figure out products that could grow, and grow effectively in the world, and has gotten to a point where they have not matured out of that stage.

I think we can get to a point where these companies could do more in that space. We just have not seen them make that transition into a more responsible set of activities.

Chairman PETERS. All right. Thank you.

Ranking Member Portman, you are recognized for your questions.

Senator PORTMAN. Thank you, Mr. Chairman. We have a good group of Members here so I will keep within the time, and I want to focus on TikTok first. We talked about TikTok being the most popular social media app in the United States. I also think it poses a risk to our national security, and I want to dig deeper into that today with both of these panels, this one and then when the TikTok representatives are here later.

My understanding is that under Chinese law the Chinese Communist Party can access data of tech companies that are run out of China or have parent companies that are run out of China. Both ByteDance and TikTok have offices and they have employees in Beijing.

Mr. Cain, under Chinese law does TikTok have a legal obligation to give U.S. user data to the Chinese Communist Party?

Mr. CAIN. Oh yes, absolutely. TikTok executives will, under Chinese law, face a minimum of 20 days' detention if they refuse to turn over data on anyone in the world, and this could be anybody in China, anybody who is traveling through China, through Hong Kong. This is a documented legal situation, and it is not something that TikTok, despite claiming to be an American company, can avoid.

I would also like to point out that TikTok does dodge this question frequently by trying to point out that it is run by a Cayman Islands holding company, a shell company essentially. This is a red herring to distract from the issue at hand.

The American company, TikTok and the Chinese company, ByteDance, both report to this Cayman Islands shell company. The company has never said how many people actually work for the shell company, the holding company, but we do know that the CEO of ByteDance and the CEO of TikTok are the same person. This is listed on the Cayman Islands registry. The CEO is the same person running the ByteDance company in China, according to their website.

Senator PORTMAN. Let me delve a little deeper here because we are going to hear from TikTok later, and based on the testimony we have received from them in advance I think they are going to say they have not provided data to the Chinese Communist Party. Even if CCP requested data, they said they would not share it with them. Again, does China need to make a request to access this data, or does China have the capability to access it at will?

Mr. CAIN. I am not aware of the Chinese government having the ability to simply open a computer and access it at will. It would usually happen through somebody in ByteDance or in TikTok. This has already been demonstrated and documented. There was a Buzzfeed report that came out a few months ago which contained 20 leaked audio files from internal meeting at TikTok in which TikTok employees said that they had to go through Chinese employees to understand how American data was being shared. It also pointed out that employees were saying that there is an individual in Beijing who is called the "Master Administrator." We do not know who that is yet. But this person, according to them, had access to all data in the TikTok system.

When they say that this data is being kept separate, it is simply a point that has been disproven already, because we have documentation that shows that the data has been shared extensively.

Senator PORTMAN. OK. We will get a chance to talk to TikTok about that, but I appreciate your work on this and your testimony today. We have cause for us to legislate more in this area, generally. We talked about that earlier, regulations, legislation.

My concern is that we really do not know what is behind the curtain, the black box, so to speak. We proposed this legislation called the Platform Accountability and Transparency Act to require the largest tech platforms to share data, again with vetted, independent researchers and other investigators. We know what is happening with regard to user privacy, content moderation, product development. We talked about the bias that I believe is out there in social media today, in many of the companies, and other practices.

Mr. Boland, you talked a little bit about this in your testimony. I see in your written testimony you said, "To solve the problem of transparency we must require platforms to move beyond the black box with legislation like the Platform Transparency and Accountability Act." Can you explain why that legislation is needed and how it would be used?

Mr. BOLAND. Thank you, Senator. Yes, I believe the Platform Accountability and Transparency Act is one of the most important

pieces of legislation that is before you all. It is not sufficient because we have to address the incentives that we have been talking about.

But to begin with, we are at a point where we are supposed to trust what the companies are telling us, and the companies are telling us very little. I think Facebook, to their credit, is telling us the most, but it kind of like a grade of a D out of an A through F grading system. They are not telling us much, but they are telling us more than everybody else, especially YouTube and TikTok.

In order to understand the issues that we are concerned about with hate speech and the way that these algorithms can influence people, we need to have a public understanding and a public accountability of what happens on these platforms.

There are two parts of transparency that are very important. One is understanding what happens with moderation, so what are the active decisions that companies are taking to remove content or make decisions around content. There is another critically important part that is around what are the decisions that the algorithms built by these companies are taking to distribute content to people.

If you have companies reporting you what they would like to, and I am sure you will hear from them this afternoon, a lot of averages, a lot of numbers that kind of gloss over the concerns, if you look at averages across these large populations you miss the story.

If you think about 220-some-odd million Americans who are on Facebook, if one percent of them is receiving an extremely hate-filled feed or radicalizing feed, that is over two million people who are receiving really problematic content. In the types of data that you are hearing today, that you are receiving today, you get an average, which is incredibly unhelpful.

By empowering researchers to help us understand the problem we can do a couple of things. One, we can help the platforms, because today they are making the decisions on their own, and I believe that these are decisions that should be influenced by the public. Two, then you can bring additional accountability through an organization that has clear oversight over these platforms. Whether that be through new rules or new fines that you levy against the companies, you have the ability to understand how to direct them.

Today, you do not know what is happening in the platforms. You have to trust the companies. I lost my trust with the companies, of what they were doing, and what Meta was doing. I think we should move down trust to helping our researchers and journalists understand the platforms better.

Senator PORTMAN. OK. To the other two quickly, Mr. Cain and Mr. Roetter, do you disagree with anything that was said about the need for more transparency? Just quickly. I have very little time.

Mr. CAIN. I 100 percent agree.

Mr. ROETTER. I 100 percent agree, and I think that this Committee is uniquely poised, given its subpoena powers, to enforce transparency.

Senator PORTMAN. OK. I will have additional questions later. Again, we have so many Members here I want to respect the time, but I appreciate your testimony.

Chairman PETERS. Thank you, Ranking Member Portman.

Senator Johnson, you are recognized for your questions.

OPENING STATEMENT OF SENATOR JOHNSON

Senator JOHNSON. Thank you, Mr. Chairman.

Listen, I think we all agree this is a big problem. It is my definition of a problem, that there is no easy solution. As Chairman of this Committee I would meet with Facebook, and I appreciated what they were trying to do to hold down Islamic terror type of content. I think we all agree that we do not want to be disseminating extremist, violence-inducing type of behavior through these platforms, but we also need to protect free speech as well. It is a real tension. It is a real balancing act.

Mr. Boland, I think you talked about the term extremism and harmful content, but I guess that is all in the eye of the beholder, is it not? It is difficult to define.

I guess what I want to focus on a little bit is, where do we draw the line in terms of taking down content that we all would agree is extreme and could induce violence, versus censoring legitimate political debate?

Mr. Roetter, do you have any idea of what percentage of Twitter employees are conservative versus liberal?

Mr. ROETTER. I have no idea.

Senator JOHNSON. You think it is probably pretty heavily tiled to the left. Correct?

Mr. ROETTER. I do not know.

Senator JOHNSON. I think you do.

Mr. Boland, would you want to answer that question?

Mr. BOLAND. I do not know.

Senator JOHNSON. OK. Mr. Cain.

Mr. CAIN. It is obvious, but my only knowledge is China tech and TikTok. I am not as familiar with that area.

Senator JOHNSON. OK. Let me move on. Let me use an example that I think we are all aware of, the 800-pound gorilla in the room. Let us talk about the Hunter Biden laptop. Mr. Roetter, do you believe that, like The Washington Post, that there is authentic information on that laptop?

Mr. ROETTER. I am not sure. I will say that these are massive platforms. There are billions of people.

Senator JOHNSON. I have very little time.

Mr. ROETTER. I do not know.

Senator JOHNSON. Mr. Boland, do you assume that is authentic information on the laptop?

Mr. BOLAND. I do not have an opinion on the laptop.

Senator JOHNSON. OK. Twitter was actually very effective when they blocked the New York Post articles on the Hunter Biden laptop. We had Jack Dorsey in front of the Senate Commerce Committee back in, I think, October 2020, and Senator Cruz and myself asked him, because we were talking about Russians using the platforms to impact our elections, and everybody agrees that could happen.

We asked Mr. Dorsey, "Do you believe Twitter could impact the election?" Mr. Dorsey said, "No." Mr. Roetter, do you believe Twitter has the capability of impacting an election?

Mr. ROETTER. I think all of these social platforms, they are so massive it is hard to believe that they are not impacting.

Senator JOHNSON. Mr. Boland, do you believe that as well?

Mr. BOLAND. Yes, these platforms absolutely have influence.

Senator JOHNSON. Mr. Cain.

Mr. CAIN. Absolutely.

Senator JOHNSON. OK. There is a problem right there, OK, and I appreciate you acknowledging that fact. We had 51 former intelligence officials. I have no idea on what basis they wrote this letter, that came out immediately. I think it might be because the Federal Bureau of Investigation (FBI) had a scheme in August 2020, to downplay the derogatory information on the Hunter Biden laptop. But they came out and said that the laptop had all the earmarks of a Russian information operation. It seems to me like that letter itself was an information operation.

We have the platform censored that and Facebook throttled it back. We actually took a poll on this. I did not but a company called Media Research Center poll—this was after the election—1,750 voters in seven swing States, of Biden voters who were unaware of the emails, texts, testimony, banking transactions on the laptop, as well as Senator Grassley's and my report, which was based on interviews with U.S. persons and U.S. documents.

Seventy nine percent of those Biden voters said they would still vote for him, but 16 percent said they would not, four percent said they would switch their vote to President Trump, four percent would vote for a third party, four percent would skip voting altogether, five percent would not have voted at all. Pretty strong evidence that what Facebook and Twitter did impacted the 2020 election to a far greater extent than anything Russia ever could have hoped to do in 2016 or 2020.

I want to talk about other disinformation coming out of this Committee. A day or two after Senator Grassley and I issued our report, based on U.S. documents and interviews with U.S. persons, our now Committee Chairman, who was then Committee Ranking Member, issued a press release. It said, "Peters response to a Republican effort to amplify Russian disinformation." He said, "I generated a partisan political report that is rooted in Russian disinformation."

Mr. Chairman, do you want to retract that false allegation now, now that we know that the Hunter Biden laptop is accurate, that there has not been one scintilla of information provided in Senator Grassley's and my report that has ever been refuted. It was 100 percent accurate. Yet you, as Ranking Member of the Committee, accused me repeatedly of soliciting and disseminating Russian disinformation. Do you want to retract your false allegation here that you issued in your press release on September 23rd?

Chairman PETERS. No. Let us focus on what we are trying to——

Senator JOHNSON. I am focusing on this because this is exactly the type of harm we can do to our political process when you have these big tech companies engaging in political debate, censoring one side of the political spectrum and amplifying the false allegations of another side. Do any of you want to dispute that?

Mr. BOLAND. Senator, I think it is important that we get the data to know. This is why the Platform Accountability and Trans-

parency Act is so critical to our globe and our Nation, is that if you were able to look at the data, to understand what had happened from content moderation, and you were able to see the distribution, you could compare that data across the platforms and see what sort of impact that it had.

Senator JOHNSON. One part of the transparency would be to at least have people who at least used to work or work for these platforms to at least acknowledge the highly political nature of the individuals that work in them. Just acknowledge it. It is obvious to everybody. Mr. Zuckerberg spent, what, about a half a billion dollars impacting the 2020 election? Took over the Green Bay election system, in a highly partisan fashion. About 95 percent of the money he spent was in Democratic strongholds in Wisconsin.

Can we at least acknowledge that there is enormous political activity going on, partisan activity going on, within these social media companies, rather than just trying to bury it? Let us be honest. Let me be transparent. But let us be honest in our transparency.

Mr. BOLAND. I agree with you on the request for transparency. My experience, outside of whether someone had a certain political leaning or not, I did not see political leanings shape the decisions that were made inside the company, per my experience, and what I saw.

Senator JOHNSON. OK. I sat it in their censure of the New York Post article prior to the 2020 election, and I think it is pretty obvious.

Thank you, Mr. Chairman.

Chairman PETERS. Senator Lankford, you are recognized for your questions.

OPENING STATEMENT OF SENATOR LANKFORD

Senator LANKFORD. Thank you, Mr. Chairman. Mr. Cain, you have spent a lot of time studying authoritarian regimes and how they use social media to be able to literally control their own populations. The Chinese government obviously doing this with the Uyghurs and what they have done. In that case you have spent a lot of time studying and going through.

One of those features that is in TikTok, for instance, and in several platforms, is the permissions. When you join it you use this free platform. The user gives ByteDance, TikTok, whoever it may be, the opportunity to be able to open their microphone, to use facial recognition, to be able to store data on that. How is that information used in an authoritarian regime?

Mr. CAIN. An authoritarian regime such as China will attempt to get access to that data and use it to build artificial intelligence capabilities, capabilities that might involve espionage, spying on military officials, government officials. This is a major Trojan Horse that needs to be dealt with, and the Chinese government has made clear, in its National Artificial Intelligence Strategies, that it does need data, that data is its biggest target.

Senator LANKFORD. Right. One of the things that I have seen from TikTok even recently is the ability to be able to keep up with keystrokes. If you use their app to be able to then go to other websites so they can then track your keystrokes, that would in-

clude credit card numbers, that would include passwords, user IDs, all of those things as well. Factual or not factual?

Mr. CAIN. Factual. You are absolutely correct, Senator.

Senator LANKFORD. They have made the statement publicly, "We do not use that for any other purposes. We just maintain that." That is now owned by the Chinese government. At that point if it is going through TikTok they have access through that to be able to get user names, passwords, facial recognition, everything else, on this. That is the building of a database system.

This is not some hypothetical, possible thing. This is actually occurring.

Mr. CAIN. Precisely, Senator. This is occurring. TikTok does get gather large numbers of data, and there was one recent study by Citizen Lab, which does work on this, which found that TikTok does gather unusually large amounts of data from its users.

The key login software that was found recently and reported on, TikTok has said that they do not use this. But it is there, and if the Chinese Communist Party wants to get access to it they have the power to do it.

Senator LANKFORD. OK. Thank you.

Mr. Roetter, I want to ask you a little bit about value system. You have a unique perspective from coming from Twitter and then outside of it to be able to look backwards on this. It is unusual to me that Twitter is blocked in many authoritarian countries, but yet the leaders of those countries are allowed to be on Twitter and to be able to put out authoritarian propaganda, basically. They are still allowed to be able to do that.

Twitter's value system seems to shift from country to country, based on that country. Even if that country blocks them from a platform, they are still allowed to be able to put out the propaganda from that platform. Am I wrong on this? What have you seen?

Mr. ROETTER. Twitter certainly is obligated to follow the laws of all the countries where it is operated. That is a consequence of what you are seeing.

Senator LANKFORD. But it also seems to be a patchwork of values in these countries, where in our country they will say, "We really stand strong for this principle," but in another country they do not.

Mr. ROETTER. I think that is a fair characterization.

Senator LANKFORD. Is that a problem long-term or giving authoritarian regimes a platform through that is just a matter of having customers there, even if those countries actually even block the use of Twitter in their country.

Mr. ROETTER. I think the bigger problem is the consequences of their overwhelming incentive model. Not being deeply specific about that example, of per-country variation. That is a consequence of trying to get everyone to use the platform and then being subject to some constraints, whether it is local governments or some other constraints.

I think the broader problem is the consequences of who sees what content and what that does to people in the real world as a function of the incentive structure that is created for these companies.

Senator LANKFORD. Recent testimony on Twitter has come out that they have had on their time Chinese spies, individuals that work for the Indian government, individuals that work for the Saudi government, that were on the staff and were funneling information back to those authoritarian regimes from the staff.

I would assume Twitter has a process of actually going through and vetting their employees. I am making the assumption that while you were there you saw some of the vetting of how this actually happens. How are they vetting their employees to be able to evaluate individuals so they do not end up with Chinese spies, Saudi employees from the Saudi government, or required from the Indian government, for instance?

Mr. ROETTER. It may have changed in the time that I was there. There are background checks and other things you go through when you get hired, but there was nothing I saw that made me think that process was designed to counteract a threat model of governments inserting spies. It was much more pedestrian of a process than that.

Senator LANKFORD. It seems to be a different issue when you talk about the Indian government saying we are requiring our individuals to actually be on your staff, to be able to be in the process, or to allow individuals, as was accused from the Saudi government, to be there on the staff. That does not seem to be a vetting issue. That seems to be a requirement. If you are in our country we also require backdoor access, basically.

Mr. ROETTER. I am not sure I am familiar with the specific rules from India and Saudi Arabia in terms of operating in the country.

Senator LANKFORD. Fair enough. But this is an issue we will have an opportunity in the second panel to be able to talk about this.

Mr. Boland, you have spoken out often on the algorithms that are out there and dealing with basically how the platforms seem to engage angry comments. The angrier that you become, the more it helps the algorithm to be able to engage and to be able to place this.

I have done a recommendation to Facebook for years to say why could not the page owner, in that sense, be able to take the comments, if people want to make comments, those comments come to the page itself, that individual, but other individuals cannot see it. There is basically an option that you could create to turn off the public viewing of all the comments. If you want to make comments to me, and we want to have dialog, you can do that. But it prevents people that make comments on my page from attacking other people that make comments on the page.

Basically what Facebook has created is a place for people to scream at each other in a lot of the political dialog, and it is pretty hostile, and it reinforces the anger comments to continue to be able to drive that.

What I am describing to you, of giving the user, the owner of a page, the option to be able to make the comments between myself and those that are making comments so they cannot attack each other, is that technically possible, to be able to do?

Mr. BOLAND. That is technically possible to be able to do. I think an important step for a lot of the product development work is back

to this transparency point of if we could have researchers and academics involved and evaluating the different types of scenarios here and the tradeoffs, I think that could move us a lot faster forward.

Senator LANKFORD. Yes. The value system of Facebook seems to be that we want that engagement across and that anger and that attack to each other because that keeps people engaged. Instead of trying to lower the temperature and saying on this page the temperature is going to be lower, it seems to be high temperature in as many places as possible, and the anger emoji or anger responses seem to build in that algorithm to continue to accelerate coming back to that page over and over again.

Mr. BOLAND. Yes, and I think an important point that Alex has made is that the algorithm knows no temperature. It does not know if something is charged or not charged. It just knows whether it gets a result. It knows if it gets engagement.

Without there being a qualitative view over the kinds of content, the algorithms will just chase what they are told to chase, and they are really good at it. They are really good at going after the metrics that they are given, and as machine learning techniques improve you will see more and more of that. The idea that it chases that kind of content, if that is what gets engagement and that still gets reaction, that is what is going to grow in the system.

Senator LANKFORD. OK. Mr. Chairman, thank you.

Chairman PETERS. Thank you, Senator Lankford.

Senator Romney, you are recognized for your questions.

OPENING STATEMENT OF SENATOR ROMNEY

Senator ROMNEY. Thank you, Mr. Chairman. Mr. Roetter, your comments about the incentives of a corporation are accurate, which is they are trying to make as much money as they can, and to do so they are trying to get as many eyeballs as they can. Every newspaper, every magazine, every TV show, every broadcaster, every radio station does the same thing, which is how can we get more eyeballs? What gets more attention?

What we are seeing with social media is not entirely unprecedented. I was not around in the early days of broadcast, but I presume it was Wild West initially, and then there was threat of government intervention to tell you who could advertise and what they could advertise and what words could be said and how much sex and violence could be allowed.

The industry came together and said, OK, we are going to start grading things and establishing rules. Ultimately, from what I understand, the government also established an entity to establish rules for broadcasting, saying what you could show at certain hours of the day, how much sex and violence, and so forth, could be on broadcast networks.

We have not done that with regard to social media. Social media is far more engaging and captivating of our young people, as well as many adults, than broadcast was. I wonder whether we need to do this, one, whether the industry should not come together and talk about its own decisionmaking, the rules, where they draw the lines, and say, yes, these are things we are all going to agree to. If the industry does not do that whether government should,

whether we should establish an agency to say, hey, these are the rules, and you are all going to have to follow them.

Is the industry willing to take action of that nature? Could it? Should it? If not, should government? I will ask you, Mr. Roetter, first, and then Mr. Boland, and Mr. Cain.

Mr. ROETTER. I think probably the best predictor here is just past behavior. My observation is the industry will share information, which is not the information you would share if you were generally interested in providing transparency. Brian has talked somewhat about sharing an average, when you have a distribution which is so non-uniform that an average is not a useful statistic. We see that. We see exact numbers being shared.

I view this as a two-step process. The first is we actually do not know what is happening on these networks today, and that is why a lot of the conversation about networks devolves into cherry-picking examples to prove something that I believe these networks are doing. You can also prove that they have a bias in the other direction, depending on which example you cherry-pick.

The first step is we need statistically representative, unbiased, raw data that can be processed and then we understand what they are actually doing. I think if we had a better view of what they are actually doing in a representative way we could then talk about do we believe their incentives are going to create the right outcome, what is the true impact of that. Then, faced with that shared understanding of what the networks are doing, it is possible, to your point, that maybe the companies would come together and decide to self-regulate because they realize the specter of someone else regulating them is worse. I am not sure.

But I think really the first step is demanding the transparency so we have a shared view on what is actually happening on these things.

Senator ROMNEY. Yes. Thank you. Mr. Boland.

Mr. BOLAND. Yes, I think that usually you see industries take self-regulatory steps when they feel like external pressures or legislation is impending, and I do not believe that they feel that with the United States today. I think that drives decision from them to kind of let the status quo go.

I think what is particularly terrifying is that we do not know what is happening on these platforms today. The nice thing about broadcast is that you know what was broadcast. Everyone could see what was being broadcast. With the way that our feeds work today you have such a distribution of content that it can look, on an average basis, that things are getting better. The industry can tell you, "We are making improvements. Here is the average that shows you what we are doing." But for the person who has the 99th percentile most hate-filled feed or the group of people that have that 99th percentile most extremist feed, they may be seeing an increase in the types of harmful content that we will never know.

Meta provided a lot more transparency three years ago than they provide today; that transparency is decreasing—none of the other platforms are taking steps in that regard to increase transparency, and you have an increased TikTok-ization of media in that Facebook is now moving toward a TikTok model, where it is not just friends and family content. It is unconnected, algorithmically

driven content. These kinds of fringes, these pockets, are going to grow. We will never see them unless we mandate that we get to see them.

Senator ROMNEY. Thank you. Mr. Cain.

Mr. CAIN. Thank you, Senator Romney, for the question. TikTok has seen an alarming number of leaks get out in the press, which does suggest, and I have spoken personally with former TikTok employees who have been disgruntled, who say that the company is not transparent, that it is not doing what it promises. Since it has this connection to the Chinese Communist Party it will not be transparent about what it is doing, that this algorithm, we do not know what is behind the curtain.

China is not a place that values transparency. It is a one-party, authoritarian state, and the most sophisticated police state in the world. I do not think we can count on a company such as TikTok or its parent, ByteDance, to do anything that will actually address the problems at hand.

I think that, to be honest, this Committee is uniquely placed to address this problem of transparency because the subpoena power that can be used here I think would require TikTok to open up its emails, show us what is really going on, and show us what the China-based executives are saying with the American executives.

Senator ROMNEY. Yes, I must admit I share your view in that regard, although I am probably even more alarmist than you, which is I question whether we should allow an authoritarian regime to have a social media capability of the scale they have in our country, gathering the data they have. I think that is a huge risk for us.

I have a lot of kids and grandkids, particularly grandkids these days. I am very concerned about their exposure to social media. Have other countries figured a better way to try and reduce the draw and the compelling nature of social media? I understand China, for instance, between various TikTok segments they have a five-second gap where the screen just goes blank or something. Are we not doing even what other nations are doing to try and protect our kids? I will let anyone that wants to respond to that. Maybe Brian, if you want to take that.

Mr. BOLAND. Yes, it is a tricky one in that you are dealing with incentives you would mandate versus steps you would like companies to take on their own. There are steps. You are describing friction, right, that slowdown——

Senator ROMNEY. Right.

Mr. BOLAND [continuing]. The process. There are known steps you can take to introduce friction in the products.

Senator ROMNEY. Have some other nations done some of those things?

Mr. BOLAND. I do not know of mandated friction. I think Europe has done a very good job of starting to move toward required transparency, so we will see how that moves. But I have not seen the prescriptive type of products.

Senator ROMNEY. Thank you.

Chairman PETERS. Thank you, Senator Romney.

Senator Hawley, you are recognized for your questions.

OPENING STATEMENT OF SENATOR HAWLEY

Senator HAWLEY. Thank you, Mr. Chair. Thanks to the witnesses for being here.

Mr. Boland, let me start with you. Can you tell me, when were you at Facebook?

Mr. BOLAND. I was there 2009 through November 2020.

Senator HAWLEY. November 2020. Was it normal in the time you were at Facebook for executives or team members at Facebook, not even have to be executives, to coordinate closely with the United States government?

Mr. BOLAND. I am not aware of that.

Senator HAWLEY. You were never in any such meetings?

Mr. BOLAND. No.

Senator HAWLEY. You never had any contact with U.S. Government employees in your time at Facebook?

Mr. BOLAND. Not that I can recall.

Senator HAWLEY. Would you be surprised to know that on July 16, 2021, an employee at Facebook wrote to the Department of Health and Human Services (HHS) saying, "I know our teams met today to better understand the scope of what the White House expects from us on the misinformation going forward."

On July 23, 2021, Facebook employee thanked HHS for "taking the time to meet earlier today. I wanted to make sure you saw the steps we just took this past week to adjust policies in what we are removing with respect to misinformation." That included, and I am quoting, "increasing the strength of our demotions for Coronavirus Disease 2019 (COVID–19) and vaccine-related content."

On April 16, 2021, Rob Flaherty at the White House circulated a Zoom meeting invitation stating, "White House staff will be briefed on vaccine misinfo."

On April 7, 2021, a Facebook employee thanked the Centers for Disease Control and Prevention (CDC) for responding to misinformation queries. "We will get moving now to be able to remove all but that one claim as soon as the announcement and authorization happens."

On July 28, 2022, this year, a Facebook employee reached out to CDC about "doing a monthly misinfo/debunking meeting." CDC responded, "Yes, we would love to do that."

On May 11, 2021, Facebook employees organized a be-on-the-lookout meeting with CDC officials.

On July 20, 2021, Clark Humphrey at the White House emailed Dave Sumner and others at Facebook asking, "Any way we can get this pulled down," and cited an Instagram account. Within 46 seconds, Facebook replied, "Yep. We are on it," and down the account went.

Is that normal? Is that normal in your time at Facebook?

Mr. BOLAND. I do not have experience around that.

Senator HAWLEY. You have no knowledge of anything like this. Nothing like this ever happened, and then, presto, it started happening just suddenly in 2020, as soon as you left?

Mr. BOLAND. I did not have personal experience with that, or I did not hear about it.

Senator HAWLEY. You do not know anything about it at all? You have never heard of anything like this happening, ever?

Mr. BOLAND. I do not.

Senator HAWLEY. That is remarkable. I thought that you were the former Vice President of Partnerships Product Marketing, Partner Engineering, Marketing, Strategic Operations and Analytics at Facebook.

Mr. BOLAND. That is true.

Senator HAWLEY. None of this ever happened. Why did it start happening, do you think, as soon as you left? What do you think drove this kind of collaboration, where you have Facebook becoming an arm of the United States government, more specifically the White House, to censor private information, personal speech, at the behest of government officials?

Mr. BOLAND. It is hard for me to comment on the specific context of the content we are talking about, whether it was public content or whether it was personal content. I do know that from what I have read, and probably the same documents that you have access to, that there were a lot of steps taken around COVID response and COVID misinformation that may have presented a unique scenario and a unique situation where the company took steps to coordinate that way.

Senator HAWLEY. Took steps to coordinate, by which you mean to censor the speech of ordinary Americans at the best of the President of the United States and his Administration.

I commend to everyone who is interested these emails which were discovered as part of litigation led by the State of Missouri and other States as they are suing these tech companies, including your former employer, Mr. Boland, which, for the record, is one of the worst companies in America. They have discovered this trove of information, extensive coordination, extensive, between Facebook and the Biden administration, targeting the speech of ordinary Americans. By the way, for standards that are ever-shifting.

Early on, if you questioned that COVID had anything to do with a lab you were marked as disinformation, you were censored, only to have the President of the United States later admit the possibility that COVID has some lab nexus is, in fact, a very distinct possibility that our intelligence communities (IC) think is actually quite a viable theory. We have seen the same thing with people who have questions about masks, who have questions about vaccine efficacy. It is really quite remarkable.

Let me ask you this. What safeguards, when you were at Facebook, were in place to protect Americans from having their speech censored or having government censors like this access personal information?

Mr. BOLAND. During my time there my experience was the company was more reluctant to take down speech and very careful about trying to remove content. I also do not think the company studied the content on the platform as heavily as you would like.

Senator HAWLEY. They did not do things like, not like what they were doing later, when they were looking at particular private Instagram accounts and removing them at the behest of the White House? You are saying that did not happen while you were there?

Mr. BOLAND. That is not a scenario that I ran across. It is hard for me to comment on COVID pandemic response, which I think a lot of things were outside of the norm.

Senator HAWLEY. I will just say this. I find it hard to believe that suddenly Facebook became an entirely different entity and was interfacing with the United States government in an entirely different way only when COVID happened. I mean maybe, but I doubt it.

Let me ask you this, Mr. Roetter? You were an engineer at Twitter. Is that right?

Mr. ROETTER. Correct, yes.

Senator HAWLEY. You were the Senior Vice President for Engineering?

Mr. ROETTER. Yes.

Senator HAWLEY. Yesterday Mr. Zatko testified to another Committee I sit on that 4,000 engineers at Twitter had access to all of the personal information, user data, geolocations, of Twitter users. Is that accurate?

Mr. ROETTER. I have never met him, and he joined the company after I left, so I do not know if that particular claim is accurate.

Senator HAWLEY. But he said all the engineers. You were an engineer. Did you have access to user data?

Mr. ROETTER. When I was there, I do not know if it was all the engineers.

Senator HAWLEY. Did you have access to user data?

Mr. ROETTER. I was head of engineering for the whole company.

Senator HAWLEY. Did you have access to user data? I am looking for a yes or a no.

I will just remind you.

Mr. ROETTER. No.

Senator HAWLEY. You did not have access to user data?

Mr. ROETTER. I think I could have gotten it.

Senator HAWLEY. I am sorry?

Mr. ROETTER. I think I could have gotten it.

Senator HAWLEY. OK. If you can get it, that is what we call access. You did have access to user data. Is that a yes?

Mr. ROETTER. When I was there I probably could have. Yes, so that is probably right. Yes, I probably could have.

Senator HAWLEY. OK. You did. Did you ever access any user data?

Mr. ROETTER. No.

Senator HAWLEY. Were you aware of Twitter engineers ever doxing members, of users?

Mr. ROETTER. No.

Senator HAWLEY. Were you aware of Twitter engineers ever taking over an account and tweeting out or altering the content of that account? Mr. Zatko said he thought that had happened.

Mr. ROETTER. I am not aware of that.

Senator HAWLEY. OK. Lots to unpack there. Thank you, Mr. Chair.

Chairman PETERS. Thank you, Senator Hawley.

Senator Rosen, you are recognized for your questions.

OPENING STATEMENT OF SENATOR ROSEN

Senator ROSEN. Thank you, Chairman Peters, Ranking Member Portman. Thank you to the witnesses for being here today.

Transparency and accountability, I guess those are the words of the day, because we know that social media companies, of course, what we do. I am a former computer programmer. Data is power. How you analyze the data, the data tells a story if you are smart enough to listen to it.

You collect the demographic behavioral data from consumers in order to enhance the predictive engagement algorithms to target the consumers with ads, recommendations based on other content, perceive interests, or even vulnerabilities. This is really great when you are shopping for a new outfit or some new furniture. Maybe not so great when you are on an extremist or violent website or harmful or illegal content.

When it comes to that harmful or illegal content there has to be greater transparency into the platform promotion mechanisms, and how the content ultimately spreads from platform to platform. We have small businesses, hospitals, schools, everyone are on these platforms in some form or fashion.

When we say "consumer," we can go from the individual right up to our full national security, that we understand better the algorithms that amplify the content and how these things reach their feed. Some social media platforms, for example, they have standards in place for moving content that promotes Holocaust denial or distortion. They are often inconsistent with implementing the policies, but the content flourishes.

I am going to cut right to the chase. Mr. Roetter and then Mr. Boland, is there a difference now in how predictive user engagement algorithms behave for harmful, illegal, or extremist content versus other content, and how might we modify or regulate an otherwise agnostic algorithm—It is math; it is agnostic, to your point—an agnostic algorithm to identify illegal, certainly illegal, or extremist content? How do we take the agnostic out of the math? Mr. Roetter.

Mr. ROETTER. Today the algorithms are doing exactly what they are incented to do, which is maximize attention on the platform. If you changed what those companies were accountable for, these companies are very smart. They have a lot of engineers and a lot of money and a lot of computational power. They would change what the algorithms do.

For example, if companies were penalized for sharing certain types of content, these algorithms would no longer promote that content because it would be not optimal for them to do so. The extra benefit they get from the attention and the usage would be outweighed by whatever the penalty was.

This is all possible. I think the two takeaways are, one, without transparency we are not going to know what it is doing today, and two, they will behave optimally in the face of any incentive structure they have. Today it is just maximizing intention, and they are doing exactly what you would expect, but you could change that.

Senator ROSEN. Thank you.

Mr. BOLAND. I think it is important to note that not only do we not know what is happening on the platforms, the platforms do not know what is generally happening on their platforms. The turning point for me, to go from having concerns to being publicly vocal about my concerns, was when Facebook said, "Nothing on January

6th happened on our platform." Then it turns out, after the fact, that we find out that there is a lot of Stop the Steal on the platform, and there were internal concerns around it.

In order to change these algorithms part of it is understanding what is happening, and as a society having conversations about what do we think the right distribution is. Facebook has proven that with things like QAnon, after the fact, after the fire was lit and burned through, they could then adjust it and actually manage the distribution of that type of content. It is possible when we know what we are managing toward.

The problem is that it is all after the fact. It is all after the damage has been done that you then go back and say, OK, there has been this set of articles or conversations, and finally we go back to address them, rather than saying we have a whole community of researchers and people who can quickly spot thing, raise the issues, and then adjust them. It is doable. It is an incredibly hard problem. Like I am very sympathetic to the fact that human speech is very complicated and very nuanced.

Senator ROSEN. But the platforms have an unwillingness. They actually want to have this lack of understanding so they have some deniability on the back end, if that is what you are saying is true. We do not know it is happening. Oh my gosh, it happened after the fact. Their lack of wanting to do the analysis ahead of time and understand their own platform, they are setting themselves up for deniability, in my estimation.

But we are going to move on to cybersecurity because I have a few minutes left. We know the whistleblower complaint from Twitter's former head of security depicted they were unable to protect the 238 million users, government agencies, influential figures, heads of State, from spam and security breaches. The complaint alleged the company servers were running out-of-date and vulnerable software and withheld dire facts about the breaches and lack of protection for user data.

I am really concerned about cybersecurity. Companies are laser-focused on growth, not laser-focused on protection, in my estimation, so individuals—again, small businesses, hospitals, schools, critical infrastructure, all of those things we are responsible for her are at potential risk.

Again, both of your experiences working at Facebook and Twitter, is cybersecurity a high enough priority for the large social media platforms, and do the social media platform security teams, do they work alongside product development, application development to protect cyberattacks? Do you have a hunt forward? Are you looking for these breaches? How are you working that, and how does this threaten our own security, even our national security?

Mr. ROETTER. The teams, they do work alongside engineering, but it is not a primary driver the way product and growth and revenue is. You need to build something that drives usage and revenue and then make it secure enough.

In terms of your question, is it a high enough priority, the answer to that can only be known if you know the nature of the threat and if the bad actors trying to break in are being successful.

Senator ROSEN. There is no hunt-forward operations built into these things for people trying to breach the data. There is no kind

of hunt forward. There is no way that people are really actively looking for data breaches. You are finding it after the fact, in many cases.

Mr. ROETTER. No, there are in some cases penetration testing and people trying to simulate breaking into something to learn. That happens.

Senator ROSEN. Can you speak to that, Mr. Boland?

Mr. BOLAND. My sense from Meta's standpoint is that they are quite good and quite invested in protecting people's data, and from a cybersecurity standpoint, which goes to show you that where there is a will and a desire to make progress on issues I believe they can. This is an area where, in my experience, they were quite strong.

Senator ROSEN. Thank you. I see my time is up so I will yield back. Thank you.

Chairman PETERS. Thank you, Senator Rosen.

Senator Hassan, you are recognized for your questions.

OPENING STATEMENT OF SENATOR HASSAN

Senator HASSAN. Thank you, Mr. Chair and Ranking Member Portman for this hearing, and thank you to our witnesses for being before the Committee today. I really appreciate it.

I want to start with a question that builds a little bit on what Senator Rosen was discussing, and this is to Mr. Boland and Mr. Roetter. Terrorists have horrifically livestreamed their attacks on social media. These livestreamed attacks, in turn, inspire other individuals to commit future attacks. Are there ways for social media companies to quickly identify terrorist content so that it is not shared in real time?

Mr. Boland, we will start with you.

Mr. BOLAND. I know that particular for livestreamed videos that Meta has put considerable resources in AI to try to spot these types of attacks and take them down quickly. I think they have gotten a lot better, obviously, than they were with Christchurch years ago. I think it is an incredibly hard problem. I am not an expert on the extent of what is possible there, but I do think they have made strides.

Senator HASSAN. Mr. Roetter.

Mr. ROETTER. It is certainly possible. It is a hard technical challenge but you can build algorithms to figure out, in real time or near real time, the content of videos. They will not be perfect, like any sort of classification or segmentation algorithm, but you could certainly do so.

Senator HASSAN. Yes. I mean, this is an ongoing issue because, of course, we are seeing the acceleration from idea to action happening much more quickly in part because of the influence of social media too. I thank you for that and I would look forward to following up with you both on it.

Another question for the two of you. Facebook is currently running an advertising campaign which is touting the thousands of employees and billions of dollars that the company says it spends on safety and security. These numbers, however, are pretty meaningless without proper context, right?

What specific information or metrics should these companies provide this Committee to help us fully understand their actual commitment to safety and security? Mr. Boland, again I will start with you.

Mr. BOLAND. Yes. You are 100 percent correct on the context of the numbers matter. When they first announced their $13 billion over five years safety and security number it was in the context of $50 billion in stock buybacks, so like a massive imbalance of investment. They also will give you numbers of employees. Numbers of engineers matter. If you think about these issues, you can have employees who are non-technical who can be in what would be a review queue or a process to look at content. But the really important thing is engineering resources and how many engineers are put on these problems.

I would really try to get from the companies an understanding of where they allocate their engineers, for these types of problems. They do not need to show you their entire organization chart and you get to know how many are working on Metaverse and whatnot, but these are the numbers working on these safety and security issues, these safety issue, and this is how they are allocated by country, by topic, et cetera. I think that is justifiable to understand, and to feel like we have a sense of whether that is adequate relative to the total number of engineering employees.

Senator HASSAN. OK. Thank you. Mr. Roetter.

Mr. ROETTER. I think what is important is that we get metrics that are of the form that show what results they are getting, not metrics that basically equate to, "We are trying really hard. Give us a break." That would never work in Wall Street. You can tell them we tried really hard to make profit this quarter. You have to actually show what the results are.

If you have transparency over the content and how the content is shared, and the engagements on that content, we will be able to study. Independent people can look and see certain content spreads very widely, other content does not—and then after this investment that they have made has this changed or not?

We need metrics where we can measure the actual result, not just, "Oh, I tried really hard, so please be happy."

Mr. BOLAND. I am sorry. One more quick thought there, is that I worry that a lot of times, because it is so painful, we focus on these extreme examples of content, or the livestream shootings. There is a broad swath of content that influences people that does not feel as scary. That is the stuff that terrifies me, and that is the stuff we do not get to see without transparency.

Senator HASSAN. OK. I thank you both. I thank the whole panel for your testimony, and I am very grateful for this hearing, and I will yield back.

Chairman PETERS. Thank you, Senator Hassan.

During your opening statements I think each of you discussed the product development process at these companies, and we have talked at length about that process through the hearing.

Mr. Boland, you discussed how Facebook does not incentivize limiting the spread of harmful content but, of course, prioritizes growth and revenue. Could you tell the Committee generally what

metrics inform employee compensation at the company? What goes into that?

Mr. BOLAND. Employees at Facebook, it is about rewards, right, so the rewards that you receive is your cash compensation, your stock compensations, and promotions. Generally if you are building products you are rewarded on the success of that product, and that product success is defined by some set of metrics around whether that product is being used more.

Let us pretend that you are building a video product. The things that you will care about are the metrics around what are the total watch hours, how many hours are being spent watching videos, what is the user growth, how many people are using that video product, where is that spread geographically, et cetera. You are incentivized on those hard metrics, and then you are not incentivized around, what kind of content are you growing your video with? What is the stuff underneath the hood that is showing up, that is driving this growth? That is not your problem. That is somebody else's problem.

The problem is that does not drive individual behavior. Company goals are kind of there. You do not think about them. You think about what you, individually, and your team are goaled to deliver. That is always metrics. That is always product growth metrics and success metrics of the product and not success in are we keeping people safe.

Chairman PETERS. There is not a trust and safety metric?

Mr. BOLAND. For the trust and safety team, actually my understanding is they have been disbanded and moved into a central team. I did not experience products like video or others carrying a metric that was incentivizing trust or safety.

Chairman PETERS. That is not there. Mr. Roetter, is that the case as well?

Mr. ROETTER. Yes. I agree with all that. There is that promotion system, compensation system, review system. The problem with trust and safety metrics, typically companies may have five top-level goals, let us say, and maybe one of them is trust and safety. The problem is that is at odds with the other metrics, and the other metrics always win.

If I am an engineer building, say, a new livestreaming video service, if I launch that product and it gets some usage, that is a feather in my cap. That is something I can say that I did. I can point to its effect. That will help me with promotions, compensation, career advancement.

If, at the last minute, I decide not to launch that product because I realize I cannot control some of the safety aspects and we should not do it if we cannot do it without certain safeguards, I get zero credit for that. It is as if I have done nothing for the company over the last X months.

Chairman PETERS. So you are, in effect, punished, and your future advancement will probably be questionable as well.

Mr. ROETTER. A product that I build and then I do not launch because it might not be safe is no different than if I just did not show up to work, in terms of the future credit that I get.

Chairman PETERS. Not a good place for an employee to be.

Mr. ROETTER. Correct.

Mr. BOLAND. You can change incentives and you can change the way that people show up, but not even just through goals but through process. There was an example where when Facebook started as a desktop site and moved to mobile, Mark Zuckerberg required that all products that were demoed to him showed mobile in their demonstration, that they had designs around that. He kicked the first team out. They came in without that design, and suddenly everybody was thinking about mobile designs.

If, in your process, you create an incentive where you said, as part of every product design discussion is what are all the harmful ways this product can drive hate or drive extremism or drive polarization, you would have a radical change in the way that people showed up to those meetings, and in the process thought about the negative impacts of their product.

Chairman PETERS. All right. Mr. Boland, before you left it is my understanding—and correct me if I am wrong—that you voice objections about how Facebook recommendation algorithms were actually promoting extreme, hateful, and racist content. Is that correct?

Mr. BOLAND. That is correct. My concern was specifically around racist content.

Chairman PETERS. What was the reaction from your senior leadership within Facebook when you expressed these concerns?

Mr. BOLAND. It was disappointing. I raised issues, particularly around the distribution of racist content that I was seeing in the CrowdTangle tool, and my concerns that we did not understand it, and brought forward three steps that I felt would be very good internal steps to actually help mitigate the problem: one, more internal researchers, two, more data to external researchers, and three, beefing up CrowdTangle to share more information.

I had a range of responses from, "You are wrong and that is not the case that this is driving this," with no evidence, mind you. Just, "I believe you are wrong," and no counter-evidence. Too, some, "Yes, this might be a problem but not something that we are working on right now."

Chairman PETERS. But when you say "no evidence" and those statements that say you are wrong, you worked for a company that looks at a lot of data and makes decisions based on data, but this is something they wanted to ignore, basically.

Mr. BOLAND. Yes, so a particularly concerning moment for me was when I had my moment where I really came to terms with believing that the product could be causing harm. I started to look at a variety of things that research teams were doing internally, to understand what they were seeing.

The internal dialog and the internal documents, many of which Frances Haugen has shared, were troubling. There was a particular document that was an overview of polarization research from, I think, June 2020, and that talked about political polarization. One of the lines said, "We have not researched and have very little understanding of racial, ethnic, or religious polarization." That underinvestment was significantly concerning to me.

Chairman PETERS. Yes. Mr. Roetter, one of the documents submitted by a Twitter whistleblower to the U.S. Securities and Exchange Commission (SEC) last month was a 2021 study that he

commissioned of the site integrity team's capabilities. The study found that Twitter planned to launch a new product, Fleets, just weeks before the 2020 elections.

The integrity team, according to that document—and I am quoting the document—said, "Had to beg the product team not to launch before the election because they did not have the resources or capabilities to take action on misinformation or disinformation on the new product." The report also found, "While product teams do elicit feedback for new product launches, product managers are incentivized to ship products as quickly as possible and thus are willing to accept security risk."

Are these findings consistent with the pattern of decisionmaking that you saw?

Mr. ROETTER. With the caveat that that specific example happened after I was there and I cannot speak to it, that is absolutely consistent. In fact, I would be surprised, given the incentives at play, if the product team had done anything else.

One way we used to talk about product managers is they are, the "mini CEOs" of their product, and they get consultation from other teams—trust and safety, legal, finance, compliance—but it is their decision to launch or not. Again, there is no possible credit or reward for not launching, whereas there is possibly a credit or a reward from launching. Because they probably had more confidence that it would at least get some usage and potentially drive revenue, there is every reason to launch and not worry about the other issues.

Chairman PETERS. Thank you.

Ranking Member Portman, any remaining questions?

Senator PORTMAN. Let me follow up on that particular issue. Twitter Spaces, an audio function newer to the platform, was allegedly rolled out in such a rush, to your point, that it had not been fully tested for safety. Twitter lacked real-time audio content moderation capabilities when they launched it.

We are told that in the wake of our withdrawal from Afghanistan it was exploited by the Taliban, and Taliban supporters used this platform to discuss how cryptocurrency can be used to fund terrorism.

First of all, is that accurate? Mr. Roetter, maybe I will start with you. Second, is that common for Twitter to launch products that lack content moderation capabilities? You said that sometimes they are under pressure to ship products as soon as possible. Was that why this happened?

Mr. ROETTER. I is accurate that they are under pressure to ship products as soon as possible, and Twitter, in particular, has a history of being very worried about user growth and revenue growth. It is not the runaway success that Facebook or Google are, and so there was often very extreme pressure to launch thing.

A saying we has is that if you walk around and ask enough people if you can do something, eventually you will find someone who says no. The point of that was really to emphasize you just need to get out and do something.

Again, the overwhelming metrics are usage, and you would never get credit or be held up as an example or promoted or get more compensation if you did not do something because of potential neg-

ative consequences on the safety side or otherwise. In fact, you would be viewed probably as someone that just says no or has a reason not to take action. There is a huge bias toward taking action and launching things at these companies.

Senator PORTMAN. Yes. Are you aware of this Twitter Spaces issue and the Taliban having exploited it?

Mr. ROETTER. That specific example I am not.

Senator PORTMAN. OK. Do you think, assuming my example is correct, which I believe it is, that PATA would have been helpful there, to at least get behind the curtain and figure out why the decisions are being made?

Mr. ROETTER. I have not read the draft of that, but if my understanding is correct, yes, having more understanding of what these products do and what sort of content is promoted and what the internal algorithms are that drive both decisionmaking and usage of the products, I think that would be extremely valuable. Without any of that I would expect examples such as this to keep happening.

Senator PORTMAN. On this trust and safety issue, and specifically the product development and business decisionmaking processes—Mr. Boland, I will maybe direct this to you—Meta disbanded its responsible innovation team just last week it was announced. Did you see that?

Mr. BOLAND. I did. It was extremely disappointing.

Senator PORTMAN. My understanding is they have been tasked with addressing harmful effects of product and development processes. You are saying it was concerning to you. Why are you concerned about it, and tell us about how you interacted with integrity teams while you were at Facebook.

Mr. BOLAND. I know the people who led that team. Very high integrity, very intentional about responsible design of products, as the team was named. Without that kind of center of excellence that is helping to shape other teams I fear that Meta is not going to continue to have that as a part of their conversations.

You can think about that group as influencing and indoctrinating, if you will, the engineers that come to the company of how to start to think about some of these issues. It is less hard-coded into the incentive structure, which I think is a missing element, but would have driven really important conversations on how to ethnically design products.

I do not believe them when they say that they are making it a part of everything, that they are going to interweave it into the company. That is a very convenient way to dodge the question in my view. I do not believe that they are going to continue to investing in it if it is not a team.

This comes at a time when Meta building the Metaverse. We do not know how the Metaverse is going to play out. I am extremely concerned because the paradigms we have seen in the past, that we have started to understand, around content and content distribution, are very different in the Metaverse. That is an area that if I were this Committee I would spend a lot of time really trying to understand the risks of the Metaverse. It feels very risky to me. It feels like the next space where there will be underinvestment,

and without a team of responsible innovation helping to guide some of that thinking that is concerning.

Senator PORTMAN. Again, same question to Mr. Roetter, with regard to how to evaluate these trust and safety efforts in general, and specifically something like the responsible innovation team and what impact it is having. Do you think that it would be helpful to have this legislation called the Platform Accountability and Transparency Act?

Mr. ROETTER. I think so. If we get from that more information to illuminate what these algorithms are doing and what the incentive structures are, that would be extremely helpful.

I think today we are operating in a vacuum, and what we see, a lot of the public conversation about this is people will cherry-pick one example and use it as evidence of whatever their theory is of these companies are doing, that, of course, it must be true because this here is one example.

The fact of the matter is these companies are so massive and there is so much content you can cherry-pick examples to prove almost anything you want about these companies. Without broad-scale representative data from which we can compute what is being promoted and then reverse-engineer what the incentives must be, we are never going to see a change into the things they are optimizing for.

Senator PORTMAN. What are your thoughts on that, Mr. Boland?

Mr. BOLAND. Yes, no, I think the issue that we face today is we have to trust, and without having robust set of data to understand what is happening, and making these public conversations, not company conversations is critical.

Meta would like to tell you that they do not want to put their thumb on the scale when it comes to algorithmic distribution. The challenge is that these algorithms were built in a certain way that you are kind of leaning on the scale already. You just do not realize you are leaning on it. These algorithms today are already doing a lot to shape discourse and to shape what people experience. We do not get to see it, and we have to trust the companies to share with us information that we know that they are not sharing.

As I said earlier, I think the Platform Accountability and Transparency Act is a critical first step. We need to do it quickly, because these things are accelerating, to understand what is actually happening on these platforms.

Senator PORTMAN. Mr. Cain, you have the last word.

Mr. CAIN. I do believe that there are a number of issues that were addressed here today have will have significance not only for our democracy within America but the position of America in the world. There has been just major changes that I have seen personally, having been in China and Russia and recently Ukraine, in the world of technology, in the world of social media. My greatest concern is that we are ceding too much ground to authoritarian regimes that seek to undermine and malign us in whatever way they can.

The software that we are using, the AI, the apps, these are ubiquitous. This is not the Cold War where we had hardware, we had missiles pointed at each other. Now we have smartphones, and it is entirely possible and quite probable that the Chinese Communist

Party has launched major incursions into our data within America to try to undermine our liberal democracy.

Senator PORTMAN. That is a sobering conclusion, and I do not disagree with you, and I appreciate your testimony and our other experts. Thank you all.

Chairman PETERS. Thank you, Ranking Member Portman.

Let me just follow up on a brief question on transparency. It is pretty clear we need transparency. We need to have, though, the active involvement of researchers that use that data and researchers, whether academics or civil rights groups or rights organizations, journalists, everybody has to be engaged.

One pushback we could get, and I would like your response, is do you think there are ways that we can protect user data and still provide the kind of data that is necessary for these researchers? Is that possible?

Mr. Roetter and Mr. Boland?

Mr. ROETTER. Yes, you will get that pushback, as well as a bunch of other pushbacks, I am sure. One, it is possible. We can obfuscate the data. We can generate random ideas that you can hide the personally identifiable information (PII).

Second, there are examples when third-party reviewers have access to confidential information, and because they operate in a professional manner and are well trusted, that does not mean it leaks out publicly. One of the examples I gave was auditors, in the course of certifying financial statements, see a bunch of internal financial performance that if it leaked out would be extremely valuable to competitors. The reason we have third-party auditors is they are allowed to balance the public's need for information with the company's need to keep information private.

We could do the exact same thing. There are secure computing environments. For example, there is a bunch of health care data in the world, which has a bunch of personally identifiable information and very strict legislated private requirements around that, and that is managed in a way such that people can extract insights from the data without violating individual privacy. We could do the same thing here.

Chairman PETERS. Mr. Boland.

Mr. BOLAND. It is absolutely possible and doable. There are some hard aspects to it but it can be done. I think there are two components that I think are favorable there. One, with the increased TikTok-ization of service more and more content is public content, so you are really not dealing with issues around privacy and private data.

For private data, Meta was able to solve it with their ads measurement system. We built a system where we could connect the ads that people saw on Facebook with the purchases that they bought in a physical store. We were able to do that in a privacy-safe way. If we can do it for ads, you can do it for these other areas as well.

Chairman PETERS. Great. Thank you, and I would like to thank the three of you for your testimony here today. You certain provided some very insightful contributions to what is a very important conversation. We appreciate your availability to be part of the first panel for this hearing.

This hearing is going to resume this afternoon when we will welcome our second panel of witnesses, the chief product officers of Meta, YouTube, TikTok, and Twitter.

The Committee will now go into recess, and then we will reconvene at 2:30 p.m.

[Whereupon, at 11:47 a.m., the hearing was recessed, to reconvene at 2:30 p.m. this same day.

The Committee reconvened at 2:31 p.m., in room SD–342, Dirksen Senate Office Building, Hon. Gary Peters, Chairman of the Committee, presiding.

OPENING STATEMENT OF CHAIRMAN PETERS[1]

Chairman PETERS. This morning the Committee heard testimony from experts and former executives at Facebook and Twitter that provided important transparency and context for how many of the largest social media companies operate. Independent and accurate information about how companies balance competing priorities or how they do not, who within the companies make those decisions, and how they build their products is incredibly difficult to find.

This morning's testimony shed some light on many of the areas that this Committee and the public have questions about. I look forward to building on that testimony with our second panel of witnesses, who can speak directly to what steps Meta, YouTube, TikTok, and Twitter are taking to stop the spread of extremist content on their platforms, and I want to sincerely thank each and every one of you for being here today before us.

As we heard from our panel this morning, as Chief Product and Operating officers, you play key roles in your company's decision-making process. You set the agendas for the product teams who are constantly updating the apps and developing new features. You play a prominent role in setting priorities and determining what tradeoffs to make among those priorities, as product teams launch new features or make changes to the apps.

This is the first time executives in your positions have appeared before Congress, and I really do appreciate you joining us for this opportunity to hear directly about your role at these very powerful companies.

The platforms you are representing today reach billions of people around the world. Meta's platforms reach more than 3.6 billion people a month, TikTok has more than one billion users a month, YouTube reaches almost two billion people a month, and Twitter has more than 200 million monthly users. The reach is massive and so is the influence that your platforms wield.

Whether users are fully aware of it or not, the content they see on your platforms shapes their reality, and the business decisions you make are one of the main driving forces of that phenomena. This amount of influence may have a minimal impact on the average user of your platform, but we have seen firsthand how quickly dangerous and extremist content can proliferate online, especially to vulnerable communities or users already on the fringe, and alter how people view the world.

---

[1] The prepared statement of Senator Peters appears in the Appendix on page 97.

Conspiracies like QAnon and Stop the Steal, hateful ideologies like white supremacy and anti-Semitism, and so many more examples of harmful content pollute your platforms. This extremist content can spread like wildfire, amplified by the recommendation algorithms and other tools your team build to increase your companies' audiences and profits. Extremists use the products you designed to recruit and to radicalize followers, and plot attacks, including the January 6th attack on the Capitol, our democracy, and our Nation.

There is no question that there is a relationship between social media amplification of this extremist content and the rise we have seen in hate crimes and domestic terrorist attacks that mark one of the gravest threats to our homeland security. Despite these serious threats, I am concerned that your companies have still not taken the necessary steps to limit the spread of the hateful, dangerous, and extremist content that has motivated real-world violence.

That we all understand exactly the type of extremist content we are discussing today and how challenging this problem is to tackle—it is clearly a challenge—I would like to take a moment to show a few examples, if you would check the screen.

[Video plays.]

This morning we heard from former executives that your companies have no incentive to effectively address the problem this content creates or prioritize the safety of your users as you build and introduce new social media products. Instead, like any for-profit company, your incentives are to prioritize user engagement, grow your platforms, and generate revenue.

I have asked you to appear before the Committee today to answer questions about your companies' incentives and priorities, how those incentives are reflected in how you compensate and promote your product development engineers, managers, and other employees, and to provide important insight into your decision-making processes.

I want to thank you again for joining us today. I am looking forward to this conversation, so that our Committee and the public can better understand this serious problem and how it threatens the safety and security of our Nation.

Ranking Member Portman, you are recognized for your opening comment.

OPENING STATEMENT OF SENATOR PORTMAN[1]

Senator PORTMAN. Thank you, Mr. Chairman. We had a very productive hearing this morning with experts on the impact of social media on homeland security, and I look forward to our discussion this afternoon, and I want to thank the representatives here from Meta, YouTube, TikTok, and Twitter. Thank you all for being here, and in anticipation of another good hearing I appreciate you being very frank with us today and providing us information we need to be able to move forward.

About 300 million Americans now use social media. We know that social media has offered unprecedented connectivity, and that

[1] The prepared statement of Senator Portman appears in the Appendix on page 101.

is often very positive, but we also know it has raised serious concerns for our children, our civic culture, and our national security. Terrorists and violent extremists, drug cartels, criminals, authoritarian regimes, and other dangerous forces have used social media in furtherance of their goals. They have exploited your platforms.

Perhaps the most concerning consequence of social media is the ability for our adversaries to exploit platforms to harm Americans for their own geopolitical gain. As an example, in this second panel I hope we will discuss China's influence over TikTok, which is a social media app that at least one-third of Americans use, and a lot of young people.

As the lead Republican and former Chairman of the Permanent Subcommittee on Investigations (PSI), I have been focused on China's malign activities for many years, and in 2019 I led year-long bipartisan investigation which found that China recruits U.S.-based researchers to steal taxpayer-funded intellectual property and research for its own military and economic gain.

Following this report, I introduced bipartisan legislation, Safeguarding American Innovation Act, which seeks to stop U.S. taxpayer-funded research and Internet Protocol (IP) from falling into the hands of the Communist Party of China (CCP).

Two months ago I issued a new report detailing China's efforts to target influence and undermine the United States and Federal Reserve. China has a pattern of economic and cyber espionage, and social media for them is just another opportunity. I am highly concerned about TikTok and how China may be leveraging their influence to access the platform's data on Americans.

Chinese law requires all companies operating under its jurisdiction to, in essence, allow the Chinese Communist Party to access every piece of data collected. Any company that refuses to comply with the CCP's demand is subject to severe consequences, as are individuals. Therefore, since both TikTok and its parent company, ByteDance, have a presence in Mainland China, an expert witness this morning told that TikTok's user data could be accessed by the Chinese Communist Party. We want to talk more about that today.

This means that the CCP may have access to 100 million Americans' personal and proprietary information. As the U.S. Government has warned, China's access to user data will allow it to extend its malign agenda and build dossier's on American citizens. The overwhelming popularity of this app with America's youth will allow China to collect never-before accessed troves of data on our children, the future generations of Americans.

But the challenges that social media poses to our children are not limited to TikTok. We continue to see the proliferation of child sexual abuse material online. I have been at the forefront of this for years. I am proud that the Stop Enabling Sex Traffickers Act was signed into law in 2018. This was the first bill to reform Section 230, by removing barriers to both criminal prosecution and civil suits against websites that knowingly facilitate online sex trafficking.

Because of this change in law, courts are beginning to affirm that Section 230 cannot shield internet companies when they fail to respond to images of child exploitation and continue to profit from exploitation on their platforms. A specific case against Twitter is now

being considered by the Ninth Circuit Court of Appeals, for example, and will show if the law needs to be expanded in order to properly protect children.

But it is not just Twitter. The fight continues on other platforms that are used to exploit children. Meta announced earlier this year that they would not report all explicit images of children and would instead, and I quote, "err on the side of an adult," end quote, when moderating explicit images of could-be children. In other words, when the age of an individual in a sexual image is uncertain, content moderators are told to put their thumbs on the scale of that individual being an adult.

To me this is shocking. Let us be clear what we are talking about. This is child sexual abuse material, images of a minor's rape, exploitation. Somehow, at least what we have been told, is that Meta has decided that these should not be referred to law enforcement. The National Center for Missing and Exploited Children (NCMEC) has made it clear that images must be reported if they appear to involve a child so that law enforcement can intervene and stop the abuse and prosecute perpetrators.

I worked with colleague across the aisle to draft this legislation, Stop Enabling Sex Traffickers Act (SESTA), and we crafted it narrowly so that it would be focused on ending trafficking and exploitation online. But it may, in fact, be too narrow if companies continue to turn away from keeping the exploitation of children off of their platforms. I hope my colleagues will take up the challenge of revisiting SESTA and tightening the standard so that entities showing a reckless disregard for the sexual exploitation of children are held accountable. I am ready to be an ally in this fight, even after I leave the Senate this term.

I look forward to discussing these matters, especially regarding how product development processes appear to be at odds with user safety as well as the need for more detailed transparency from companies, and again, I look forward to the testimony.

Thank you, Mr. Chairman.

Chairman PETERS. Thank you, Ranking Member Portman.

Our first witnesses is Chris Cox, Chief Product Officer at Meta. Mr. Cox joined Meta in 2004, as a software engineer, and has helped build the first versions of signature Facebook features, including the News Feed. He later served as Director of Human Resources (HR), leading the direction and tone of Meta's company culture.

In 2008, he began serving as Vice President of Product, and in this role Mr. Cox built out the initial product management and design teams before being promoted to Chief Product Officer in 2014, and began his role overseeing the family of apps in 2016.

Before I ask you to have your opening comments, we skipped over an important part of the Committee, and that is that it is the practice of this Committee to swear in witnesses. If the four of you would please stand up and raise your right hands.

Do you swear that the testimony that you will give before this Committee will be the truth, the whole truth, and nothing but the truth, so help you, God?

Mr. COX. I do.

Mr. MOHAN. I do.

Ms. PAPPAS. I do.

Mr. SULLIVAN. Yes.

Chairman PETERS. All four answered in the affirmative. Thank you. You may be seated.

Mr. Cox, I have already had your introductions so please proceed with your opening comments.

TESTIMONY OF CHRIS COX,[1] CHIEF PRODUCT OFFICER, META

Mr. COX. Thank you, Chairman Peters, Ranking Member Portman, distinguished Members of the Committee. Thank you for the opportunity to appear here before you today. My name is Chris Cox. I am Meta's Chief Product Officer, overseeing our apps and privacy teams.

I first joined the company in 2005, as one of our first 15 software engineers. I care deeply about the work we do to help people connect with things and the people they care the most about. It is important to us that we help people feel safe on our apps, and we stand firmly against the exploitation of social media by those committed to inciting violence and hate. That is why we prohibit hate speech, terrorism, and other harmful content.

To enforce these rules, we employ tens of thousands of people and we use industry-leading technology. We regularly publish transparency reports so people can see how we are doing over time and how we compare to other internet platforms.

I am proud that we have invested around $5 billion last year alone and have over 40,000 people working on safety and security, more than any other tech company, even adjusted for scale. Our efforts are making a difference. For example, we reduced by more than half the amount of hate speech people see on Facebook over the last 18 months.

People often talk about these types of issues as safety issues, but at Meta we also refer to them as integrity issues. Integrity is our way of referring to the work we do to prevent bad actors from abusing our platforms. This includes working to stop terrorists and violent extremists and also bullying and harassment, scams, and other types of harm.

As the Chief Product Officer I am proud that safety and integrity are key to the product experience. We build products and continually update them with safety and integrity in mind. It is a core part of our ethos, that as we develop products we constantly think about how people are going to use them and work to make sure they can do so safely.

I know you have questions about our algorithms. Like most platforms, Facebook and Instagram use different algorithms for various features. For example, we use algorithms to help keep our community safe by identifying and removing content that violates our policies, including hate speech, incitement, and terrorism. This work often happens before anyone reports content to us, sometimes even at the point of creation. We use algorithms to rank the content that appears in people's feed and search results, to help deliver relevant advertising, and a whole lot more.

[1] The prepared statement of Mr. Cox appears in the Appendix on page 125.

I also want to stress that our goal is to help people see what they find most valuable. It is not to keep people on the service for a particular length of time, and it is certainly not to give people the most provocative or enraging content. In fact, key parts of those systems are designed to do just the opposite. We reduce the distribution of many types of content, including because they may be misleading or are found to be false by independent fact-checking partners.

At the end of the day, our job is to build the best product for people, and that is a product that is reliable, fast, safe, secure, and relevant, a product that connects people to content relevant to their interests and connects them to their family and friends. That is the product that people want, and that is the product we wake up every day trying to build.

We appreciate your attention to these important issues and look forward to continuing to work with you to find ways we can continue to improve our products, our processes, and our partnerships.

Thank you, and I look forward to your questions

Chairman PETERS. Thank you, Mr. Cox.

Our next witness is Neal Mohan, Chief Product Officer at YouTube. In his role, Mr. Mohan is responsible for YouTube products and user experience on all platforms and devices globally, including YouTube's core mobile applications, YouTube, YouTube Kids and Music, YouTube Red, and YouTube TV, as well as other designs, policies, and services.

Previously Mr. Mohan was Senior Vice President of Display and Video Ads at Google, and prior to joining Google Mr. Mohan served as Senior Vice President of Strategy and Product Development at DoubleClick, an advertisement company that developed and provided internet and ad-serving services before its acquisition by Google. In that role, he built the company's strategic plan, led the product management team, and grew the business rapidly.

Mr. Mohan, welcome to the Committee. You may proceed with your opening remarks.

TESTIMONY OF NEAL MOHAN,[1] CHIEF PRODUCT OFFICER, YOUTUBE

Mr. MOHAN. Thank you, Chairman Peters, Ranking Member Portman, and distinguished Members of the Committee. Thank you for the opportunity to appear before you here today. As the Chairman mentioned, my name is Neal Mohan, and I am the Chief Product Officer of YouTube. In my role I am responsible for all of YouTube's products, our user experience, and trust and safety globally.

YouTube's mission is to give everyone a voice and show them the world. Our openness is core to that mission and enables us to help billions of people around the world to learn new skills, discover emerging music and artists, and enjoy videos from their favorite creators.

We are also proud to be a place where creative entrepreneurs can build thriving small businesses. Last year, YouTube's creative ecosystem contributed over $25 billion to the U.S. gross domestic prod-

[1] The prepared statement of Mr. Mohan appears in the Appendix on page 129.

uct (GDP), and we supported more than the full-time equivalent of 425,000 jobs across the country.

Our commitment to openness works hand-in-hand with our responsibility to protect our community from harmful content. Responsibility is central to every product and policy decision we make, and is our No. 1 priority.

To that end, I want to make clear that there is no place on YouTube for violent extremist content. Our policies prohibit content that promotes terrorism, violence, extremism, and hate speech. Not only is the type of content harmful to our community, the overwhelming majority of creators, viewers, and advertisers do not want to be associated with it, meaning it is also bad for our business.

In my testimony today I will provide more information on our approach to responsibility as well as our policies and technology that enable our skilled enforcement efforts to combat terrorist content online.

We have four pillars of responsibility. We call them the Four R's: we Remove content that violates our policies as quickly as possible, we Raise up authoritative sources, we Reduce the spread of content that does not violate our policies but brushes up against our lines, and we Reward trusted creators and artists. My written submission explains each of these Four R's in much more detail.

This framework enables us to uphold our responsibility to the YouTube community and society while preserving the opportunities of an open platform. For us, safety and growth are intertwined. Violative content undermines user trust and satisfaction, deters advertisers from investing in ads on YouTube, and harms the creators that have built businesses on our service.

Let me discuss YouTube's policies prohibiting terrorist, violent, and extremist content. Our Community Guidelines set the rules of the road for content on YouTube. These policies explicitly prohibit terrorist organizations using our services for any purpose, and we routinely remove such material. We rely on a combination of people and technology to enforce these policies. In fact, machine learning is a critical tool in our effort to remove violative content at scale before it is widely viewed.

As a result of our ongoing investments in teams and technology, in the first six months of 2022 we removed close to 8.4 million videos for violating our policies. Over 90 percent of this violative content was first detected by machines, the majority of it removed before receiving ten views.

Our policies are complemented by our work to raise up authoritative content and reduce the spread of content that comes close to but does not quite violate our policies. For news and information topics our systems elevate authoritative sources such as news outlets and public health authorities and search results and Watch Next panels, and what we call borderline content is not widely recommended.

We also share best practices on counterterrorism with our industry peers through the Global Internet Forum to Counter Terrorism (GIFCT), which is dedicated to disrupting terrorist abuse of digital platforms. Responsibility is and will continue to be YouTube's No. 1 priority. Our business literally depends on it.

Thank you, Mr. Chairman, for convening this important hearing. We look forward to continuing to work with you to address these challenges. Thank you.

Chairman PETERS. Thank you, Mr. Mohan.

Our next witness today is Vanessa Pappas, Chief Operating Officer at TikTok. In her role she is responsible for overseeing content, operations, marketing, and product teams. She previously served as interim head of TikTok globally. She also has experience serving as Global Head of Creative Insights at YouTube, where she oversaw YouTube's global creative research and trends, audience development, creative strategy, and growth teams.

Before joining YouTube, Ms. Pappas served as Vice President of Programming and Audience Development at nextnewnetwork, later acquired by YouTube and Google, where she spearheaded business partnerships and audience development efforts.

Ms. Pappas, welcome to the Committee. You may proceed with your opening comments.

TESTIMONY OF VANESSA PAPPAS,[1] CHIEF OPERATING OFFICER, TIKTOK

Ms. PAPPAS. Great. Thank you for having me.

Chairman Peters, Ranking Member Portman, and Members of the Committee, thank you for the opportunity to appear before you today to discuss how TikTok is delivering on our commitment to provide a safe and welcoming experience for our community while also combating some of the real-world harms that are the focus of this Committee's important work.

My name is Vanessa Pappas, and I am the Chief Operating Officer for TikTok. I live in Los Angeles with my family and I have been in the United States for 20 years, and have spent my career in entertainment and media. Prior to joining TikTok, I was an executive at YouTube. I am passionate about creating safe online communities where people can express themselves creatively and discover entertaining and useful content.

At TikTok our focus on safety starts at the top, with a leadership team goal to strengthen safety and build trust, and this focus on safety and security flows through our product decisions.

As a person who believes in the potential of online platforms to create amazing opportunities for individuals, for businesses, and for society, I am personally invested in this goal, and as an executive there is no responsibility greater than protecting the people of our platform.

TikTok's mission is to inspire creativity and bring joy, and more than 1 billion people around the globe enjoy the authentic, entertaining content that TikTok is known for.

We know that with success and growth comes responsibility. We are committed to being an industry leader in safety and security, and earning trust through the transparency of our actions. Let me talk first about safety and security.

At TikTok, creating a safe environment means we make decisions that prioritize the well-being of our community and limit the potential of online polarization or real-world harm, even if those

[1] The prepared statement of Ms. Pappas appears in the Appendix on page 136.

choices come at the expense of short-term commercial success. Our trust and safety teams have an active seat at the product development roadmap and before launch.

Our terms of service and community guidelines are built to help ensure our vision of a safe and authentic experiences. Our policies have zero tolerance for disinformation, violent extremism, and hateful behavior. Enforcement of these policies in the United States is led by our U.S. safety team in Los Angeles, which reports directly to me. TikTok has thousands of people working across safety, privacy, and security, and we invest heavily in technology to detect potential violations or suspicious accounts at scale.

We also work to prevent the spread of harmful content. For instance, with the help of partners, including the U.S. intelligence agencies, we identify groups and individuals in the United States and abroad who promote violent extremism and hateful ideologies, and we work to eliminate that content associated with them. Examples include foreign terrorist organizations, drug cartels, and groups such as Three Percenters and Oath Keepers. Anyone who searches for this content or related hashtags or keywords will instead be redirected to our community guidelines.

Notably, TikTok was not the platform of choice for those who organized the violence at the Capitol on January 6, 2021. Of 686 references in the Department of Justice (DOJ) charging documents, TikTok was mentioned in only 18.

I will next talk about trust and transparency. Trust is a huge component of safety, and it is hard to earn but easy to lose. We hold ourselves to a high standard when it comes to being transparent about our work on safety and security in order to build trust.

You may be familiar with Project Texas, a critical and industry-leading initiative we have been pursuing for over a year. We are making progress toward a final agreement with the U.S. Government to further safeguard U.S. user data and fully address U.S. national security interests. We look forward to finalizing this arrangement and sharing more when we are able.

Since 2019, we have released community guideline enforcement reports which detail the type and volume of the content we remove. For instance, in the first quarter of 2022, more than 95 percent of the time we discovered and removed problematic content before receiving any reports. We also disclose the data on the requests we receive from law enforcement and/or governments.

Finally, our Transparency and Accountability Centers show how we moderate content and recommend content. We would be happy to arrange for a tour for Members and Committee staff as we have for others at your convenience. Last month, we confirmed that our content moderation and recommendation models will be vetted and validated by Oracle. We recognize your questions and concerns and strive to lead the industry in meaningful transparency.

Thank you again for inviting me today. We know that our work in safety and security is never done. These issues are of the utmost importance to TikTok, to our community, and to our industry. We are glad to be a part of a forward-looking conversation such as this one so that we can better work together to address these critical challenges.

I look forward to answering your questions. Thank you.

Chairman PETERS. Thank you, Ms. Pappas.

Our final witness today is Jay Sullivan, General Manager of Bluebird, Twitter's Consumer Products. He concurrently serves as the interim General Manager of Goldbird, Twitter's Revenue Products, and previously served as Vice President of Consumer Product at Twitter.

Prior to joining Twitter, Mr. Sullivan worked at Facebook where he led the development of Realty Labs' AI Assistant, and then led the privacy, integrity, and systems product teams for Messenger and Instagram Direct, launching many user-focused features and improvements.

Mr. Sullivan, welcome to the Committee. You may proceed with your opening comments.

TESTIMONY OF JAY SULLIVAN,[1] GENERAL MANAGER OF BLUEBIRD, TWITTER

Mr. SULLIVAN. Chairman Peters, Ranking Member Portman, and distinguished Members of the Committee, thank you for the opportunity to speak with you today about this important issue. My name is Jay Sullivan. I joined Twitter in November 2021. In April of this year I became General Manager (GM) of Twitter's Consumer Product team. This team is responsible for the main features that people use on Twitter's mobile apps and website. I am also the General Manager of Twitter's Revenue Products team.

Twitter's purpose as a company is to serve the public conversation. The open nature of our service gives a voice to a world of diverse people, perspective, ideas, and information. We believe that Twitter is a force for good in the world.

In the past year we have seen people come to Twitter to get on-the-ground information about the conflict in Ukraine, to access life-saving information during natural disasters, and to exchange ideas about diverse topics ranging from news to culture to sports. The goal of the Consumer Product team is to increase healthy participation in the public conversation. We measure our success by how many people use Twitter and the health and safety of the platform. These two priorities go hand in hand. If people do not feel protected from hate, abuse, and harassment they will simply leave the service.

By the same token, advertisers do not want their brands, products, or services to appear anywhere near harmful content. They will simply pull their ads. This is why it is not in our interest to have harmful content on our platform, and this is why we set out to build features that promote a healthy conversation, and why we will pause, delay, or stop a product rollout if we have health or safety concerns.

My written testimony outlines many steps that Twitter has taken to safeguard our service and improve health on our platform. I would like to use this time to explain my team's overall approach to health, which is built on three main pillars.

The first pillar is integrating health and safety considerations into the product design and development process. We proactively

[1] The prepared statement of Mr. Sullivan appears in the Appendix on page 154.

and methodically assess risk and potential unintended consequences before we begin development of a new feature and through the development process.

We also develop new features that incentivize healthy discourse. Some recent examples are the development of prompts that encourage people to read articles before sharing them; interstitial labels that provide context; Birdwatch, a community-powered annotation feature; Twitter Circle; and many more health and safety features.

But we cannot always prevent bad behavior so we also build tools that enable us to identify and take action on harmful content, including machine learning software to help detect it. These tools enhance the customer experience by decreasing the burden on individuals to report content for review, and they improve the platform overall.

The second pillar is developing and enforcing policies designed around health and safety. We have a team responsible for developing the Twitter Rules, the policies and governance frameworks that prevent and mitigate harm to the people who use Twitter. This team does not report to me but we work closely together.

Twitter's policies prohibit terrorist and other violent organizations on our platform, inciting violence, harassment targeted at individuals or groups, and hateful conduct. Our platform integrity and authenticity policies address efforts to spread misinformation relating to civic integrity, moments of crisis, COVID, and synthetic and manipulated media.

The third pillar is transparency and accountability. We directly engage with outside experts, formal public feedback processes, and research. For example, in 2018, we were the first in the industry to release an archive of potential foreign influence operations identified on Twitter, enabling a host of research on this important issue. This effort has now evolved into the Twitter Moderation Research Consortium. We also provide industry-leading research access to the Twitter application programming interface (API).

Underpinning all of this work are our culture, our processes, and our technology. I, and other senior leaders at the company, work to encourage and empower every employee to contribute to this shared goal.

I look forward to the discussion today. Thank you.

Chairman PETERS. Thank you, Mr. Sullivan.

Mr. Cox and Mr. Mohan, the question is going to be directed to you. The dangerous QAnon conspiracy that started in 2017, spread unchecked on your platforms for years before you started to downrank and then ban them on your platforms. A February 2022 poll found that 16 percent of Americans now believe in this conspiracy theory.

We heard this morning that your algorithms push sensational content. Mark Zuckerberg, in fact, said in 2018, that this is, quote, "basic incentive problem," end of quote, and he goes on to say that, "when left unchecked people will engage disproportionately with more sensationalist and provocative content," from Mr. Zuckerberg.

My question is, if your recommendation algorithms are focusing on engagement—we understand the business reasons for doing that—and provocative content that increases that engagement—because that is certainly the most engaging—is it inevitable that they

will promote extreme content that you have not yet labeled violative, like QAnon?

Mr. Cox, we will start with you, and then Mr. Mohan.

Mr. COX. Thank you, Senator. To start with QAnon, this is an organization that today is labeled as a violence-inducing network and so is not allowed on our platform. In general, we believe there is no place for terrorism, for violence-inducing content, for extremism across the network. We work hard to take that content down, and as we have talked about, we publish our results and work with law enforcement to make sure that we can do so quickly.

Chairman PETERS. Mr. Mohan.

Mr. MOHAN. Senator, thank you for the question. QAnon on YouTube is deemed to be a harmful, criminal conspiracy. We do not allow that content on our platform. We have been removing that type of content from our platform for years, given the nature of potential incitement to violence.

But we do not just stop there. Not only do we remove the content because there is no place for hate, harassment, violent extremism, or graphic violence of any kind on YouTube, but we also make sure that when users are looking for information on our platform around a particular topic, news topic, what have you, we raise up content that comes from authoritative sources, that includes typically mainstream media outlets, et cetera, that can put that particular news story in context.

We have a combination of a number of tools, those Four R's that I alluded to before, that work comprehensively to make sure that this type of content has no home on YouTube.

Chairman PETERS. I appreciate that and I wanted to hear your response from both of you. I appreciate it has no home and you are aware of that now. The intro to the question was that it started to spread in 2017. This stuff was on your platforms for years. It took you a long time to come to the conclusion that both of you have just come to.

But I want to get back to, really, the question that I have, is that we have a quote from Mr. Zuckerberg. We understand the business model, although I think Mr. Cox said that it is not necessarily to keep people on platforms. It is to engage. The more people that are engaged in your platforms, the better from a business perspective. You will be able to serve up more ads for folks, generate more revenue. As Mr. Zuckerberg said, people engage disproportionately in sensational and provocative content.

That content is actually good for your business. If more sensational, provocative content is put forward, get people to stay on the platform longer, you are going to be able to show them more ads.

My question is, is that not inevitable that that is going to happen when you continue to put this content? I appreciate after the fact, and we are going to talk about your model up front to try to prevent some of this stuff from happening at the front end. I appreciate at the back end that you are going to take some action. How many people saw the QAnon false or the conspiracy theory? I said 16 percent of the American people now think this conspiracy theory is real. You caught it, but not until 16 percent of the American people are part of this insidious theory.

Tell me about the up front. Why are you not engaged up front before you launch products, to understand and perhaps anticipate how things could be misused?

I will start with you, Mr. Mohan, this time, and then Mr. Cox.

Mr. MOHAN. Yes, Senator. I appreciate the opportunity to clarify what I think might be a potential misconception in terms of our business model and what our incentives are.

To be very clear, we have no incentive to post this content, to promote it in any nature, and that goes to the fundamental nature of what YouTube's business actually is. We are fundamentally an advertising platform. We generate revenue through advertising partners. We share that revenue, the majority of that revenue, with our creators.

When we talk about the creator economy, all those over 400,000 jobs that we have created in this country, it is through that business model that generates revenue on behalf of our creators. Our advertisers have told us, in no uncertain terms, that they do not want to be associated with content that promotes hate, violent extremism of any sort, terrorism, or what have you. I have firsthand experience myself, over the years. When they feel that sort of content is on our platform, they walk away.

We have not just a moral imperative—that is my top priority, living up to our responsibility—but also it aligns with our business goals.

Chairman PETERS. I appreciate that, and my question is how long it takes to be able to identify. This was up for two-plus years before the changes there.

Mr. Cox, Facebook currently has an advertising campaign touting that you have spent $16 billion over the last six years on safety and security. I do not think it is the amount of money that is spent. It is about the results that are most important. You are a very large company, and, in fact, I think over the last six years a revenue close to $450 billion, a massive amount of money, a relatively small amount relative to the total revenues of your company. In fact, I think it is equal to basically what you have spent is $1 per user per year for the entire globe. In comparison, Meta spent over $85 billion on stock buybacks over the last six years, considerably more than you pay on trust and safety.

Why is your company willing to spend so much more per year to drive up a stock price but not willing to spend the money necessary to be able to pull down this dangerous content a whole lot quicker and perhaps actually be forward-leaning and design products from the get-go to eliminate the abuse of these platforms?

Mr. COX. Thank you, Senator. This is an issue for the whole company, not just for the safety teams or for the specific investments that we talked about there. I expect every engineer, product manager, designer, researcher, data scientist, whether it is building a new product or whether it iterating upon an existing product, to pay attention to safety. That is something that is built into the deoxyribonucleic acid (DNA) of the company. It is something I personally care very deeply about. It is something that we expect folks who are designing products to think about as a part of their work.

In addition to the specific investment of 40,000 folks who work directly on safety and security for the company, it is part of the culture of how everybody at the company thinks about their work.

Chairman PETERS. I am going to turn it over to the Ranking Member because we have a number of Members here, but I am going to drill down a little further on that comment in a further round.

Ranking Member Portman, you are recognized for your questions.

Senator PORTMAN. Thank you, Mr. Chairman. I look forward to getting into this issue of the balance between free speech and the hate speech that leads to violence because it is a line that has to be drawn. I know it is not easy, but I am going to talk about one that I think is easier, and that is child sexual exploitation. I talked about it my opening statement a little bit.

We all know that the threat of this sexual abuse material is a persistent threat. In fact, we note that last year over 29 million reports came in of child sexual exploitation. That was a 35 percent increase from just 2020, so this is an increasing problem across the board, but particularly with regard to our kids.

That is why I thought it was so unfortunate, Mr. Cox, when I learned about the Meta policy directing content moderators, and I quoted this earlier, but it is to "err on the side of the person involved in sexual exploitation being an adult" when they are unsure about the age of the person. Let me give you a chance to respond to that. This has been in the public media. It does not mean that it is true, I suppose, but is that truly what you directed your content moderators to do?

Mr. COX. Senator, I know that, first of all, this is an incredibly serious issue and I appreciate your work on this issue. As the father of two kids, this is something I personally care about, making sure that we pay attention to as well.

The work that we do here is in consultation with NCMEC. We have been the most aggressive of the tech companies there. We have referred more content to them, I believe, than all the other tech platforms combined. That is both through the work we do on WhatsApp and Messenger as well as across the family of apps.

My understanding on this specific question is that we received direction from NCMEC to prioritize known Cyber Security Assessment & Management (CSAM) content, which was the nudge that they gave us and where they wanted us to focus our time. I have not been focused on that specific conversation and I would be happy to have the team follow up.

Senator PORTMAN. Yes. Let me make sure I understand this. You are blaming the National Center for Missing and Exploited Children for changing your approach of moderators saying that we are going to assume that kid are adults if we do not know? NCMEC has said you have a responsibility, all of you do, to report all images that appear to involve a child so that law enforcement can intervene to stop the abuse and prosecute the perpetrators, period. I cannot believe that you are saying that NCMEC would want you guys to send out instructions to your moderators saying err on the side of this being an adult if you are not sure.

Did I misunderstand what you said?

Mr. COX. Senator, I have not been in that specific conversation with NCMEC, but I would be happy to follow up on the details. I agree it is a very important issue.

Senator PORTMAN. Given your role, would you commit to, one, getting back to me on it, and two, ensuring that if that is true that you change that policy?

Mr. COX. Senator, I could commit to getting into the details of the policy and make sure we follow up with the team to work on it.

Senator PORTMAN. OK. You are the Chief Product Officer. I would hope that this is one that you would follow up on and ensure it is not the direction you are giving your moderators, because that is what has been publicly reported.

With TikTok, we talked about this earlier, again in the opening statement, nearly half of American kids use TikTok, as you know. That is your audience. There are a lot of risks there to privacy and national security, in my view. Ms. Pappas, I understand that TikTok is subject to the laws of the United States but it is also subject to the laws of other countries in which it operates—United Kingdom, Germany.

But with regard to China, is it true, yes or no, does TikTok have an office and employees in Beijing?

Ms. PAPPAS. I think this is another one for clarification——

Senator PORTMAN. Just yes or no.

Ms. PAPPAS [continuing]. Of which TikTok does not operate in China. You are right in saying that TikTok is subject to the laws in the United States, as we are incorporated in the United States and California.

Senator PORTMAN. Do you have employees in Beijing?

Ms. PAPPAS. Yes, we do, as do many global tech companies, including those——

Senator PORTMAN. I was asking you, do you have an office in Beijing?

Ms. PAPPAS. Yes.

Senator PORTMAN. OK. Is your parent company ByteDance headquartered in China?

Ms. PAPPAS. No, they are not.

Senator PORTMAN. ByteDance is not headquartered in China?

Ms. PAPPAS. No. ByteDance is founded in China but we do not have an official headquarters. It is a global company.

Senator PORTMAN. Where is the headquarters of ByteDance?

Ms. PAPPAS. We are a distributed company. We have offices around the world. Our leadership team is largely in Singapore, but we do have an official headquarters.

Senator PORTMAN. You have to be headquartered somewhere, and I think it is in the Cayman Islands. Is that correct?

Ms. PAPPAS. The parent company was incorporated in the Cayman Islands. That is correct.

Senator PORTMAN. OK. You are headquartered somewhere, and it is the Cayman Islands, but you have a presence in China, and of course, you comply with Chinese law with regard to your people presence in China. Correct?

Ms. PAPPAS. That is not correct. Again, TikTok does not operate in China. The app is not available. As it relates to our compliance

with law, given we are incorporated in the United States we comply with local law.

Senator PORTMAN. Yes. Do you believe that the Chinese Communist Party has the right to access data collected by your company because you have a presence in China?

Ms. PAPPAS. Sorry, again, Senator Portman. TikTok, the app, is not available in China.

Senator PORTMAN. No. You said you have an office in Beijing and you have employees in Beijing. That is a presence.

Ms. PAPPAS. Yes, so as we have said on the record, we do have employees based in China. We also have very strict access controls around the type of data that they can access and where that data is stored, which is here in the United States. We have also said under no circumstances would we give that data to China.

Senator PORTMAN. Yes. I am glad that you say that. It does not seem to square with what we know about the Chinese national security law, but I appreciate that approach. U.S. military banned their own servicemembers from using TikTok for this reason, as you know, and last month the House of Representatives warned lawmakers of the risk of using TikTok. These are Members of Congress were told not to use it. Our military was told not to use it out of concern for the user's privacy and national security.

Do you think those decisions were wrong?

Ms. PAPPAS. I would not opine on the needs for an entertainment platform on Federal devices, but I would say that TikTok is an entertainment platform first and foremost, and this is part of the joy that we bring to millions of people around the world. We are very much committed to the security of our U.S. users and citizens, which is why we are investing so heavily in this area.

Senator PORTMAN. According to a leaked audio obtained by Buzzfeed news, which I am sure you saw, there are TikTok and ByteDance employees in China who can gain access to U.S. user data, so this Committee will now be looking into the assurance of what you said, that TikTok would not give U.S. data to China. Do you have any response to the Buzzfeed news story?

Ms. PAPPAS. Yes. Those allegations were not found. There was talk of a master account which does not exist at our company, period.

Senator PORTMAN. Yes. Will TikTok commit to cutting off all data and metaflows to China, Chinese-based TikTok employees, ByteDance employees, or any other party located in China that might have the capability to access information on U.S. users?

Ms. PAPPAS. Again, we take this incredibly seriously in terms of upholding the trust with U.S. citizens and ensuring the safety of U.S. user data. As it relates to access and controls, we are going to be going above and beyond in leading initiative efforts with our partner, Oracle, and also to the satisfaction of the U.S. Government through our work with Committee on Foreign Investment in the United States (CFIUS), which we do hope to share more information on.

Senator PORTMAN. Can you make the commitment, though, that I just asked you to make, that you will commit to cutting off all data and metadata flows to China, Chinese-based TikTok employees, ByteDance employees, or any other party located in China?

Ms. PAPPAS. What I can commit to is that our final agreement with the U.S. Government will satisfy all national security concerns, yes.

Senator PORTMAN. But you will not make a commitment to agree to what I have now twice asked you about?

Ms. PAPPAS. Sorry. Given the confidentiality of CFIUS I am not able to talk specifically about that agreement.

Senator PORTMAN. Forget CFIUS. I am not talking about CFIUS.

Ms. PAPPAS. I am happy to share more——

Senator PORTMAN. I am asking whether you would make the commitment today. Will you make that commitment?

Ms. PAPPAS. I am committing to what I have stated, which is we are working with the United States government on a resolve through the CFIUS process in which we will continue to minimize that data, as well as working with Oracle to protect that data in the United States.

Senator PORTMAN. This is part of the United States government too. This is our oversight function, and——

Ms. PAPPAS. I appreciate that.

Senator PORTMAN [continuing]. I am concerned that you are not able to answer the question except to say that you will not make the commitment to cutting off this data to China. We think that all data collected relating to Americans and then accessed in China is a problem. We think it should be safe from exploitation by the Chinese Communist Party. If the data is accessible in China, as you have testified, then it could be exploited. That concerns us.

I have gone over my time. I apologize, Mr. Chairman, but I thought it was important to get the answers.

Chairman PETERS. Thank you, Ranking Member Portman.

Senator Carper, you are recognized for your questions.

OPENING STATEMENT OF SENATOR CARPER

Senator CARPER. Thanks very much. A warm welcome to our witnesses. Mr. Mohan, thank you very much for spending time with my staff and me earlier this week. We appreciated that.

I am a former Governor of Delaware, and the last Vietnam veteran serving in the U.S. Senate. Not everybody knows this but the first State to ratify the Constitution was Delaware. Before anybody else did, we did, and for one whole week Delaware was the entire United States of America. Those times were a little less complex than they are today.

But I thought throughout the many years I have now lived in Delaware a lot about our democracy, the formation of our country and the formation of our government and how we have rolled with the punches over many years. I never imagined how fragile our democracy could really be.

Right after the Founding Fathers—they used to work on the Constitution up in Philadelphia—Ben Franklin was leaving Independence Hall, as I recall, and he was asked by someone, a passerby, who said, "What have you wrought?" he responded, "A republic, if we can keep it."

Churchill had his sense. He described democracy as "the worst form of government devised by word of man except for all the rest."

It is certainly a hard way to govern, and we live it and feel it every day, and see it every day.

Jefferson, who wrote the Declaration of Independence, largely, as you know, also wrote these words, and you already said this, "If people know the truth they will not make a mistake."

I think one of our challenges today, as we try hard to reserve our democracy, is people do not really know what the truth is, and they are not sure. I am not sure how to put that genie back in the bottle, but we need to certainly try.

I have a question if I could. I would like to start off with Mr. Cox. Mr. Cox, again we thank you for joining us today. In your testimony you mentioned that Meta has identified more than, I think, 1,000 militarized social movements and 270 white supremacist organizations on your platform, and removed some 2.3 million pieces of content from Facebook that are tied to organized hate. These statistics, which are deeply disturbing, are only from the second quarter of this, 2022.

While I am glad that they all are tracking and removing harmful content, these statistics are indicative of a troubling trend of bad actors using social media to organize and mobilize their followers.

To that end, Mr. Cox, what has Meta done to address the larger threat of these various groups or organizations and the content that they share? What more can and should be done?

Mr. COX. Thank you, Senator. As I mentioned, we have community standards which outline that there is no place for terrorism, for militarized social movements, for violence-inducing conspiracy networks across our family of apps. We have 350 experts, folks who work with law enforcement to identify terrorist organizations, to identify for violence-inducing conspiracy networks (VICNs), to make sure that we have up-to-date information that we share with law enforcement in order to understand, on a real-time basis, which of these networks to prioritize and pay attention to.

We publish our results quarterly. We publish a transparency report that outlines the prevalence of various categories of bad content. In case it is useful, around 2 in 10,000 pieces of content, 0.02 percent, is the prevalence of hate speech on the platform as of the most recent report from this last quarter. That is down from 50 percent 18 months ago. Each quarter we have released the report over the last several years we have been able to improve on prevalence, which is a sign that our AI systems, our processes, our human systems, et cetera, are improving. We believe that is the most important thing, in addition to having a transparent report on how we are actually doing on these numbers, so that outside experts, so that law enforcement, et cetera, can evaluate along with us.

Senator CARPER. Thank you for that. I like to say, if it is not perfect, make it better, so keep working on it.

A question, if I could, for Mr. Sullivan, on misinformation on Twitter. You stated, I believe, that Twitter makes it clear in its guidelines that the promotion of disinformation is against your platform's policies. We have seen numerous examples, instances if you will, of users sharing this disinformation during the coronavirus pandemic, previous U.S. elections, and the January 6th insurrection right here at this Capitol, just to name a few.

Could you take a moment please and explain for us what policy changes Twitter has made in light of the rapid spread of false information and how it has been effective?

Mr. SULLIVAN. Yes. Thank you for the question, Senator, and I appreciate the historical context. It adds gravity to what we are talking about today.

Senator CARPER. Harry Truman once said, "The only thing new in the world is the history we forgot or never learned." It is pretty good. It is timely.

Mr. SULLIVAN. These are serious matters. We have policies against COVID misinformation that have been evolving as we learn from what is happening in the world, and the same for the spread of misinformation. After our election work has been ongoing, but most recently we have been beefing up all of our policies against the spread of misinformation.

Then we have tried to be more proactive as well. For example, with what we call interstitials and prompts that give positive, valid, truthful information about things like voting, where you can vote, when election ballots will be counted, and things like that, so that people not only debunk false information but can receive vetted information in a way that feels more authoritative, so they know what is real and what is not real.

There is always more to do, but our policies are always evolving, and our software is always evolving to catch these things earlier so that people do not have to report them. We want to catch them before a person needs to even see it. Also adding these prompts and other user interface elements to prevent the spread of harmful information. For example, not being able to retweet something or tweet something that is up against that line.

We are continuing to evolve the product every day. We acknowledge these are critical, societal issues, as you have said. Thank you.

Senator CARPER. If it is not perfect, make it better.

My time has expired. I just want to say, Mr. Chairman and to our Ranking Member, this is a valuable and invaluable hearing, a timely hearing, and I just applaud you holding it and express our thanks to our witnesses for being here. Thank you.

Chairman PETERS. Thank you, Senator Carper.

Senator Johnson, you are recognized for your questions.

OPENING STATEMENT OF SENATOR JOHNSON

Senator JOHNSON. Thank you, Mr. Chairman. Again, I think based on the morning panel and this panel as well we are talking about a problem that is my true definition of a problem, one that does not have easy solutions.

Again, what is harmful content is all in the eye of the beholder. We all abhor violence. I certainly condemn white supremacists. I condemn immediately, forcefully, repeatedly, the violence on January 6th.

But I am concerned about the bias of your platforms. The earlier panel would not give me any indication whatsoever, the percentage of liberals versus conservative in your organizations. I do not expect you will be any more forthright in that.

But let me ask you this question. I know the Chairman likes to talk about the white supremacists and January 6th. Democrats

love talking about that. What about the 570-plus riots that occurred in the summer of 2020, 2,000 law enforcement officers injured, $1 and $2 billion worth of property damage, a couple of people killed in Kenosha, Wisconsin, dozens of buildings burned, a couple dozen people also lost their lives during those riots.

Mr. Cox, have you de-platformed, have you throttled back or censored anybody that was involved in the organization of the summer riots?

Mr. COX. Senator, these were——

Senator JOHNSON. Give me a pretty quick answer. I have a lot of territory to cover.

Mr. COX. Senator, domestic terrorism and extremism and calls for violence are against our community——

Senator JOHNSON. Did you censor anybody that organized the summer riots, 570 of them?

Mr. COX. Senator, I can look at the specifics——

Senator JOHNSON. Mr. Mohan, did you YouTube throttle back or censor any of the rioters in the summer of 2020?

Mr. MOHAN. Senator, our policies apply equally regardless of who the——

Senator JOHNSON. Did you throttle back? Did you censor some of the organizations that were responsible for the summer riots?

Mr. MOHAN. We would have applied our policies equally, regardless of where the riots were, if it was a violation of——

Senator JOHNSON. Can you provide me with the names of people you throttled back that were responsible for the summer riots?

Mr. MOHAN. Senator, I am happy to follow up on that.

Senator JOHNSON. Good. I appreciate that.

Mr. Sullivan, what about Twitter?

Mr. SULLIVAN. Twitter would have removed any incitement to violence that was on our platform, regardless of——

Senator JOHNSON. Did you throttle back on organizers of the summer riots?

Mr. SULLIVAN. For any case where we see incitement to violence we would remove that content and take action.

Senator JOHNSON. Your CEO, Mr. Dorsey, was before the Commerce Committee, I think it was in October 2020, and both Senator Cruz and I asked him whether your platform, Twitter, could impact our elections, and he denied it. Our three witnesses earlier this morning completely disagreed with Mr. Dorsey. They said absolutely Twitter and these platforms can.

Do you believe that Twitter can influence, in fact, our elections?

Mr. SULLIVAN. I think Twitter plays an important role in the public conversation.

Senator JOHNSON. It is really kind of a yes-or-no answer. Can they impact our elections?

Mr. SULLIVAN. We are taking the——

Senator JOHNSON. Really, yes or no. I have a lot to cover.

Mr. SULLIVAN. As I described to the previous question, we have put in place many actions and mitigations relating to elections.

Senator JOHNSON. Can you impact the elections?

Mr. SULLIVAN. People try to use our platform to get messages out regarding elections, and we are doing our best to——

Senator JOHNSON. This morning I simply talked about the fact that you censored the New York Post article about Hunter Biden. We have polls that said that had the American public known that we would not be in the ditch that we are right now.

Mr. Mohan, do you believe YouTube can impact the elections?

Mr. MOHAN. Senator, there is an open public debate that happens on YouTube every day, whether it is——

Senator JOHNSON. It is really a pretty simple yes-or-no answer.

Mr. MOHAN. Senator, YouTube is an open platform where there is debate——

Senator JOHNSON. Mr. Cox, do you believe that Facebook can impact the elections?

Mr. COX. Senator, I think the public discussion that happens on our platform are a part of the public discourse.

Senator JOHNSON. You do not think by censoring material that specifically your management can impact the elections? Not what is occurring on your platform, I mean, but your management, your decisions, your censoring of information. Can you impact the elections?

Mr. COX. Senator, respectfully, with respect to the New York Post story you reference we did not——

Senator JOHNSON. I know you throttled it back. But, again, I am not asking about that specifically. Do you have the power? Our earlier witness, Mr. Roetter, said, "A small group of people run these companies and have substantial power over shaping the reality for billions of people." I mean, can you just be honest?

By the way, our earlier panel basically, to paraphrase, said do not believe a word you guys are going to tell us. But can you at least be honest with the American public and say, yes, you had that power. You can impact the elections. Can you be honest with them or are you going to sit there and say, "People talk about things on our platforms."

Mr. COX. Senator, we do think transparency about our decision-making around our content, around not just terrorism but around misinformation, we take all of these areas of content seriously and we publish our work. We do think the public deserves to know what our policies are and how we enforce them.

Senator JOHNSON. We are going to have another round, it sounds like. I have a lot more ground to cover. But let me just start, and I think Senator Hawley did a great job of talking about how Federal health agencies were in direct communication with, in particular, I think it was Facebook, possibly Twitter. We have also heard that, of course, Mr. Zuckerberg, or Facebook, was contacted by the FBI as it relates to Russian disinformation, and we covered that in the morning.

When it comes to how miserably we failed handling COVID, I think one of the problems was the lack of robust information using this marvelous device we call the internet. I mean, doctors could have been testing out different theories of the case and sharing that experience, but they were shut down. They were censored.

I want to ask each one of you. I am 67 years old. As long as I have been alive I have always been told if you have a serious medical condition you really ought to seek a second opinion because nobody has perfect information. I wish that there would have been

some modesty exhibited by our Federal health authorities. Quite honestly, your platforms, in acknowledging the fact that we do not have perfect information, maybe we ought to let some information flourish. I mean people were censored, eminently qualified doctors who has the courage and compassion to treat COVID patients with cheap, generic, widely available drugs.

I just want to ask the question, Mr. Cox, do you believe that people ought to get a second opinion when it comes to complex medical conditions?

Mr. COX. Senator, we absolutely believe that building a product where people have the ability to express their point of view is critical to what we do. It is critical to what people expect from the product and what people expect.

Senator JOHNSON [continuing]. Can you just answer the question?

Mr. Mohan, do you believe you ought to get a second opinion when it comes to complex medical conditions?

Mr. MOHAN. Senator, as we all recall, when the pandemic started it was an unprecedented event in history where science was being created——

Senator JOHNSON. Maybe you ought to seek a second opinion. I mean, do you think it is a good idea to get a second opinion, or do you only go to one authority and put all your faith in one authority? No other opinion is going to be valid. Is that your belief?

Mr. MOHAN. Senator, we worked with a wide variety of health authorities in this country and all over the world.

Senator JOHNSON. OK. Mr. Sullivan, do you believe you ought to go get a second opinion? If you get diagnosed with cancer today are you going to rely on just one authority?

Mr. SULLIVAN. As a patient that sounds like common sense. Our COVID information——

Senator JOHNSON. Are we not 330 million patients here?

But we were not allowed a second opinion, were we?

We were not allowed by your platforms for that second opinion, and I think hundreds of thousands of people lost their lives because you did not allow a second opinion to be published on your platforms.

Mr. SULLIVAN. Our COVID misinformation policy only——

Senator JOHNSON. It was highly flawed, and I will point that out in the second round of questions. Thank you.

Mr. SULLIVAN [continuing]. It only looked at information that was demonstrably and widely believed to be true.

Senator JOHNSON. I need a second round of questions and I can point that out, that it was not demonstrably false.

Chairman PETERS. There will be a second round, Senator Johnson.

Senator Sinema, you are recognized for your questions.

OPENING STATEMENT OF SENATOR SINEMA

Senator SINEMA. Thank you, Mr. Chairman, and thank you to our witnesses for joining us today.

Every day cartels post on social media platforms and recruit teenagers in Arizona to act as drivers for illegal operations. Lured by the promise of easy cash, these teens, some as young as 14

years old, take their parents' cars to the border and participate in smuggling and trafficking. Innocent bystander and migrants have even died while these teens, recruited by cartels on social media, flee law enforcement at high speeds.

The Department of Homeland Security (DHS) must do more to crack down on dangerous cartels' use of social media, secure the border, and keep Arizona families safe.

My first question is for Ms. Pappas. According to my office's conversations with Border Patrol leadership in Arizona, TikTok is the platform that cartels use most frequently to recruit Arizona teens. What steps does TikTok currently take to ensure its algorithms do not promote cartel-sponsored content, and can you tell me why have those efforts not been more effective?

Ms. PAPPAS. Certainly, Senator. It starts for us with our policies. Obviously that type of content, any illicit activity, organized crime, including drug cartels, is strictly prohibited from our platform.

In that regard we work with our trust and safety moderation teams to ensure that we are detecting that content through our technologies and also through human moderation, to remove that content when found.

Furthermore, as a platform, we do not have the same product features that other platforms do in terms of being able to have that type of organized behavior. For example, we do not allow links through direct messages or images to be sent through direct messages. We also do not have group chat available.

Those types of behavior to help organized crime are limited in terms of TikTok's platform. Any of the content that gets posted on TikTok has to go through our content moderation. Obviously, our work is never done there, but we are constantly working to identify that content at scale and remove it when found, and all of those numbers are also available through our transparency reports.

Senator SINEMA. A follow-up question then. If you are doing that content moderation and reviewing each of those posts, how is it that there are so many efforts that are successful on TikTok to recruit young teens to assist cartels?

Ms. PAPPAS. We are striving to get that number to zero. Obviously, this is a challenging area for everybody in the industry in terms of being able to moderate our platforms, but this is something that we heavily invest in from a technology perspective as well as a people perspective. I am happy to look into any of those cases. But I do know that when reports have been sent, that content is immediately taken down.

Senator SINEMA. Thank you. Mr. Cox, my next question is for you. What is Meta doing, both on Facebook and on Instagram, to prevent cartels from using your platforms to recruit teens along our Southwestern Border?

Mr. COX. Thank you, Senator. This is an important issue. It is sad. I really appreciate your leadership on this issue. We prohibit human trafficking. We prohibit these cartels. We work with law enforcement to identify the names of the cartels, and then we fan our systems to help us find instances of them across our platforms and take them down right away.

Senator SINEMA. My next question is for Ms. Pappas, Mr. Cox, Mr. Mohan, and Mr. Sullivan, the whole panel. As Chair of the

Border Subcommittee in this Committee I believe it is critical that each of your platforms work with the Department of Homeland Security to identify cartel content and prevent Arizona teens from being targeted for recruitment. I would ask you to answer yes or no. When you discover that cartels are using your platforms to recruit are you willing to commit to sharing that information with the Department of Homeland Security as quickly as possible.

Mr. Sullivan.

Mr. SULLIVAN. Yes, with the appropriate privacy and oversight I believe we could do that.

Ms. PAPPAS. Yes. Similarly, following legal process and privacy policy.

Mr. MOHAN. Yes, Senator. We would cooperate as long as there is a due legal process with the DHS and other law enforcement as well.

Mr. COX. Similarly, we would commit to that provided privacy and legal concerns were addressed.

Senator SINEMA. Thank you. Back to you, Ms. Pappas. Today's hearing is about product development, and in the case of TikTok there is no product more important than the, "for you" algorithm that offers content recommendation to users. There is a real risk that TikTok could alter its algorithm to promote or censor content on Beijing's behalf, whether that means silencing voices that are critical of China or promoting conspiracies or extremist content.

Has TikTok ever altered its algorithm or promoted or downranked content based on the actual or perceived wishes of the Chinese government?

Ms. PAPPAS. No.

Senator SINEMA. In your privacy policy it says that TikTok, "may collect biometric identifiers such as face prints and voice prints." Has the biometric data of an American ever been accessed by or provided to any person located in China, and if not, is biometric data able to be accessed by anyone in China?

Ms. PAPPAS. Let me clarify because I think biometrics is one that is a topic that is hard to define and everybody has their own definition of what biometrics means. I will be clear in how TikTok sees this.

We do not use any sort of facial, voice, or body recognition that would identify an individual. There is no way that we would be able to identify. The way that we use facial recognition, for example, would be if we are putting an effect on the creator's video. You are uploading a video and you wanted to put sunglasses or dog ears on your video, that is when we do facial recognition. All of that information is stored only in your device, and as soon as it is applied to, like that filter is applied and posted, that data is deleted, so we do not have that data.

Senator SINEMA. You are assured that there is no opportunity that during the time between the use of the imprint of the face print or voice print, and the deletion, that there is no ability for anyone other than that device to access or capture that information?

Ms. PAPPAS. That is my understanding, yes. I know it is a technical area, so to the best of my knowledge the data is stored on the devices and deleted immediately once you post your video.

Senator SINEMA. I would like follow-up. Neither of us are experts on this technological issue, but I would like to get some follow-up from those who are.

Ms. PAPPAS. Happy to, yes.

Senator SINEMA. Thank you. Mr. Chair, I see that my time has expired. I have a few more questions that I will submit for the record. Thank you.

Chairman PETERS. Thank you, Senator Sinema.

Senator Padilla, you are recognized for your questions.

OPENING STATEMENT OF SENATOR PADILLA

Senator PADILLA. Thank you, Mr. Chair, and I want to thank you for holding this important hearing today. The companies testifying today offers users an unprecedented ability to access, consumer, and distribute information. Mr. Chair, you are right to focus our attention on how corporate product design and investment choices influence the content that is produced and distributed.

My first question for a couple of you is relative to content moderation, and we have been talking about that throughout the hearing here. Last year, Frances Haugen disclosed that at Facebook 87 percent of all spending combating misinformation on Facebook was spent on English language content, despite the fact that only nine percent of Facebook's users are English speakers. She also disclosed that trust and safety investments for users in countries other than the United States were abysmal.

An audit of Twitter's disinformation and misinformation work, disclosed by Pieter Zatko, who testified just yesterday in Senate Judiciary Committee, found that Twitter's integrity lacked language expertise in the countries it was serving, even though 80 percent of Twitter users are outside the United States.

Your companies make commitments to all of your users who are not just linguistically diverse but culturally diverse as well.

A question first for Mr. Cox. In your testimony you state that you have over 40,000 people working on trust and safety issues. How many of those people focus on non-English language content and how many of them focus on non-U.S. users?

Mr. COX. Sure, Senator. I am happy to take your question. Our safety and security teams are deployed to help our users all around the world. Specifically on the question of misinformation, which you mentioned, we have 80 fact-checkers operating in 60 countries around the world. Those fact-checkers are certified by independent fact-checking organizations.

In the United States we have 11 fact-checkers, six of whom support Spanish language content. We also have partnerships with Univision and Telemundo to connect people with Spanish language authentic information around elections. We also offer an election voting center in Spanish language to Americans to help folks get authoritative information about where to vote, that is tailored to their specific ZIP code.

Senator PADILLA. I appreciate the information you are sharing. Some of it has been in your testimony. I welcome more. More is better. Do you have any idea of the breakdown of the 40,000 people I referenced? First of all, is that number roughly accurate? If it is significantly higher, let me know. If it is significantly lower, let us

know. What I am looking at is the ratio of English versus non-English. Do you have that data?

Mr. COX. Senator, I would be happy to follow up on the specifics of how those 40,000 folks are broken down.

Senator PADILLA. Thank you. I would greatly appreciate that.

Mr. Sullivan, how many members of your trust and safety team have non-English language expertise and focus on issues outside the United States?

Mr. SULLIVAN. Yes, thank you for the question, Senator. As a global company this is important to us. We have about 2,200 people working on content moderation globally. I do not have the exact breakdown but we can take that back to our team.

Senator PADILLA. Please. Ms. Pappas, how large is your trust and safety team and how much does TikTok invest in your non-English users, and I guess non-Western users?

Ms. PAPPAS. I do not have those numbers at hand but I am happy to get back to you on those as well.

Senator PADILLA. OK. Mr. Mohan.

Mr. MOHAN. Senator, we have over 20,000 people that work on content moderation all over the world. We are a global platform, as you know, supporting a couple billion users all over the world, and we endeavor to enforce our policies as well as make sure that our recommendation algorithms work equally well for all speakers, all over the world. We support dozens of languages on our platform in all the countries that we operate.

To give you a couple of more concrete examples, here in the United States our support across all those Four R's I described in my initial testimony are not just about English but other languages as well. For example, in Spanish our policies are enforced. We serve up information panels not just in English but in Spanish. Those relate to optics like elections, how to vote, where to vote, et cetera, COVID-related information, because families are looking for that content not just in English in this country, but we recognize in a number of different languages, including Spanish.

Senator PADILLA. Thank you. I would appreciate more detail and data from all of you.

Speaking of data, as some of you may or may not know, my background is in engineering, so I am a big believer in data-informed and data-driven policymaking. In reviewing your testimony—and I appreciate some of the data that you did provide, especially around dangerous content found in your platforms, whether it is incomplete or desire for additional data, let me just jump into a couple more questions.

Mr. Cox, in your testimony you say that Meta found and removed 95 percent of hate speech content before it was ever reported. Of the remaining five percent, how many users were recommended the content in their news feed? Do you have data along those lines?

Mr. COX. Senator, I can offer data on prevalence, which would be the amount of content that appears across the averages of the content on our platform. For hate speech the prevalence in our last report is 0.02 percent, or two out of every 10,000 pieces of content.

Senator PADILLA. But you see where I am going, right? Ninety-five percent is a good number. The five percent that you did not catch before it was reported, if those recommended, 1, 2, 3, 5 times,

that is one thing. If it is recommended tens of thousands of times or more that is a different dynamic. That is what we are trying to get at. If you do not have the data at your fingertips, a follow-up would be welcome.

Mr. COX. Senator, we would be happy to follow up on that.

Senator PADILLA. Great. Ms. Pappas, in your testimony you say that 88.4 percent of removals under TikTok's violent extremism policy occurred within 24 hours of being posted. Again, a good number but it is not 100. For the other 11.6 percent, do we have a gauge of how long it took to find and resolve those items?

Ms. PAPPAS. No. I would have to get back to you on that, but similarly we look at the prevalence of content authoritarian would be violative, and for violent extremism it is 0.01 percent.

Senator PADILLA. OK. Thank you.

Mr. Chairman, my time is up. Similar to Senator Sinema I will have some additional questions I will submit for the record.

Chairman PETERS. Very good. Thank you, Senator Padilla.

Senator Hawley, you are recognized for your questions.

OPENING STATEMENT OF SENATOR HAWLEY

Senator HAWLEY. Thank you very much, Mr. Chairman. Thanks to all the witnesses for being here. Ms. Pappas, let me start with you.

I have to say it is great to see you here today. I have repeatedly invited your company to testify before Congress. I invited them to testify to the Judiciary Subcommittee on Crime and Terrorism in November 2019. I invited them to testify again in September of the following year. Both times we were stiffed. TikTok told me that they would set up a meeting with the CEO. They did not want to testify in public but they set up a meeting with the CEO after November 2019. They then canceled that meeting.

It nice to see TikTok be willing to answer questions in public. It is a pleasant change. Let us dig into a few things, if we could, specifically about TikTok's links to the Chinese Communist Party.

In response to a letter from some of my colleagues, TikTok claimed earlier this year that the company has never shared data with the Chinese government. Is that correct?

Ms. PAPPAS. That is correct, yes.

Senator HAWLEY. And has never shared data with the Chinese Communist Party. Is that correct?

Ms. PAPPAS. We will never share data, period.

Senator HAWLEY. My question was in the past tense. Has TikTok ever shared data with the Chinese Communist Party?

Ms. PAPPAS. We have never shared data with the Chinese government. Correct.

Senator HAWLEY. With the Chinese Communist Party.

Ms. PAPPAS. Yes, correct.

Senator HAWLEY. Have you ever shared it with members, to members of the Chinese Communist Party?

Ms. PAPPAS. We have said many times, Senator, that we do have Chinese engineers based in Chinese. I do not think there is any platform up here that would be able to speak to what you are talking about as it relates to the political affiliation of an individual. But I am happy to assure you that we are ensuing the access con-

trols around our data as well as the storage of that data in the United States.

Senator HAWLEY. I think you are telling me that there are TikTok employees or ByteDance employees who are members of the Chinese Communist Party. Is that what you are saying?

Ms. PAPPAS. No. I am saying I would not be able to verify that.

Senator HAWLEY. Let me ask you affirmatively. Are there TikTok employees or ByteDance employees who are members of the Chinese Communist Party?

Ms. PAPPAS. Senator, I am saying that nobody that is sitting on this panel could tell you a political affiliation——

Senator HAWLEY. I am not interested in anybody's opinion. I am asking you a factual question. Are there members of the Chinese Communist Party employed by TikTok and ByteDance? Yes or no.

Ms. PAPPAS. I would not be able to tell you the political affiliation——

Senator HAWLEY. You do not know?

Ms. PAPPAS [continuing]. Of any individual. What I can tell you is how much we are investing in the——

Senator HAWLEY. No. Membership in the Chinese Communist Party is not exactly like membership in the Democratic Party. I am looking for an answer. You are telling me you do not know? TikTok does not know.

Ms. PAPPAS. Here is what I can tell you. I can tell you that in our United States and Singapore leadership, there are no CCP members.

Senator HAWLEY. You do know that. But you are telling me that you do not know if there are any members who are employed by TikTok or ByteDance, members of the Chinese Communist Party?

Ms. PAPPAS. Senator, I am happy to share that we are putting access controls——

Senator HAWLEY. That is not my question.

Ms. PAPPAS [continuing]. As well as——

Senator HAWLEY. That is not my question. My question is are there any TikTok employees or ByteDance employees, members of the Chinese Communist Party? Yes or no.

Ms. PAPPAS. Senator, I am saying nobody could sit up here and give you that.

Senator HAWLEY. You are saying you do not know? But you do know your leadership is not but you do not know about your employees. Is that your testimony?

Ms. PAPPAS. I know that everyone who makes a strategic decision at this platform——

Senator HAWLEY. Yes.

Ms. PAPPAS [continuing]. Is not a member of the CCP.

Senator HAWLEY. A strategic decision. OK. It is interesting. It is interesting to me that you are quite confident that anyone who could make a strategic decision—how many people is that?

Ms. PAPPAS. It is our leadership team.

Senator HAWLEY. The number?

Ms. PAPPAS. Again, the leadership team is based in the United States and Singapore. Our CEO is based in Singapore. He is not Chinese. I am happy to go into the efforts that we——

Senator HAWLEY. Would it surprise you to learn that Forbes magazine recently reported that at least 300 current TikTok or ByteDance employees were members of Chinese State media and affiliated with the Chinese Communist Party?

Ms. PAPPAS. Again, we do not look at the political affiliations or cannot speak to individuals, but what I can tell you is that we are protecting the data in the United States.

Senator HAWLEY. But apparently, though, you do look at political affiliation because you are quite willing to sit here and tell me that no one who has strategic input or makes strategic decisions is a member of the Chinese Communist Party. You do know very well, as a matter of fact. You just do not want to answer my other question.

Ms. PAPPAS. We have thousands of people that work at the company so I am not going to vouch on the political affiliation of any particular individual. What I can vouch for——

Senator HAWLEY. Have you seen the videos of Chinese Communist Party members conducting training for TikTok and ByteDance employees?

Ms. PAPPAS. No.

Senator HAWLEY. That is fake?

Ms. PAPPAS. I do not know what you are referring to. But what I can tell you——

Senator HAWLEY. Has that happened?

Ms. PAPPAS [continuing]. Is any decision——

Senator HAWLEY. Has that happened?

Ms. PAPPAS. Sir, I just said that I would not be able to tell you. I have not seen it. I am not sure what you are referring to, but I am happy to follow up. But what I can tell you——

Senator HAWLEY. Wait. I am sorry. Let us go back. Let us see if we can cut through the mumbo-jumbo. I am asking you if the Chinese Communist Party has conducted training sessions ever for employees of ByteDance or TikTok. Yes or no.

Ms. PAPPAS. Not for TikTok. TikTok, the app, does not operate in China.

Senator HAWLEY. You have employees in China and ByteDance has employees in China. Listen, we have been through this song and dance.

Ms. PAPPAS. We have.

Senator HAWLEY. Let us just skip that. I have heard it all before.

Ms. PAPPAS. Senator, I appreciate——

Senator HAWLEY. Answer my question. Yes or no. Have they conducted training for ByteDance employees or TikTok employees.

Ms. PAPPAS. I can speak on behalf of TikTok, and the answer is no.

Senator HAWLEY. No. That is interesting. Do any TikTok employees based in China have access to U.S. user data?

Ms. PAPPAS. As we have publicly said, yes, we have engineers in China, and we are working on the access controls——

Senator HAWLEY. None of them are members of the Chinese Community Party?

Ms. PAPPAS [continuing]. We are working on the access controls to minimize that data access——

Senator HAWLEY. I have heard that, and frankly I do not believe it.

Ms. PAPPAS [continuing]. Working with the United States and through the CFIUS——

Senator HAWLEY. Wait. So your testimony is that you do have TikTok employees based in China who do have access to U.S. user data, but you are confident that none of them are members of the Chinese Communist Party and they never accessed it? Is that your testimony?

Ms. PAPPAS. Anyone who has access to U.S. user data has and does so to perform daily duties, so if it is for the performance of site management, bug handling. But we have strict controls in terms of who and how our data is accessed.

Senator HAWLEY. None of that is accessible to any member of the Chinese Communist Party. Is that your testimony?

Ms. PAPPAS. We believe we have the strictest controls out there——

Senator HAWLEY. That is not my question.

Ms. PAPPAS [continuing]. Actually we are working with Oracle——

Senator HAWLEY. My question is does anyone who has access to user data, are they members of the Chinese Communist Party?

Ms. PAPPAS. I feel like I have answered your question.

Senator HAWLEY. You have not, and I feel like you are avoiding it——

Ms. PAPPAS. No.

Senator HAWLEY [continuing]. At every opportunity.

Let me give you another one, since you are on the record and under oath.

Ms. PAPPAS. Can I be as clear as——

Senator HAWLEY. I would welcome you being clear.

Ms. PAPPAS. Thank you.

Senator HAWLEY. Does any person who has access to U.S. user data, are they members of the Chinese Communist Party? Yes or no.

Ms. PAPPAS. Let me be clear again.

Senator HAWLEY. Yes or no.

Ms. PAPPAS. For our U.S. users, the data is sorted and housed in the United States. We have access controls in place——

Senator HAWLEY. You are not answering my point. Let the record reflect you will not answer my question. Why not?

Ms. PAPPAS. Any of that data, it is overseen by our U.S. led security team.

Senator HAWLEY. That is not my question.

Ms. PAPPAS [continuing]. And monitored daily.

Senator HAWLEY. That is not my question.

Ms. PAPPAS. Furthermore——

Senator HAWLEY. My question is does any employee who has access to U.S. user data, are they members of the Chinese Communist Party? You will not answer that.

Ms. PAPPAS. Again, as a global technology platform there is no other company that could make that assertion either.

Senator HAWLEY. That sounds like a yes to me. I think that is news.

You are familiar, I know, with this Buzzfeed article that says that according to leaked audio at more than 80 internal TikTok meetings, China-based employees at ByteDance have repeatedly accessed non-public data about U.S. TikTok users. "Everything is seen in China," said a member of China's Trust and Safety Department in a September 2021 meeting. In another September meeting a director referred to one Beijing-based engineer as a "Master Admin who has access to everything." These reports show data was accessed far more frequently and recently than previously reported. Your testimony is that this is false?

Ms. PAPPAS. Correct.

Senator HAWLEY. All of this is false.

Ms. PAPPAS. That is correct. Everything that you just stated, there is no such thing as a Master Account.

Senator HAWLEY. That is not what it says. It says that someone is referred to as "Master Admin."

But you are telling me that China-based employees have never accessed non-public data of U.S. TikTok users.

Ms. PAPPAS. No. I have already said on the record that we have Chinese employees who have accessed data.

Senator HAWLEY. That is what this is saying. So you agree?

Ms. PAPPAS. If you want to clarify on each individual statement. I am saying that there are strict access controls around the data that is accessed in the United States. That is overseen by our U.S. led security team. We are working with Oracle.

Senator HAWLEY. That is not what this article says.

Ms. PAPPAS. We disagree with the categorization in that article, wholeheartedly.

Senator HAWLEY. Here is the point. I know there are other Senators who want to ask questions. I think we are going to have a second round. The truth appears to be, besides the fact that we cannot get a straight answer on any of these questions, is that you have hundreds of employees with, it appears, access to U.S. user data, that may very well be members of the Chinese Communist Party. You have no way to assure me that they do not have access to our citizens' data. You will not answer my question in a straightforward way about whether a CCP has ever gained access or not.

I think, for my own point of view, that is a huge security problem.

Ms. PAPPAS. Senator, if I may. We are one of the most highly scrutinized platforms. There have been many cybersecurity experts who have researched our platforms, including Citizen Lab, which is a leading academic research unit based in the University of Toronto, who have said, and I am happy to submit this for the record for the Committee, that, "Our research shows that there is no overt data transmission to the Chinese government by TikTok."

Senator HAWLEY. Overt.

Ms. PAPPAS. "TikTok's features and codes do not pose a threat to national security."

Senator HAWLEY. Wait a minute. Overt data transmission?

Ms. PAPPAS. There are also——

Senator HAWLEY. Ms. Pappas, this is not a hearing for you to testify at will. You are here to answer questions.

Ms. PAPPAS. I am providing you with information.

Senator HAWLEY. No, you are not. You are talking over me, and you are submitting information from—who knows who funds this entity, who knows who is behind it, who knows what it contains? I do not know.

What I do know is you will not give me straight answers to my questions, and the reason, I think, is pretty clear, because your company has a lot to hide. You are a walking security nightmare. For every American who uses this app, I am concerned.

Chairman PETERS. Senator Hawley, thank you.

Senator Ossoff, you are recognized for your questions.

OPENING STATEMENT OF SENATOR OSSOFF

Senator OSSOFF. Thank you, Mr. Chairman, and thank you to our witnesses today.

Mr. Sullivan, in disclosures he has made publicly and to the Congress and in his testimony yesterday, former Twitter employee, Mr. Zatko, alleged that Twitter has made willful misrepresentations to the Federal Trade Commission with respect to its compliance with past regulatory action. Is that true?

Mr. SULLIVAN. I am familiar with the allegations. I would point you to our statements that we made as a company that the company disagrees with much of the allegations. Now it is connected to an ongoing lawsuit, so I am not able to——

Senator OSSOFF. My question to you, Mr. Sullivan, is has Twitter willfully misrepresented facts to the Federal Trade Commission?

Mr. SULLIVAN. I can tell you that Twitter disputes the allegations, is all I can tell you about those particular allegations.

Senator OSSOFF. You cannot tell me definitively, Mr. Sullivan, that Twitter has not willfully misrepresented facts to the FTC.

Mr. SULLIVAN. I would point you to what I just said.

Senator OSSOFF. Noted. You do not deny that Twitter has willfully misrepresented facts to the FTC. Understood.

I want to ask you about the logging of access to user data and the extent of privileged access to user data for Twitter personnel. Does Twitter, Mr. Sullivan, have in place a system by which you can determine definitively which Twitter employees have accessed private user data, for example, to include history of use of the platform, browsing history, direct message, geolocation data, Indo-Pacific addresses?

Mr. SULLIVAN. Thank you for the question. I can tell you what I have observed. I have been in my role since April of this year. Our current leadership for infosec, the privacy and access controls has a robust process for access to data.

For example, people have to have a business need to access certain datasets, so we have to operate the service, some number of people need access to certain datasets. Our goal, that is aligned with our privacy objective, is to minimize that access to that necessary to do your job function.

We have access controls, monitoring, logging. I receive, for example, new employee approvals, this person needs to be able to run this report.

Senator OSSOFF. Mr. Sullivan, I appreciate the overview, but the specific question to which I am seeking an answer is, is there a log event any time a Twitter employee accesses the private user data

of a specific user? Can Twitter determine every time one of your employees has accessed such private user data? Do you have that functionality? It is really a yes-or-no question.

Mr. SULLIVAN. We have monitoring and logging and access control. It is always evolving and improving. But what I can tell you is I have observed it in action. I cannot speak to every single system. We have a team that can.

Senator OSSOFF. Mr. Sullivan, I will look for that in the follow-up. I want to say, respectfully, this. You are here before the U.S. Senate. Serious allegations were made yesterday by one of your former employees, and I am open-minded. I am here pursuing the facts. Certainly in your responses for the record it is going to help you to be clear, definitive, and precise responding to yes-or-no questions like that one. Can you commit that in your written responses we are not just going to get talking points and generalities, we are going to get precision and yes-or-no answers to yes-or-no questions?

Can I get a yes-or-no answer to that question?

Mr. SULLIVAN. Yes. I am trying to explain——

Senator OSSOFF. Thank you. No. I just need a yes to that question.

Mr. SULLIVAN. Yes, I understand. Thank you.

Senator OSSOFF. So yes?

Mr. SULLIVAN. Yes.

Senator OSSOFF. Great. Thank you.

Let me ask you, please, Mr. Cox. There has been substantial public reporting controversy and concern about the Metapixel product and the possibility that its deployment on various hospital system websites, for example, has enabled Meta to collect private health care data, some of it potentially that would typically be Health Insurance Portability and Accountability Act of 1996 (HIPAA) protected, from U.S. persons.

Does Meta possess or collect any health care or medical data related to its users or to U.S. persons?

Mr. COX. Senator, not to my knowledge, but I would be happy to follow up on that specific issue.

Senator OSSOFF. OK. I would like you to follow up, and please, Mr. Cox, submit to this Committee a comprehensive and precise answer to that question, which I will recharacterize in writing. We need to understand, as the U.S. Congress, whether or not Meta is collecting, has collected, has access to, or is storing medical or health data for U.S. persons or your users. Will you get me a comprehensive and precise answer to that question?

Mr. COX. Senator, yes, we would be happy to follow up.

Senator OSSOFF. OK. Thank you very much.

Ms. Pappas, I overheard some of the responses to Senator Hawley's question. I would like you to answer a question. There has been a significant topical focus on this throughout this hearing. In what ways does the government of the People's Republic of China, if at all, exercise influence over TikTok's corporate behavior or corporate policies? I am going to ask the Chairman's indulgence and follow-up for as much precision as I can get, so I am going to humbly and respectfully ask you not to give me the immediate topline talking points but to give me a precise, particularized answer to that question.

Ms. PAPPAS. In no way, shape, or form, period.

Senator OSSOFF. In no way, shape, or form, period, does the government of China exercise any influence over TikTok's corporate practices or policies.

Ms. PAPPAS. Correct.

Senator OSSOFF. For example, if you receive a response from the government of China to take down certain content for reasons that they State are related to their national security, do you comply with such requests?

Ms. PAPPAS. No.

Senator OSSOFF. Do you comply with such requests if you receive them from the U.S. government?

Ms. PAPPAS. If it follows due legal process, yes. We actually include all government requests for takedown in our transparency reports, in which you can see that China has not requested.

Senator OSSOFF. Thank you, Ms. Pappas. There will be some follow-up questions for you there for the record. I appreciate all of your testimony. Thank you for answering questions, for those which were answered, and Mr. Chairman, I yield back.

Chairman PETERS. Thank you, Senator Ossoff.

Senator Lankford, you are recognized for your questions.

OPENING STATEMENT OF SENATOR LANKFORD

Senator LANKFORD. Mr. Chairman, thank you. Thank you to all of you and your testimony. You have been here a long time. There are a lot of questions. You have gone through a lot of different issues. I apologize I had to be able to run in and out real quick.

Ms. Pappas, I want to be able to follow up on a couple of things real quick. You have answered a lot on China. Obviously it has been a big issue. You know that. It is not like you went to TikTok and were shocked there were issues with China and the possibility there.

There are a couple of questions that have come up recently on this. One of them is the ability for TikTok to be able to track keystrokes after you leave the app, to be able to be on the app, click a link to be able to go to another site, and be able to track keystrokes. Is that a part of the app's design that you can do that?

Ms. PAPPAS. No, it is not.

Senator LANKFORD. That is not used, because it has been widely reported that is part of the app currently and its structure. Has it been part of the app and has it recently been taken off?

Ms. PAPPAS. The keystrokes one, to my knowledge, was basically an anti-spam measure, and so that was never collecting the content of what was being typed.

Senator LANKFORD. Was there an ability, though, to be able to track keystrokes on it as you are on the app, click on a link to be able to go to another page, to be able to track?

Ms. PAPPAS. I do not believe so, no.

Senator LANKFORD. OK. We will follow up on that "do not believe so" on that.

The other one is you have offices all over the world. As you mentioned, a lot of your offices are in Singapore. The original development of TikTok, it is my understanding it came from ByteDance.

It was a Chinese development originally and then it spread all over the world. Correct?

Ms. PAPPAS. Yes, it was originally developed by the parent company, ByteDance, but also Musical.ly, the app, the two were combined. But currently, and for a while now, there have been separate apps, separate codes, separate servers.

Senator LANKFORD. Are any of the developers that still work on the design still based in China, in your Chinese office?

Ms. PAPPAS. Yes. We have said that we have engineers in China. Correct.

Senator LANKFORD. That will be one of the conversations we will have in the days ahead to be able to follow up on, what those access points are. There is an obvious consideration here with this Committee and with others on national security issues. It is just well known that China has, as a part of their law, they get access to anything with technology. For China to have the possibility to have access to 100 million Americans, including most of our young people, that is an issue for us, and it is the reason we ask hard questions.

Ms. PAPPAS. We understand that concern and I appreciate your question, Senator, which is why we are investing heavily in ensuring strict access controls, and we are working with Oracle. We recently announced that 100 percent of our user data is now stored in Oracle's cloud infrastructure, and we have further said that they will be vetting and validating our content moderation and recommendation systems. We really are committed to transparency and security on these topline issues, and we are happy to provide further information.

Senator LANKFORD. Great. We will continue to be able to follow up.

Mr. Cox, thank you, as well, for being here, as for all of you in this conversation. I have a couple of questions here. One is dealing with the experts, as you mentioned in your testimony as well, that are actually helping with the fact-checking process. We did a little bit of digging in some of this, and obviously you have a diverse group of nonprofits and think tanks and other folks that help some of the experts in fact-checking. But there are also some that make us scratch our head a little bit on it.

There was one of the groups that was dealing with coronavirus and some of the fact-checking early on on that was actually a group of journalists. As we went through and looked at some of the credentials, all of which were public on all these individuals—thanks for the transparency on that—none of them were medical professionals on it.

Not to be pejorative on journalists, but I do not run into a lot of conservative journalists. There are a few out there. The consistent fear is that conservative voices are silenced, and when I look at some of the groups that actually do the fact-checking I do not find a lot of conservative groups that do this.

Ms. Pappas, on the same kind of issue, as I go through for TikTok they list as one of the fact-checking groups, or the experts that are out there, the Southern Poverty Law Center as one of the places they go. The Southern Poverty Law Center is considered the Family Research Council and the Alliance Defending Freedom,

which are just pro-family groups and religious freedom groups, as hate groups. If TikTok is dependent on the Southern Poverty Law Center to be able to find what is a hate group, the Family Policy Council is a hate group, suddenly, on TikTok.

The question is, how do you develop your expert groups? How do you make sure that they are actually balanced and that the advice you are getting on what that looks like is actually fair?

Mr. Cox, do you want to jump in first on the Meta side?

Mr. COX. Yes, I would be happy to, Senator. On the issue of misinformation, we know that people do not want misinformation on the platform, and that is why we have developed a program to work with independent fact-checkers that are certified by the Independent Fact-Checking Network (IFCN).

Senator LANKFORD. How do you make sure it is a balanced perspective, philosophically?

Mr. COX. I know that the IFCN has specific policies around looking for balance. I also know that there are folks on both sides of the aisle who are members of that network.

Senator LANKFORD. I would only say, how do you make sure it is balanced, not that organization make sure it is balanced, because again, there is a perception—and I would tell you, I understand their perception because I have a lot of conservative organizations—churches, faith-based nonprofits, all kinds of entities—that reach out to my office, at home and here, and they will reach out to me and say, "I just got blocked from Facebook. We are trying to figure out why." They are not terrorist organizations. They are not violent. They are not anything else. They just got blocked, and they are trying to figure out if conservative ideology is the reason why.

What I am trying to figure out is who fact-checks the fact-checkers for you to be able to make sure that you are getting a fair perspective on this? You have millions or billions of pages IFCN needs to be able to track on this. When someone gives you counsel, how do you take advantage of checking that first to make sure it does not have a bias?

Mr. COX. Senator, on the question of checks and balances among the fact-checker network, the system that we have set up allows fact-checkers to check each other and resolve dispute claims in that way. We believe that that helps the system be more fair.

Senator LANKFORD. But I guess I am asking, so the same question. How do you make sure that that perspective is balanced, that it is not all fact-checkers that think alike?

Mr. COX. Senator, ultimately we believe that our platform is best for people when it can be a place for all voices and for all political points of great.

Senator LANKFORD. Great.

Mr. COX. That is in our interests and that is in the interests, we believe, of the Nation.

Senator LANKFORD. I 100 percent agree. I am trying to say to you it is not, that there are entities that really believe their voices are being blocked out, and that the individuals that are fact-checking them have a bias against them politically, not necessarily for violence or something else. That is part of the challenge here, that I

would challenge you on that, and for all of us, to be able to make sure it is going to be fair and balanced on that.

Let me move on to a couple of other issues. Mr. Chairman, do I have an extra minute here I can go on it? Thank you for that. I have two quick other things that I want to be able to address. In Meta's terms of service you state, in terms of service, 3.2.1, "You may not use our products to do or share anything that is unlawful, misleading, discriminatory, or fraudulent, or assists someone else in using our products in such a way."

But you also have stated, as Meta, "We prohibit content that offers to provide or facilitate human smuggling, which includes advertising a human smuggling service, but we do allow people to share information about how to enter the country illegally or request information about how to be smuggled."

Now I am trying to align those two, where you say you cannot use our platform for any illegal activity, promoting illegal activity, facilitating that unless you are illegally crossing our border. Then we are going to facilitate the use of our platform, and in fact, has been accessibly used to facilitate connecting with the cartels and the traffickers to be able to facilitate illegally crossing our Southern Border.

Help me understand between those two. Which one is correct?

Mr. COX. Senator, we have been working with law enforcement for a while now on the very serious issue of human trafficking across our borders. We have folks at the company who specifically speak to law enforcement and border officials to make sure we have an up-to-date list of which cartels that we can use in order to fan out our systems and make sure we take them down. We have policies against human trafficking, and we have policies against those cartels, to make sure that we are able to remove them as soon as they pop up.

Senator LANKFORD. But this is a Meta statement that you said, "We allow people to share information about how to enter the country illegally or request information about how to be smuggled." That is allows already, based on Meta's policy, but you also say, "We do not allow you to use this for illegal activity." That is what I am trying to figure out, is you either allow for illegal activity or you do not allow for legal activity, but it looks like you are trying to do both. We do not like smuggling, but we are facilitating people who are illegally coming into the country.

Mr. COX. Senator, the policy here, as I understand it, is specifically about human trafficking and cartel networks that are facilitating illegal trafficking of people. I would be happy to follow up with you on this.

Senator LANKFORD. Let us do this. There is not a person that crosses our Southern Border into the United States that does not pay the cartel. As our Border Patrol will tell you, the border is secure. It is secure on the south side.

When I was in McAllen, Texas, a couple of months ago they said in that area the cartels, just in that area, make $153 million a week trafficking people across the border. Many of those meet up with those people that are moving them across the border illegally, through a Facebook platform. That is a big issue to me, and it

seems like Meta is being inconsistent in their terms of service about illegal activity.

My last comment on this, and I really will make it my last comment, and I really do appreciate the time on it, I have, for years gone back to Facebook and said, "I have all kinds of constituents at home that say to me, 'I would comment to you on your page except when I comment I get just ruthlessly attacked by people that politically disagree.'" They click the angry button, they yell at them, they say all kinds of mean things to them when they comment on my page. So they just do not comment.

What has happened is, political Facebook pages—and that is for everyone here, both sides of the aisle—have become places for anger and aggression. When you disagree you go and attack people that comment, that like someone politically, you go attack them instead.

What I have asked Facebook for years is, allow those of us that we know our pages are place where there is wide disagreement to be able to have the option to say, "You can comment to me but you cannot attack the people that comment to me." We can have dialog and interaction but you cannot have this angry interaction with each other on the page. Give us the option to turn that off so we see comments, we can respond back to people and have that dialog and interaction, but you turn down the volume.

What I have heard, year after year, from Facebook is, "That is not really what we do. What we do is interaction." But everyone knows the interaction there is angry, bitter, aggressive interaction. That is not healthy.

My request again to Facebook, which we have done in writing and in follow-up and in conversation, is you have the ability to turn down the volume, to be able to have fewer angry emojis flying at people, by allowing that interaction. Please do.

Chairman PETERS. Thank you, Senator Lankford, for the Senate one minute. I appreciate it. [Laughter.]

We will do a second round, and because of the late nature, and all four of you have been here a long time, if everyone can hold within—I was very generous in the first round. We will try to make sure that we do the seven minutes in this round, if you would.

I want to get back to where we left off which was on the actual design of these products up front, not dealing with problems after they have already arisen and sometimes waiting years before you fix the problems. How do we put it in the initial design?

My question, Mr. Cox, is Wall Street Journal (WSJ) reported last week that Meta shut down its Responsible Innovation Unit. These two dozen employees were charged with identifying potential harms at the conceptual stage of new product design and change the design culture. Why was that eliminated?

Mr. COX. Senator, thanks. Respectfully on this team, because I saw this reporting as well, the work here was not eliminated. The specific team named here was a small team of about 20 people that was overlapping with our much broader integrity, safety, civil rights efforts across the company. This was a case of just moving that work into the teams where it was best——

Chairman PETERS. You just moved them into a different part. They are still there. Is it safe to say that this is happening across

your platform, with our design team? Is everybody on the design team, are they compensated based on the trust and safety of the products that you are putting out? Is that part of the metric?

Mr. COX. Senator, when we look at how health of any product we will look at trust and safety as a part of that. We will look at security. We will look at relevance. We will look at a holistic set of metrics, both quantitative and qualitative.

Chairman PETERS. No, I know. Let me interrupt because of time, because I am going to hold everybody to seven minutes here. In the time, with the metric to determine compensation, does an individual in that metric, do they get compensated based on something related to safety and trust. When they get their bonus at the end of the year, are safety and trust part of that compensation?

Mr. COX. Senator, so for bonuses we would have——

Chairman PETERS. Just say yes or no.

Mr. COX. Is it a part of how we look at the health of the product, which is related——

Chairman PETERS. It is a part. It is an actual line.

Mr. COX. Excuse me, Senator?

Chairman PETERS. It is an actual line, related to safety and trust.

Mr. COX. Trust and safety metrics are part of——

Chairman PETERS. If a product goes out that causes a lot of problems, they are going to be penalized for that, financially?

Mr. COX. Senator, we would not launch a product if we believe it was about to be unsafe.

Chairman PETERS. You would not.

Mr. COX. Once we do launch products we evaluate things like prevalence, things like reports, and a whole host of metrics in order to understand the health of a product from a safety perspective.

Chairman PETERS. But do people get compensated related to safety and trust? Just yes or no. You said yes, they do. I will go down. Mr. Mohan.

Mr. MOHAN. Senator, building in trust and safety into our products is not just an integral part of our goal, so topline metrics, it is our No. 1 priority. But it is also built into the product development process.

Chairman PETERS. I know it is in the process. I just want to know the compensation. Are they compensated specifically because they are working on trust and safety? Is every employee in your product team doing that?

Yes or no.

Mr. MOHAN. If an employee builds a product that does not factor into account trust and safety we simply would not launch that product for our users.

Chairman PETERS. We heard today that people had questions about launching of products, and they still got launched, that trust and safety folks—thank you for your opinion but we have to launch this product because people are compensated—second question, they are compensated based on growth and profitability, like other companies do. You are not the only company that does that, based on growth and profitability, but that is really the main driver.

Mr. MOHAN. Senator, as the Chief Product Officer of YouTube I look after both our product development process and our trust and

safety operations, and I can tell you, unequivocally, that we would not launch a product or grow a product that was at the detriment of our users' trust and safety.

Chairman PETERS. If you launched a product and it turned out it was not like you thought, and it was not trustworthy or safe, would the product designer lose their bonus or they would be compensated?

Mr. MOHAN. It would factor into their performance reviews.

Chairman PETERS. OK. Ms. Pappas. Yes or no.

Ms. PAPPAS. Safety and trust is a core priority for us.

Chairman PETERS. I understand. Is it a part of compensation?

Ms. PAPPAS. Every trust and safety, like every feature, rather, has trust and safety as a seat at the table. As we do our product development and launch process we have actively delayed launches based on not meeting the merits of safety. It is a top priority for us. We invest heavily in this to ensure the safety of our products at launch. In regard to performance, it is one of the factors.

Chairman PETERS. Mr. Sullivan? Be brief, please. Yes or no.

Mr. SULLIVAN. The health and safety is a topline metric for the Consumer Products organization, so that will affect how people's performance is graded.

Chairman PETERS. You have mentioned examples of where you may not have launched or have not launched. I assume that you all have examples where you have not launched because some issues were raised. I would certainly like to have that information. Would each of you commit to giving us an example so we have a sense of what actually is caught before it is actually released?

Mr. Cox, not now, but would you provide an example for us on that?

Mr. COX. Senator, we would be happy to.

Chairman PETERS. Thank you. Mr. Mohan.

Mr. MOHAN. Yes, Senator, I am happy to follow up.

Chairman PETERS. Ms. Pappas.

Ms. PAPPAS. Yes.

Chairman PETERS Mr. Sullivan.

Mr. SULLIVAN. Yes. I also have examples.

Chairman PETERS. Great. Thank you. The other thing is how you deploy resources, and we have heard a lot of numbers here. I think the most valuable resource is just the number of engineers. I am going to go and ask you three questions, each of you to answer.

We sent this to you last week. We have been trying to get this information for a long time. We said I am going to ask you this question today, so I am sure you are prepared for the question because we asked it Friday.

Each of you, what is the total number of full-time engineers you have in your company? How many of those engineers work full-time on ensuring trust, safety, or integrity of your platforms? Three, how many engineers work full-time on product development?

Mr. Cox.

Mr. COX. Senator, the total number of engineers at the company is on the order of tens of thousands.

Chairman PETERS. No. That is not what I asked. We asked very specific questions on Friday. We have been trying to get this infor-

mation for a long time. We said we are going to ask you this question in the hearing, and you are saying you did not get it. You do not have it for me? OK. Mr. Mohan.

Mr. MOHAN. We have thousands of engineers that work at YouTube.

Chairman PETERS. OK. You do not have a specific answer for me either. Ms. Pappas.

Ms. PAPPAS. I do not have the engineer numbers but trust and safety represents our largest labor expense for TikTok.

Chairman PETERS. OK. You do not have numbers. Mr. Sullivan, you do not have numbers as well, or do you? I hope you do. Please, one of you do. We have been trying for months to get these answers. This is why we get so frustrated.

Mr. SULLIVAN. We have about 2,200 people working on trust and safety across Twitter.

Chairman PETERS. What is the total number of full-time engineers?

Mr. SULLIVAN. I am sorry. That was not an engineer number. This is those that build and enforce the Twitter rules. We have several thousand engineers at Twitter.

Chairman PETERS. So the same thing. You do not have specific numbers, as we asked. OK.

Would you commit to get me those numbers, Mr. Cox?

Mr. COX. Senator, I am happy to have the teams follow up.

Chairman PETERS. That is a yes. Mr. Mohan.

Mr. MOHAN. Senator, I will have my teams follow up as well.

Chairman PETERS. Thank you. Ms. Pappas.

Ms. PAPPAS. We are actively working to get you those numbers. We will follow up as appropriate, yes.

Chairman PETERS. Thank you. We are trying to work together. This is really a complex problem. We get it. I understand the complexity of the problems you have to deal with each and every day. We want to work with you, but we need to be able to have this kind of dialog to get a better sense of what is going on as we go forward, so please do that.

Ranking Member Portman, you are recognized for your questions.

Senator PORTMAN. Thank you, Mr. Chairman. Not to leave Twitter out, I wanted to ask a question regarding the sexual material online that we talked about earlier. As I said, this Committee has been a leader in stopping human trafficking, and specifically sex trafficking underage kids, and we have passed some legislation that is making a difference.

Based on a website called Bark, that advises parents on how to get their kids safe, among the top five severe sexual content sites was Twitter. This year it was widely reported that Twitter considered monetizing sexual content, meaning, as I understand it, people could actually get paid for pornography, basically, for putting sexual content online. My understanding is this project has now been put on ice because a group of Twitter employees found that the platform could not effectively separate out child exploitation content, and I appreciate you did not go forward with this plan.

According to Verge, the Twitter employees have said that despite executives knowing about the child sexual exploitation problems on

the platform they have not committed sufficient resources to detect, remove, and prevent this harmful content. This is a news story that I would like to ask be made part of the record.[1]

There are lots of issues here. One is you made the right decision not to monetize this explosive conduct at this time, which is really pursuing a pornography scheme as I see it. But I wonder if you can give us a commitment today to halting this program indefinitely so as to prevent the platform and bad actors from making money off of child sexual material?

Mr. SULLIVAN. First may I say that we abhor CSAM, the sharing of sexual material. I appreciate your work there. I worked on this here and also at Meta, so I have been working on this for years.

I made that decision to pause this idea. It was not a product. It was a set of people had an idea that they thought they might want to pursue. I said I want to look at all the information here and learn about where we stand, what the risks could be. I think this is how the system should work. We looked at a product in its very early ideation and did the analysis and got the perspectives, and said this is not appropriate for us to be doing. So that is how the process went.

Senator PORTMAN. OK. So you made a commitment today not to pursue it?

Mr. SULLIVAN. We are not pursuing that.

Senator PORTMAN. You have made a commitment not to pursue it in the future?

Mr. SULLIVAN. We have no plans to pursue monetization of adult content. That is correct.

Senator PORTMAN. You have no plans to do it. Can you just tell us you are not going to do it?

Mr. SULLIVAN. I am not planning to do it, no.

Senator PORTMAN. You are not planning to.

Mr. SULLIVAN. I am not doing this.

Senator PORTMAN. Just say you are not going to do it.

Mr. SULLIVAN. We are not planning to do it, no.

Senator PORTMAN. Cannot get a "planning" out of there.

Not to, again, leave anybody out, Mr. Mohan, we have not had a chance to talk yet. I want to ask you about something that is important to this Committee, and I hope a way forward in terms of legislating and regulating platforms. Your platform's algorithms have been described as a "black box," according to experts and researchers, meaning there is little to no transparency in the algorithms. I am sure you have heard that before.

The question is, is there a way to come up with a transparency approach that makes sense as calls grow for Congress to pass legislation? I like the idea of having much better information than we have, getting behind the curtain and getting into that black box.

That is why along with Senator Chris Coons I drafted this legislation called the Platform Accountability and Transparency Act. It would require the largest tech platforms to share data with vetted, independent researchers and other investigators so that we can know exactly what is happening with regard to the privacy issues

[1] The information referenced by Senator Portman appears in the Appendix on page 163.

we talked about today, or content moderation, product development, sexual exploitation issues, key industry practices.

My question for you, Mr. Mohan, would you be supportive of legislation like PATA to get at this need for transparency and for us to be able to legislate with better information?

Mr. MOHAN. Yes, Senator, I would be supportive of the spirit behind that regulation. The reason why is because I agree with you. I do think that transparency around our practices, how we go about them, is an important thing. It is the reason why we have invested so heavily in our quarterly transparency report, which you may be familiar with.

It is also the reason why we, just a few weeks ago, launched the YouTube research program, which is similar, in my understanding, to what the act that you are referring to is trying to get at, which is giving academic researchers access to our raw data, obviously in a user privacy-sensitive way, where they can derive metrics or derive insights of their own based on that data. We have taken it a step further where we will also provide technical support that these researchers might need to get at the insights that they are looking for.

I am very bullish about that transparency program, and based on the feedback that we hope to get from researchers look forward to enhancing it in the future as well.

Senator PORTMAN. We are following your YouTube research program carefully. We are glad you created it. We want to see what the results are and we want to be sure these are independent individuals who will give actual information about, what the algorithms are, again, what is in the black box so that citizens can understand it better, and as legislators we can legislate better. I think that is a positive step.

With regard to PATA, can I hear from the other members of the panel how you feel about this legislation? We have shared it with all of you. We hope to introduce it soon. Again, it would be bipartisan, and it would be one that would, I hope, give us a way forward as a first step. Mr. Cox.

Mr. COX. Senator, thanks. I know our teams have been in contact with yours on this. We are aligned that more transparency about content on our platform is a good thing. It is a good thing for the public. It is a good thing for the company.

We also have an academic research program called Fort, where we have designed privacy-protected ways of sharing information with outside academics and researchers. We have also released a widely viewed content report which helps folks get access to which content is seen the most times on the platform. We also publish quarterly community standards enforcement report which gets into categories of content by region and shows the work we are doing every day.

We are committed to working with you on this.

Senator PORTMAN. Yes. We talked about regulations needs.

Ms. Pappas, yes or no?

Ms. PAPPAS. Senator, transparency builds trust. We were the first open platform to open our own Transparency and Accountability Center for that specific reason, so people could take a look at our content moderation systems, recommendation systems as

well. Last month we announced that we will be opening our API to researchers as well, so we would be happy to support that legislation.

Senator PORTMAN. OK. Thank you. Mr. Sullivan.

Mr. SULLIVAN. Yes. We have been publishing data to researchers for years, and we are very open to anything that improves transparency. Especially as AI moves forward, it going to be very important.

Senator PORTMAN. It is important. It is needed. Thank you, Mr. Chairman. Thank you all.

Chairman PETERS. Thank you, Ranking Member Portman.

Senator Johnson, you are recognized for your questions.

Senator JOHNSON. Thank you, Mr. Chairman. Mr. Cox, just a quick little housekeeping here. Are you aware of a letter Senator Grassley and I sent to Mr. Zuckerberg on August 29th? We did get a reply on September 12th from Mr. Kevin Martin, just saying they are going to respond. Are you aware of that letter asking for information, contact between yourself, FBI, Department of Justice, documents, names, that type of thing?

Mr. COX. Senator, yes, I am aware of that letter and I know the team is working on following up as quickly as they can.

Senator JOHNSON. You will commit to full response on that?

Mr. COX. I know the team is committed to a response, yes.

Senator JOHNSON. OK. Let us put up my first chart[1] here. Back in November 2021, CDC Rochelle Walensky stated in front of the Health Committee, "We have the most robust, safe vaccine safety system we have ever had in this country."

In October 2020, before the vaccine was approved, CDC's Dr. Tom Shimabukuro stated in a web seminar, " Vaccine Adverse Event Reporting System (VAERS) is obviously something that Ms. Walensky was talking about—"VAERS traditionally has provided the initial data on the safety profile of new vaccines when they are introduced. For COVID, vaccine reports will be processed within one to five days. Depending on the seriousness of the report, CDC and FDC received updated datasets daily, and data-mining runs are planned to be conducted every one to two weeks."

This is an example of the timeliness and responsiveness of VAERS, going back to H1N1. It kind of sounds like they are really going to rely on VAERS. I remember part of that discussion when they said, "Listen, we are going to take vaccine safety so seriously, if we get a report of a couple of days of lost time because of an injury we are going to be calling that individual up and we are going to be checking on it." It really sounded like they had this all covered, right?

Let us see what they actually did. I produced this chart[2] because I took VAERS and FAERS seriously, and I started tracking this, and I started putting together this chart. I want to quickly describe what this is. The first five lines, the first five drugs, four of them are in the FDA Adverse Event Reporting System (FAERS) system, the FDA Adverse Event safety system: ivermectin, hydroxychloroquine, dexamethasone, and Tylenol. You have the flu

[1] The chart referenced by Senator Johnson appears in the Appendix on page 192.
[2] The chart referenced by Senator Johnson appears in the Appendix on page 193.

vaccines in there. That comes off the VAERS system. You have remdesivir, which comes off of FAERS, and COVID vaccine that comes off of the VAERS system.

Now I can see why the government really did not like the way I put this is their data. I did not make these numbers up. This off the VAERS and the FAERS system. But for whatever reason Twitter censored this chart.

Now just quick, and I will show you the current version, the one you censored, it showed that ivermectin, on average, over 26 years, on average, had 15 deaths reported on the FAERS system. Hydroxychloroquine had 69 deaths. Flu vaccines had 77. Dexamethasone had 618. Tylenol had over 1,000. Remdesivir, since it was approved, had 1,612, and the vaccines had 21,000 deaths. OK, these are just the facts, and Twitter censored it. Do you have any idea why?

Mr. SULLIVAN. Senator, I was not at the company at the time, but what I can tell you is that we want robust discussion on the platform, of any issue. A COVID misinformation policy was developed that seemed to me—and again, I do not develop it—but it seemed quite narrow to me.

Senator JOHNSON. You censored government information. Here are the current numbers, by the way. Over 30,000 deaths reported worldwide, 27 percent of those, by the way, have occurred on Days zero, one, or two. You did not only censor this chart, you censored, for example, radio shows that interviewed me, talking about FDA–CDC data.

YouTube took down a video of this Committee's hearing, of an eminently qualified critical care specialist who saved thousands of lives treating people, using what seems to me pretty safe drugs. After eight million views, YouTube pulled that video down. What would be the justification for YouTube pulling down a hearing of the U.S. Senate with a highly qualified doctor just giving a second opinion on how to save lives during COVID? Why would YouTube do that? On what authority, whose authority, are you censoring that information so the American public could not receive a second opinion, and access drugs that might have saved their lives? Why would YouTube do that?

Mr. MOHAN. Senator, respectfully, as I was mentioning earlier, we did not decide those policies on our own. We worked with third-party health authorities in this country. That did include the CDC or the FDA.

Senator JOHNSON. I will be sending you a letter, and I want to know who those health authorities were, and I want to see the communications between them. Will you commit to providing me that information, for transparency's sake?

Mr. MOHAN. Senator, I am happy to follow up on your request on how we developed that policy.

Senator JOHNSON. In July 2021, talk about misinformation, this should have been the 2021 lie of the year. President Biden said, "You are not going to get COVID if you have these vaccines. If you are vaccinated you are not going to be hospitalized, you are not going to be in an ICU unit, and you are not going to die." That is the President of the United States.

It just so happens we could not rely on the CDC and the FDA because they were not honest, they were not transparent, they were not giving us data, so we had to go to Public Health England. This is a chart[1] published from their Technical Briefing Number 23, that covered the period from February 1 to September 12, 2021. It shows 593 cases of mainly Delta, 2,542 deaths, 1,613 deaths occurred with the fully vaccinated.

Obviously, this was published, and they were publishing other similar information during that time period when President Biden lied to the American public that this was a pandemic event and vaxxed, and if you got vaccinated you are not going to go to the hospital, you are not going to be in an ICU unit, and you are not going to die. Well, 63.5 percent of the people fully vaccinated were dying in England at the exact same time.

Why did you not pull this? Have you ever labeled the President of the United States' comment as misinformation? Have you ever done that? Any of you? I will take that as a no.

Again, I just wonder, who are the authorities, who do you think you are to censor information from eminently qualified doctors who had the courage and compassion to treat COVID patients when the National Institute of Health (NIH) guideline was basically if you test positive for COVID, go home, be afraid, isolate yourself, do not do anything until you are so sick, we will send you to the hospital, we will give you remdesivir, where we have 1,600 deaths so far, we will put you on a vent, and we will watch you die.

You guys bear a fair amount of responsibility for hundreds of thousands of people not being treated, and I would say probably dying that did not have to die. I hope you are proud of yourselves.

Chairman PETERS. Senator Lankford. Here now in the second round Senator Lankford, Senator Hawley, and then you.

Senator LANKFORD. I will give back part of my magic minute here.

Chairman PETERS. Yes, please do that.

Senator LANKFORD. I will go short on this. I do want to follow through on a couple of things there on illegal activity. You have all been very outspoken on dealing with sexual child predators, with different issues, drug trafficking. Those were all good things to be able to engage on.

But it is fascinating to me that the platforms have chosen to say there are some illegal activities we are OK with, and, in fact, we are going to facilitate. One of those is illegally crossing our Southern Border. It is not hard for me to go to YouTube, and I just type in "how to cross the border illegally" and I get a video that says, "How to illegally cross the Mexico-U.S. border." It has 1.7 million views, and it has been there for two years.

Yes, I watched it, and it showed where to be able to cross, what highways to avoid, where the Border Patrol typically puts up stations, how to be able to look for different aspects. In detail, shows a video of here is how to illegally cross the Mexico-U.S. border, and where to be able to cross, and how to avoid border patrol. This has been up for two years, and it has had 1.7 million views.

---

[1] The chart referenced by Senator Johnson appears in the Appendix on page 195.

As I mentioned, on Facebook, Facebook has ads that I can actually show you that are human smugglers placing ads in Central America so people will know how to be able to connect with them, to be able to travel through Mexico, to be able to pay the cartels, which are a ruthless drug organization, to be able to get in the United States.

My confusion on this is I do not understand why the platforms look at illegally crossing the border as "we are going to look the other way" when your user agreements say "we do not promote illegal activity except for this one." Help me understand why that is different.

Mr. MOHAN. Senator, I do not know about that specific video. I am happy to follow up.

Senator LANKFORD. It is not just one. It is a bunch. That is just the first of many.

Mr. MOHAN. I am happy to follow up on those.

But just in general, we do have very clear policies where content that encourages dangerous behavior, not just illegal behavior but dangerous, harmful behavior is removed from our platform. We have the Four R's approach that I described in my opening testimony, where it not about removal of content but also reduction of content, raising up authoritative sources. In the context of people searching for that type of information, making news stories from prominent mainstream news outlets prominent. We do try to have a holistic approach to dealing with this type of content on our platform.

We are not perfect. We continue to improve both our policies as well as our enforcement. In this specific case I am happy to follow up. But we do have very clear policies against cartels, harmful criminal conspiracies, other types of organizations where their type of activity is not allows on our platform.

Senator LANKFORD. I would only say this particular video, which, by the way, this one is in English, this particular video even talks about how to be able to connect with a cartel and how much the cost is going to be when you get to the Southern Border.

Mr. MOHAN. Senator, I am happy to follow up, but we do take our enforcement——

Senator LANKFORD. I get it. This part is not being enforced. That is what I am trying to say to you, is that I do see all the platforms trying to deal with drug trafficking, but human smuggling and illegally crossing the border is not being enforced. I am not asking you to solve it today. I am raising it as an issue to say somehow we treat cartels different than terrorist organizations.

Cartels are transnational criminal organizations that are making money off of moving people illegally into our country and making money off of illegally moving drugs into our country. I would like for our social media platforms to engage with a criminal organization and with criminal activity consistent to your own terms of service.

That is it. I yield back my time.

Chairman PETERS. Thank you, Senator Lankford. Senator Hawley.

Senator HAWLEY. Thank you, Mr. Chairman.

Mr. Cox, I know that Facebook has said in the past that it is their position, as a private company, you are not subject to the First Amendment. I assume that has not changed. Is that right?

Mr. COX. That is correct, Senator.

Senator HAWLEY. But the United States government is subject to the First Amendment. I think we can probably all agree on. Hopefully we can. Hopefully that is still true in this country.

Is it appropriate for Facebook to work with the United States government to avoid the First Amendment, help the U.S. Government avoid the First Amendment?

Mr. COX. Senator, we do think it is sometimes appropriate to be in contact with government and with government organizations.

Senator HAWLEY. To help them avoid the First Amendment?

Mr. COX. Senator, I am not sure what specifically you are referring to.

Senator HAWLEY. Let me ask you this. Do you think it is appropriate to work with the United States government to target private individual speech that is constitutionally protected?

Mr. COX. Senator, I am not aware of that.

Senator HAWLEY. Let me educate you. On July 16, 2021, an employee at Facebook wrote to the Department of Health and Human Services, saying, "I know our teams met today to better understand the scope of what the White House expects from us on misinformation going forward."

On July 23, 2021, a Facebook employee thanked HHS, quote "for taking the time to meet earlier today, and wanted to make sure you saw the steps we just took this past week to adjust policies on what we are removing with respect to misinformation. This included"—and I am still quoting—"increasing the strength of our demotions for COVID and vaccine-related content."

On April 7, 2021, a Facebook employee thanked the CDC for responding to misinformation queries, and I quote, "We will get moving now to be able to remove all but that one claim as soon as the announcement and authorization happens."

On July 28th of this year, a Facebook employee reached out to CDC about, "doing a monthly misinfo/debunking meeting." The CDC responded, "Yes, we would love to do that." I am sure they would.

On July 20, 2021, Clark Humphrey at the White House, who was digital director of the COVID–19 response team, emailed David Sumner at your company, among others, asking, "Any way we can get this pulled down," and cited a specific Instagram account. Within 46 seconds, your company replied and said, "Yep. On it." That sounds like what, in the law, we call a pattern and practice of meeting, coordinating, and colluding with the United States government to target particular speech that no one in any of these emails alleges is incitement, which would not be constitutionally protected, no one in any of these emails alleges it directly encourages violence, which would not be constitutionally protected.

It appears to all be constitutionally protected speech on, I might add, very politically sensitive topics, that Facebook is directly working with the U.S. Government to target and remove. Is that your company policy to do this kind of thing?

Mr. COX. Senator, we were quite public about our cooperation with health organizations during the unprecedented time of COVID. We knew that people expected and wanted accurate information on our platform. We had conversations with CDC, with the World Health Organization (WHO), and with other public health organizations, not just in the United States but abroad, in order to understand how to help make sure that folks were not getting information that could cause them any harm.

Senator HAWLEY. Fair enough. You are saying that this was, in fact, company policy to have these kinds of meetings with HHS, with the CDC, with the White House directly, that you did engage in this behavior, and you think that it was entirely fine. Is that your testimony?

Mr. COX. Senator, I do believe it is appropriate for companies like ours to be in consultation with public health organizations and with government.

Senator HAWLEY. You can confirm that things like taking down a private Instagram account and adjusting your policies at the behest of the White House, and putting into place misinformation policies at the behest of CDC, that those things, you think, are appropriate, that this was company policy to do so. Is that fair to say?

Mr. COX. Senator, I am not familiar with the Instagram account specifically that you are referencing, but we do know that people expected and hoped from the platforms that we would help them get accurate information about COVID during the unprecedented time, especially at the beginning.

Senator HAWLEY. Is there not a difference between you, as a platform, putting forward information, and censoring your users at the behest of the White House, the Administration more broadly, and the CDC? Is there not a distinction there?

Mr. COX. We specifically wanted to work with public health experts to understand the relationship between information and behavior, and so we did consult with the CDC, the World Health Organization, and others to understand how the platform policies we built were affecting public health.

Senator HAWLEY. You did not just consult them to understand how they affected public health, you actually censored on their behalf. I mean, you took these emails—I am just quoting from a sample of them—which, by the way, have been disclosed in litigation—these emails show that you took censorship steps. You took down accounts. You planned misinformation policies. You adjusted your policies at the behest of the United States government. That is not just some theoretical thing. That is actually targeting your users' speech.

I appreciate your forthrightness, by the way.

But you think that is fine, and that was your policy.

Mr. COX. Senator, we have been public about our policies, on COVID misinformation specifically, as well as on misinformation generally.

Senator HAWLEY. You are not concerned about any of this. Nothing that I just read to you, you are not concerned about it at all.

Mr. COX. Respectfully, Senator, I think the balance of how to protect free expression as well as public safety is a difficult issue, but

it one we are committed to working with outside experts and publishing our work.

Senator HAWLEY. I appreciate you being so forthright. As I said, this is actually from litigation between the State of Missouri and the State of Louisiana and the Federal Government. I anticipate that your remarks under oath today are going to be very interesting and helpful to that litigation.

I will just say this. My view is that the United States government is bound by the First Amendment. They cannot encourage or coerce or incite or collude with a private party to get around the First Amendment, that you have just said to me today that that is basically what they did, that you coordinated with them repeatedly, over a pattern of months and years, to adjust and target your speech policies for protected speech at the behest of the United States government.

I have to tell you, I have a big problem with that, and I think all your users should too.

Thank you, Mr. Chairman.

Chairman PETERS. Thank you, Senator.

To our panelists, it is at 5 p.m. I know there was a suggestion for a break. We are right down to the end here. Rather than break and come back and keep you here longer we are going to power right through it with Senator Scott, and then I will wrap it up after that. Senator Scott.

OPENING STATEMENT OF SENATOR SCOTT

Senator SCOTT. All right. Thank you, Chair Peters. Thanks to each of you for being here.

It is critical for employees and officers of the FBI and DOJ to continue to have a mechanism for reporting concerns of fraud, waste, and abuse within their respective agencies to Congress without fear of reprisal from DOJ/FBI leadership.

Whistleblowers from within the DOJ/FBI have come forward with concerns about the Department of Justice's alleged political bias in the FBI's raid on the former President's home in Florida last month. FBI agents have reported similar concerns to individual Senate offices as well. We all need to ensure safeguards are in place so Attorney General Garland does not retaliate against or intimidate FBI agents and DOJ employees who come to Congress as whistleblowers.

Mark Zuckerberg recently disclosed that Facebook's restriction of a story about Hunter Biden during the 2020 election was based on the FBI's, "misinformation warnings." Additionally, emails and internal communications obtained by the journalist Alex Berenson, in his lawsuit against Twitter, have shown his removal from the social media platform was a result of pressure from Biden White House officials to silence his criticism of the Administration's COVID–19 policies.

These instances and several others show a clear and alarming pattern of speech suppression carried out at the direction of agencies and officials in the Federal Government. In other words, the Federal Government used private businesses to violate the First Amendment rights of our citizens. This also confirms that the Federal Government used officials at the FBI to interfere in the 2020

election by manipulating the normal flow of public discourse and information-sharing with false warnings about foreign interference and disinformation.

I am going to ask you a couple of questions, if you can just show by hands yes or no. By a show of hands, how many of you and your companies have been contacted by a Federal agency, an agency official, or a member of the Biden White House with a request to remove, censor, or restrict access to a post or an individual user on your platform? If you have, would you raise your hand and say yes.

Mr. SULLIVAN. I am not aware of it.

Senator SCOTT. So Meta, YouTube, TikTok, and Twitter all said never.

Mr. COX. Not to my knowledge, Senator.

Senator SCOTT. You would know, would you not?

Mr. COX. I have not been in conversations with the FBI, Senator.

Senator SCOTT. OK. So no conversations. OK.

By a show of hands, how many of you and your companies have felt pressure to remove, censor, or restrict access to a post or individual user on your platform based on that contact with a Federal agency or official?

Is the answer no from all four of you? So all four of you say no.

Mr. COX. Not to my knowledge, Senator.

Senator SCOTT. OK. By a show of hands, how many of you and your companies have received a misinformation warning issued by the FBI? Every one of you is saying no?

Mr. COX. Senator, I know that we have received warnings from the FBI and other experts about electoral misinformation, in general, and foreign interference, in general.

Senator SCOTT. By a show of hands—but the rest of you have said no.

Mr. MOHAN. Senator, we receive information from every administration about things like foreign interference and election results, et cetera. But one thing is very clear. We enforce our guidelines based on our community guidelines, so we are the ones making decisions about the content that is removed, based on the guidelines that we publish transparently, not based on what a particular administration asks us to do or not do.

Senator SCOTT. So your answer is you have never received a misinformation warning issued by the FBI?

Mr. MOHAN. Senator, no. What I am saying is that we do receive information from the FBI in terms of imminent threats, foreign actors trying to interfere with our free and fair elections here in the United States, and we take that into account in terms of the enforcement of our policies. But those decisions about the enforcement of our policies are made solely based on our community guidelines that we publish on our website.

Senator SCOTT. Mr. Cox, what I said about Mark Zuckerberg disclosed the FBI's restriction of a story about Hunter Biden during the 2020 election was based on the FBI's misinformation warning, is that untrue?

Mr. COX. Senator, I was not in conversations with the FBI so I cannot speak to exactly what the conversation was. What I do know is that we were in contact with a number of organizations who warned us in the time leading up to the 2020 election, to be

on the lookout for foreign interference in elections, and that is an issue that we take incredibly seriously.

Senator SCOTT. But you should know if Mark Zuckerberg, right, would you not know——

Mr. COX. Sorry, Senator.

Senator SCOTT [continuing]. Would you not know if Mark Zuckerberg—he said that Facebook made that decision. You would know that, right?

Mr. COX. Senator, I do know that, if you are talking about the New York Post story, I do know that consistent with our policies we made the decision to submit that story to be reviewed by independent fact-checkers. It was never removed from our service, and we never blocked anybody from sharing that story.

Senator SCOTT. I know people in Florida that were kicked off by putting a story up.

Last one. By a show of hands, how many of you and your companies have felt pressured to act upon an FBI misinformation warning you received, such as by removing, censoring, or restricting access to a particular user or post as a subject of that misinformation warning?

So the answer is no for all of you.

Mr. COX. Not that I am aware of.

Senator SCOTT. No for all of you. So every one of you said no. All right.

Mr. Cox, in 2013, DOJ shut down Silk Road, the illegal online marketplace on the dark web which featured over 13,000 illegal drug postings. In comparison, according to the 2020 Facebook transparency report, Facebook found 5.9 million illegal drug sale postings on Facebook and Instagram. That is 453 times more drug postings than Silk Road. If you found almost six million postings, how many posts are you not finding? If that is true, if Silk Road got shut down, what in the living daylights are you guys still doing in business?

Mr. COX. Senator, we release quarterly reports on the specific answers to your question around each category of bad content on the platform. We do not believe the sale of illegal drugs has a place on any of our platforms. We work hard to fight against that. We publish quarterly the updates on exactly how many pieces of content we take down, as well as how much we are about to take down proactively.

Senator SCOTT. All right. I have one more question, and this has to do with what has happened with my sheriffs. Do you collect stats on the average turnaround time for responding and resolving customer complaints like hijacked accounts or products that violate your terms of service?

Mr. Sullivan, you say yes. No one else does?

Mr. SULLIVAN. Yes. We have a goal of meeting a service level agreement to turn those around as quickly as possible.

Mr. MOHAN. Senator, we do look at how quickly we respond to requests from our creators who upload content to our platform as well as viewers, and we are constantly looking to continue to improve our processes around that sort of request handling.

Senator SCOTT. TikTok?

Ms. PAPPAS. We do as well, yes.

Senator SCOTT. And you do it also?

Mr. SULLIVAN. Are you asking whether we look at turnaround times?

Senator SCOTT. Yes.

Mr. SULLIVAN. Yes we do, Senator.

Senator SCOTT. OK. Do you collect stats on the average turnaround time for responding to subpoenas issued by law enforcement agencies?

Is that a yes for everybody?

Mr. MOHAN. Yes, Senator. We have a group that works to respond to subpoenas, evaluate them, and respond 24/7.

Senator SCOTT. OK. I wrote to all of your companies and I asked, and none of you responded. I do not know if you all realize that. Every one of you, I wrote and asked for information. None of you responded to either of those questions.[1]

Let me just tell you what sheriffs are saying. One of the Florida Sheriff's Departments mentions, "There is no point of contact to send subpoenas. They are slow to respond. There is no sense of urgency on how they respond to something that is even time sensitive."

I can tell you I have talked to sheriffs all around Florida—is that time-sensitive information that would impact a law enforcement investigation or a crime, you guys do not respond. How do you respond to that?

Mr. SULLIVAN. I work in the product organization. We would not be the ones to receive that, but I can have our team get back to you.

Senator SCOTT. Anybody else?

Mr. MOHAN. Senator, I am happy to follow up. We do have a team that responds to those requests. We balance the needs of law enforcement as well as our user privacy when we are responding to those, as you would imagine. We do take into account time sensitivity in terms of trying to respond to those requests. I will ask our team to follow up with you as well.

Senator SCOTT. The questions I asked before that I sent you all, that none of you responded to, you will respond to?

Mr. COX. Senator, I would be happy to have the team follow up. We take law enforcement requests very seriously.

Senator SCOTT. OK. Thank you. Thank you, Chair.

Chairman PETERS. Thank you, Senator Scott.

I have a couple of follow-up questions and we will wrap up and you will be on your way, and again, thank you so much for taking so much time. You have been in the seat a long time. I have been here with you a long time. I am ready to get out of the seat as well, with you.

You have all spoken about how essential trust and safety is, how it is part of your culture and something that you are focused on. But I want to reference a report that was commissioned by a Twitter whistleblower in the spring of 2021, focused on its site's integrity team and misinformation. It found that, "Project managers are incentivized to ship products as quickly as possible, and thus are willing to accept security risk."

---

[1] The Google letter in response to Senator Scott appears in the Appendix on page 197.

Mr. Sullivan, as head of Consumer Products at Twitter, would you agree with this finding in that report, commissioned by Twitter?

Mr. SULLIVAN. The dates of that report would have been before I was in the role, but what I can tell you is that I have been in multiple product reviews where I push hard, and our other leaders push very hard, and work with our teams to strike a balance of safety in all that we do. I cannot speak to that report but what I can tell you is how we operate now, and that is how we operate now.

Chairman PETERS. Obviously this report has a completely different conclusion than you have. I just have to ask you, how can you say that trust and safety are important to your development progress when Twitter launched its Spaces product, despite your predecessor publicly stating that Twitter would not be able to moderate all of its Spaces?

It is my understanding that since the launch it is documented that Spaces has been used by white supremacists as well as ISIS to spread misinformation, as shown in this poster. In fact, the internal report shown on this poster basically says, if I can read it here, "We did not prioritize identifying and mitigating against health and safety risks before launching Spaces." Do you disagree with this characterization? I have heard you say on the record, "We never, ever send anything out that we have concerns about." This is obviously very different.

Mr. SULLIVAN. Yes, I understand what you are saying. Since I have started in my role, we have been looking at health and safety across the board and working to improve it.

In Spaces, for example, we have been continuing to beef up all of our reporting, our automatic detection, our language support. We are working very hard to further improve the health of Spaces. I think that is just one example of many that I could give you for how we are operating now.

Chairman PETERS. After the fact. After some of these things are released.

Mr. Cox, you also have talked quite a bit about trust and safety as being central to Meta's development process. My question to you is, why, after several years of warnings by external organizations such as the Tech Transparency Project, does Facebook continue to automatically generate home pages for white supremacists and other extremist groups and terrorist such as ISIS, as shown in this poster right here for the Aryan Brotherhood, a page that has been created? I guess it was taken down just recently, but it was on Facebook for 12 years, for 12 years the Aryan Brotherhood.

Does not this feature allow extremist groups to basically recruit members more easily because you are putting this up?

Mr. COX. Senator, we believe there is no place for terrorism or violence-inciting networks, for militarized social movements. We believe there is no place for these on our platform. We use automated tools to find and take them down as well as teams of experts dedicated to these specific problems.

Chairman PETERS. Your automated tools and teams, they were successful after 12 years. They were able to bring it down after 12 years. Do you think that is an acceptable performance?

Mr. COX. Senator, I know that for 97 percent of terrorist content we are able to get to it before anybody reports it, and also that we have been able to improve that number, quarter over quarter, and I will continue to make sure that we aspire to.

Chairman PETERS. But certainly any content that gets through and is disseminated very broadly can have catastrophic consequences and violent actions, and particularly groups like this where they have pages that are being put up by your company that are there for 12 years. I hope you would agree that that is unacceptable.

Mr. COX. Senator, respectfully, we would not have put this page up ourselves, but we do work hard to make sure that extremist and terrorist networks are not allowed——

Chairman PETERS. Yes, this is auto-generated. This was an auto-generated page.

Mr. COX. Senator, I have not seen this specific example.

Chairman PETERS. I would love to have you comment on this. If you could look at this example, and if we could have your written comments on it would appreciate that afterwards.

My final question, and then we are going to let you go, seriously. When your product teams are testing new products or features I know that you track engagement and growth, a pretty fundamental part of the work that you do.

My question to you, and this is to each of you, do you consistently measure the impacts of these new features on societal harms like misinformation, disinformation, hate speech, and terrorism? If you could just give me a yes-or-no answer. We will start with you, Mr. Sullivan, and work that way. Yes or no, please.

Mr. SULLIVAN. Yes, depending on the feature and which of those harms might apply, we go deep into those and analyze those, yes.

Chairman PETERS. So it is yes. If you could go a little deeper I am going to ask you then, how do you characterize and measure this data, and which metrics are used? If you can be very specific, it is very helpful that we have that information. I will ask all three of you to do the same, please.

Mr. SULLIVAN. I will give you one topline metric we use for many of these, which is what we call "harmful impressions." So 0.1 percent of tweets that turn out to be violative, we want to limit the number of people that see them. It is how many people may have seen something before we identified it as violative. Those are some of the metrics that would be important to combat this harm on the platform. Thank you, Senator.

Chairman PETERS. Thank you. Ms. Pappas.

Ms. PAPPAS. Similarly the way we measure it is looking at community guidelines violation rate. Essentially we do a sample size that is view-weighted of our corpus of videos, and then we are looking at was there any violative content, and then we are looking at how do we minimize that exposure and drive it down to zero.

As I had mentioned earlier, we look at that at a per-policy basis, so things like hate speech, violent extremism, mis-and disinformation, and we are able to measure our improvement of that, quarter over quarter, week over week. We look at those metrics and reports, and doing so with regard to our features, like our For You feed.

Chairman PETERS. You do this before every launch, this is the study you will do?

Ms. PAPPAS. We would not have the metrics before launch, but in terms of setting our baselines and knowing——

Chairman PETERS. Right. Mr. Mohan.

Mr. MOHAN. Senator, I can say very clearly that our responsibility as a global platform comes before any of our growth metrics around engagement, revenue, et cetera. It is the top line of the company, and we are constantly reviewing our products.

Chairman PETERS. How do you characterize the data?

Mr. MOHAN. When we are reviewing our products on a constant basis one of the core metrics that I look at, and that the rest of the company looks at, is something called our "violative view rate (VVR)." That is the metric that we have also started to publish on a quarterly basis so that you have access to it. In fact, our most recent transparency report was just published a couple of weeks ago. The violative view rate is basically a metric that calculates how much content is up on our platform that would have violated any of our policies, across hate speech, harassment, et cetera. That number is something on the order of 9 to 11 impressions out of 10,000. It is a small number that we aim every single quarter to continue to drive down.

Chairman PETERS. I want to be clear. You do this with your AB testing (split testing). This is testing that you do before you launch a product?

Mr. MOHAN. We measure this metric on a constant basis.

Chairman PETERS. But do you do it before you launch a project?

Mr. MOHAN. We would not have the metric——

Chairman PETERS. You do testing. You do not launch without doing some testing. You do AB testing. I have not heard about AB testing here from you today. You do not test the product before you launch? You just say, "Hey, let's launch and see what happens"? I do not think you do that. Is that what you do? You launch products without testing whether or not it makes sense?

Mr. MOHAN. Senator, I did not say that.

Chairman PETERS. OK.

Mr. MOHAN. We test our products extensively before launching, in terms of usability of the products, but also the trust and safety and the impact those products will have on our users.

Chairman PETERS. OK. You do those tests during the test.

Mr. MOHAN. Correct.

Chairman PETERS. Ms. Pappas, you said afterwards. Do you do that AB testing before you launch, and do you also test safety and trust, because you are testing your product before you send it—unless you do not test it before you send it out to the world.

Ms. PAPPAS. We do do testing before launch, and we will delay products, or rather features, if they do not meet our safety standards.

Chairman PETERS. Mr. Cox.

Mr. COX. Senator, to your question, the primary way we measure and understand is prevalence, and we publish reports on the specific categories of content that violates our policies as well as regions around the world.

Chairman PETERS. You do that in your AB testing before you launch?

Mr. COX. Senator, for many of those metrics you need a specific study in order to understand that metric, but we look at lots of other metrics associated——

Chairman PETERS. Does that mean you do not do it before you launch?

Mr. COX. What we would do before the launch of any product where we had any reason to be concerned about safety is put it through a review with our integrity teams, whose job is to understand safety concerns. We would not launch a product if we believed that there was a safety issue.

Chairman PETERS. You have referenced several times the statistic that hate speech on your platform represents 0.02 percent of all views. Is that accurate?

Mr. COX. That is correct.

Chairman PETERS. Certainly that sounds like a small number. I will appreciate that. But you also have a lot of views. You are a massive platform. I am concerned that this could mask the total amount of hate speech that could be out there and viewed by an awful lot of folks. My question is, what is the total number of views that hate speech actually gets on your platform, not a percentage, but how many views, last year, for example, or last month, yesterday, whatever you may have? Do you have those numbers?

Mr. COX. I do not have those numbers right now, Senator, but I would be happy to have our teams follow up.

Chairman PETERS. Would you provide those numbers to the Committee as to the total number of views, not as a percentage?

Mr. COX. Yes, I would be happy——

Chairman PETERS. You can do the math. You have massive amount of views on your platform.

Mr. COX. Yes, Senator. I would be happy to have our teams follow up on that.

Chairman PETERS. I appreciate it. Thank you.

Thank you. I want to thank again our witnesses for joining us today. I am certainly grateful for your contributions to what is a very serious and a very important discussion, and I want to certainly thank Ranking Member Portman for holding this hearing with me here today.

I think today's hearing shed some new light on some serious problems of rising domestic extremism and violence and its relationship to amplified content on platforms.

We heard from our first panel earlier today about how user engagement and revenue generation are the primary incentives that drive product development and decisionmaking at your companies, and that the overall goals of growth and profit are always prioritized over the safety of users. That tradeoff, revenue over safety, has contributed to, unfortunately, some real-world harms, from horrific attacks and acts of violence motivated by extreme ideologies to our fundamental democratic process also being challenged.

I will be honest. I am frustrated that the Chief Product Officers, who all of you have a prominent seat at the table when these business decisions are made, were not more prepared to speak to spe-

cifics about your product development process, even when you were specifically asked if you would bring specific numbers to us here today, and that your companies continue to really avoid sharing some very important information with us. We have been working on this for quite some time and continue to be frustrated at the slow response, or the no response that we receive from you.

The testimony we heard today from both experts and from former executives as well as from the four of you have made clear the important work of the current trust and safety teams. It is simply not enough to address the problem. This problem continues to be with us today.

Although we head plenty of testimony about your companies' content moderation policies, what content gets removed and why and even how much you spend on safety measures, it is clear that those actions cannot effectively address this problem as long as the product development process and the revenue-based incentives do not change to make safety a higher priority in those structures.

We need to continue this important conversation. This will be the first of, I am sure, many conversations, and discuss possibly regulatory measures and changes to the incentive structures within your companies to build better practices, to limit the spread of harmful and extreme content before it is actually spread to users. Certainly we appreciate actions that are taken after the fact, but at that point much harm could already be released out into society with potentially catastrophic consequences. We all want to be ahead of the problem, not reacting to a problem that exists already in our society.

As Chairman of this Committee I will continue to work alongside Ranking Member Portman and Members of the Committee to find effective solutions to this growing homeland security threat. I certainly hope that each and every one of you will be part of that process to find that solution. We all need to be working together on this. It is very clear, the more we talk about this issue, the more we realize how complex it is, and it is going to take all of us putting our heads together and figuring out a path forward.

The record for this hearing will remain open for 15 days, until 5 p.m. on September 29, 2022, for the submission of statements and questions for the record.

With that, this hearing is now adjourned.

[Whereupon, at 5:25 p.m., the hearing was adjourned.]

APPENDIX

Chairman Peters Opening Statement As Prepared for Delivery
Full Committee Hearing: Social Media's Impact on Homeland Security
September 14, 2022
Panel I

In recent years, domestic terrorism, and specifically white supremacist, conspiracy related, and anti-government violence, has become one of our nation's greatest homeland security threats.

Last October, the Committee held a hearing to examine the role social media platforms play in the amplification of domestic extremist content and how that content can translate into real world violence.

We heard from expert witnesses who discussed how recommendation algorithms, ad targeting, and other amplification tools end up pushing increasingly extreme content to users because that type of content is what keeps people active on the platforms.

Unfortunately, because these platforms are designed to push the most engaging posts to more users, they end up amplifying extremist, dangerous and radicalizing content.

This includes QAnon, Stop the Steal, and other conspiracy theories, as well as white supremacist and Anti-Semitic rhetoric.

In some cases, this content may not necessarily violate a company's community guidelines. In other cases, even content that is in clear violation of company policies remains on the platforms, and is often only removed after public pressure. In both cases, this content does significant harm to our society and stokes real-world violence.

We have seen this happen time and time again. From the 2017 neo-Nazi "Unite the Right" rally in Charlottesville, Virginia that was organized using a Facebook event page, to the violent January 6, 2021 attack on the U.S. Capitol spurred to action in part by "Stop the Steal" content that repeatedly surfaced online, to the shooter who livestreamed as he massacred Black shoppers at a Buffalo supermarket, there is a clear connection between online content and offline violence.

Over the years, we have heard many explanations from social media companies about their content moderation policies, efforts to boost trust and safety, and actions taken to remove harmful accounts.

There is no question that those efforts are important. But there is a question of whether those actions are enough to effectively address the spread of dangerous content online and the resulting threats it poses to our homeland security.

The central question is not just what content the platforms can take down once it is posted, but how they design their products in a way that boosts this content in the first place, and whether they build those products with safety in mind to effectively address how harmful content spreads.

That is the focus of today's hearing where we will have the opportunity to hear from two panels of witnesses, outside experts, including former Facebook and Twitter executives, as well as current senior executives from Meta, YouTube, TikTok and Twitter, who are charged with designing social media products used by billions of people around the world.

The overwhelming majority of social media users have very little information about why they see certain recommended content in their feed, and there is very limited transparency into how social media companies balance their business decisions with the need for online safety, including what resources they invest into limiting the spread of harmful content.

Our goal is to better understand how company business models and incentive structures, including revenue generation, growth and employee compensation, determine how social media products are built and the extent to which current incentives contribute to the amplification of content that threatens homeland security.

For nearly a year, I have been pressing Meta, YouTube, TikTok and Twitter for more information on their policies to monitor and remove extremist and conspiracy content that advocates violence, as well as the relationship between their recommendation algorithms and targeted advertising tools that generate much of the companies' revenues, and the amplification of extremist content.

The companies' response to those inquiries have been incomplete and insufficient so far.

This morning, we will hear from two former executives and a technology journalist with social media expertise about the internal product development process and the business decisions these companies make, including tradeoffs between revenues and growth and their trust and safety efforts, as well as how they interact with foreign governments.

Later this afternoon we will hear directly from the Chief Product Officers of Meta, YouTube, and Twitter and the Chief Operating Officer of TikTok, the executives who are charged with making these business decisions and driving the strategic vision of the companies.

I look forward to a productive discussion.

Chairman Peters Opening Statement As Prepared for Delivery
Full Committee Hearing: Social Media's Impact on Homeland Security
September 14, 2022
Panel II

This morning, the Committee heard testimony from experts and former executives at Facebook and Twitter that provided important transparency and context for how many of the biggest social media companies operate.

Independent and accurate information about how companies balance competing priorities, or how they don't, who within the companies make those decisions, and how they build their products is incredibly difficult to find.

This morning's testimony shed light on many of the areas that this Committee, and the public, have questions about. I look forward to building on that testimony with our second panel of witnesses who can speak directly to what steps Meta, YouTube, TikTok and Twitter are taking to stop the spread of extremist content on their platforms. Thank you all for joining us today.

As we heard from our panel this morning, as chief product and operating officers, you play key roles in your companies' decision-making processes. You set the agendas for the product teams who are constantly updating the apps and developing new features.

You play a prominent role in setting priorities and determining what tradeoffs to make among those priorities, as product teams launch new features or make changes to your apps.

This is the first time executives in your positions have appeared before Congress, and I appreciate you joining us for this opportunity to hear directly about your roles at these powerful companies.

The platforms you are representing today reach billions of people around the world. Meta's platforms reach more than 3.6 billion people a month. TikTok has more than a billion users a month. YouTube reaches almost 2 billion people a month. And Twitter has more than 200 million monthly users.

That reach is massive and so is the influence your platforms wield.

Whether users are fully aware of it or not, the content they see on your platforms shapes their reality. And the business decisions you make are one of the main driving forces of that phenomenon.

This amount of influence may have a minimal impact on the average user of your platforms, but we have seen firsthand how quickly dangerous and extremist content can proliferate online, especially to vulnerable communities or users already on the fringe and alter how people view the world.

Conspiracies like QAnon and Stop the Steal, hateful ideologies like white supremacy and Anti-Semitism, and so many more examples of harmful content, pollute your platforms.

This extremist content can spread like wildfire, amplified by the recommendation algorithms and other tools your teams build to increase your companies' audiences and profits.

Extremists use the products you design to recruit and radicalize followers and plot attacks, including the January 6th attack on the Capitol, our democracy, and our nation.

There is no question that there is a relationship between social media amplification of this extremist content, and the rise we have seen in hate crimes and domestic terrorist attacks that mark one of the gravest threats to our homeland security.

Despite this serious threat, I am concerned that your companies have still not taken the necessary steps to limit the spread of the hateful, dangerous, and extremist content that has motivated real-world violence.

So that we all understand exactly the type of extremist content we are discussing today and how challenging this problem is to tackle. I'd like to take a moment to show a few examples.

[Video Plays**]**

This morning, we heard from former executives that your companies have no incentive to effectively address the problem this content creates or prioritize the safety of your users, as you build and introduce new social media products.

Instead, like any for-profit company, your incentives are to prioritize user engagement, grow your platforms and generate revenue.

I have asked you to appear before the Committee today to answer questions about your companies' incentives and priorities, how those incentives are reflected in how you compensate and promote your product development engineers, managers and other employees, and to provide important insight on your decision-making processes.

Thank you again for joining us today. I'm looking forward to this conversation, so that our Committee, and the public, can better understand this serious problem and how it threatens the safety and security of our nation.

Opening Statement
Ranking Member Rob Portman
COMMITTEE ON HOMELAND SECURITY & GOVERNMENTAL AFFAIRS
"SOCIAL MEDIA'S IMPACT ON HOMELAND SECURITY: PART I."
SEPTEMBER 14, 2022

AS PREPARED

Thank you, Chairman Peters, for holding this hearing. This past Sunday, we observed the 21st anniversary of the 9/11 terrorist attacks. Over the last two decades, the United States has adapted to combat the most pressing threats to our nation's security. The advent of social media has added a new dimension to the ever-evolving threat landscape and created new considerations for combatting terrorism, human trafficking, and other threats.

During last October's hearing on how algorithms promote harmful content, I examined how social media business models contribute to the amplification of terrorism and other dangerous activities. Since then, the Committee has identified ways in which social media companies' product development processes tend to conflict with user safety. Whistleblower testimony has revealed that in numerous occasions, the leaders at social media companies were aware that certain platform features increased threats to user safety and chose not to mitigate such concerns.

It is unfortunate that the American public must wait for whistleblower disclosures to find out about ways in which platforms are knowingly and unknowingly harming their users. The lack of transparency in the product development process, the obscurity of algorithms, and misleading content moderation statistics create an asymmetric information environment, in which the platforms know all, yet the users and policymakers know very little.

One consequence of this lack of transparency is related to the Chinese Communist Party. I have serious concerns about the opportunities that the Chinese Communist Party has to access TikTok's data on American users. There are over 100 million Americans, including 40 million under the age of 19, who use TikTok. This TikTok data remains vulnerable to the Chinese Communist Party both as the CCP tries to exploit its access to U.S. data and exert influence over the content that U.S. users see. For example, despite moving U.S. user data to servers in the United

States, TikTok and ByteDance employees in China retain the ability to access this data.

Also, we learned yesterday from Senator Grassley's opening statement in a Senate Judiciary Committee hearing with the Twitter whistleblower that Twitter failed to prevent Americans' data from being accessed by foreign governments. In fact, Senator Grassley spoke about how several Twitter employees were actually foreign agents of India, China, and Saudi Arabia, which is deeply concerning and speaks to why Congress needs more information from platforms on their securitization of user data.

Another consequence of poor transparency relates to content moderation. While I recognize that content moderation is a key component to creating safe platforms for users, it cannot be the silver bullet. Transparency reports released by companies often only detail the amount of content that has been removed for violating company policy. However, these reports do not account for violating content that is left up on the platform and goes undetected.

It also doesn't account for content that is incorrectly censored, as we often see with many conservative voices on social media. I, like many of my colleagues, have been critical of the political biases held by Big Tech platforms, which have resulted in systematic takedowns of accounts that hold ideologies with which the left and liberal media disagree. These takedowns are often done under the guise of combatting "misinformation" or "hate speech," when in fact they are really just combatting conservative viewpoints that conflict with their own. Any steps taken to address the impact of social media on homeland security must account for First Amendment protections and safeguard free speech.

For us to have a responsible conversation about the impact of harmful content on American users and homeland security, we need to talk about how current transparency efforts have been ineffective. Congress must enact legislation that will require tech platforms to share necessary data so that research may be done to evaluate the true extent of how harms from social media impact Americans.

That is why I have been working with Senators Coons and Klobuchar to create bipartisan legislation to do just that. The *Platform Accountability and Transparency Act* would require the largest tech platforms to share data with

vetted, independent researchers and other investigators so that we can all increase our understanding of the inner workings of social media companies and later regulate the industry based on what we learn.

I look forward to hearing from our witnesses and thank Chairman Peters for holding this hearing.

Opening Statement
Ranking Member Rob Portman
COMMITTEE ON HOMELAND SECURITY & GOVERNMENTAL AFFAIRS
"SOCIAL MEDIA'S IMPACT ON HOMELAND SECURITY: PART II."
SEPTEMBER 14, 2022

AS PREPARED

Thank you, Chairman Peters, for holding this hearing to examine the impact of social media on homeland security. We had a productive hearing this morning on the topic, and I look forward to this hearing with industry representatives.

As of 2021, almost 72 percent of Americans used social media. And while social media has offered unprecedented connectivity, it has also raised serious concerns for our children, our civic culture, and our national security.

Terrorists, violent extremists, drug cartels, criminals, and other dangerous forces have used social media in furtherance of their violent goals. Perhaps the most concerning consequence of social media is the ability for our adversaries to exploit platforms to harm Americans for their own geopolitical gain. It is imperative for policymakers to identify and thwart China's exploitation of technology that furthers its espionage campaigns. In this second panel today, we will discuss China's influence over TikTok, a social media app that at least one-third of Americans use.

As the lead Republican on this Committee and previously as the Chairman of the Permanent Subcommittee on Investigations, I have been focused on China's malign activities. In 2019, I led a year-long bipartisan investigation which found that China recruits U.S.-based researchers to steal taxpayer-funded intellectual property and research for its military and economic gain through its Thousand

Talents Program. Following this report, I introduced my bipartisan legislation, the *Safeguarding American Innovation Act*, which seeks to stop U.S. taxpayer-funded research and intellectual property from falling into the hands of the Chinese Communist Party or CCP. Two months ago, I also issued a new report detailing China's efforts to target, influence, and undermine the U.S. Federal Reserve. China has a pattern of economic and cyber espionage, and social media is just another opportunity. I am highly concerned about TikTok and how China may be leveraging their influence to access the platform's data on Americans.

Chinese law requires all companies operating under its jurisdiction to, in essence, allow the Chinese Communist Party to access every piece of data collected. Any company that refuses to comply with the CCP's demands is subject to severe consequences. Therefore, since both TikTok, and its parent company ByteDance, are located in mainland China, we are left to assume that TikTok's user data could be accessed by the Chinese Communist Party.

That means that the CCP may have access to one hundred million Americans' personal and proprietary information. As our expert witness this morning testified, China's access to user data will allow it to extend its malign agenda and build dossiers on American citizens. The overwhelming popularity of the app with America's youth will allow China to collect never-before-accessed troves of data on our children—the future generations of American leaders.

But the challenges that social media poses to our children are not limited to TikTok. We continue to see the proliferation of child sexual abuse material online. I have been at the forefront of this fight for years. I am proud that my Stop Enabling Sex Traffickers Act was signed into law in 2018. This was the first bill to reform Section 230 by removing barriers to both criminal prosecution and civil suits against websites that knowingly facilitate online sex trafficking. Because of this change in law, courts are beginning to affirm that Section 230 cannot shield for internet companies when they fail to respond to images of child exploitation and continue to profit from exploitation on their platform. A specific case against Twitter is now being considered by the Ninth Circuit Court of Appeals and will show if the law needs to be expanded in order to protect children.

But it's not just Twitter, the fight continues on other platforms that are used to exploit children. Meta announced earlier this year that they would not report all

explicit images of children and would instead "err on the side of an adult" when moderating explicit images of could-be children. In other words, when the age of an individual in a sexual image is uncertain, content moderators are told to put their thumbs on the scales of that individual being an adult. This is outrageous. Let's be clear what we are talking about: child sexual abuse material are images of a minor's rape or exploitation. And somehow Meta has decided that these should not be referred to law enforcement? The National Center for Missing and Exploited Children has made it clear that images must be reported if they appear to involve a child so that law enforcement can intervene to stop the abuse and prosecute perpetrators.

I worked with colleagues across the aisle to craft SESTA narrowly so that it would be focused ending trafficking and exploitation online. But it may in fact be too narrow if companies continue to turn away from keeping this exploitation off of their platforms. I hope my colleagues will take up the challenge of revisiting SESTA and tightening the standard so that entities showing a reckless disregard for the exploitation of children are held accountable. I am ready to be an ally in this fight even after I leave the Senate.

I look forward to discussing these matters, especially regarding how product development processes appear to be at odds with user safety, as well as the need for more detailed transparency from the companies. Thank you, Chairman Peters, for holding this hearing.

Written Testimony of Alex Roetter

Former Senior Vice President of Engineering, Twitter

Before the U.S. Senate Committee on Homeland Security and Governmental Affairs

On "Social Media's Impact on Homeland Security: Part I"

Wednesday, September 14, 2022

Good morning Mr. Chairman and Members of the Committee. Thank you for holding these hearings and for inviting me to provide testimony.

We live in a world where an unprecedented number of people consume news and information based on what is fed to them by a small number of massively influential social networks. A small group of people run these companies and have substantial power over shaping reality for billions of people. As others in front of this committee noted, viral content, misinformation, and disinformation can propagate via these platforms on a scale unseen in human history. The companies responsible for amplifying viral content, while focused solely on maximizing their profits, have either not taken responsibility for the effects this content has on society or paid lip service to them.

Regulators must understand these companies' incentives, culture, and internal processes to fully appreciate how resistant they will be to changing the status quo that has been so lucrative for them. Without increased transparency and a change of incentives, we should expect the same behavior and continued lack of meaningful, quantifiable progress on these problems.

In over 20 years of working in Silicon Valley as an engineer and an executive, I have seen first-hand how several companies prioritize features, launch products, and optimize their metrics. I was an early engineer on the Google syndicated ads product. Then, I started the ads engineering team at Twitter, growing it to $2.5B in annual revenue, before taking over as SVP of engineering for all development efforts at the company.

Everything I will talk about today is an inevitable result of the culture of the companies and the incentives they face. While it may be easy to criticize specific leaders, that is the wrong target. As a thought experiment, we would not see a meaningful change even if we were to replace leadership at all these companies. What must change is the incentive system that makes the companies act the way they do.

Today I will explain the internal systems essential to understanding how things can or won't change, and the types of data we should seek. Finally, I will recommend how to obtain and analyze that information.

Product Development

The product development life cycle works as follows:

1. Small teams of product managers, engineers, and designers brainstorm ways to meet and exceed specific internal goals that are measured with high precision. Metrics assigned to them come from various user growth and retention metrics (churn, engagement, time on sight, new user signups) and financial health metrics (revenue, ad price, click-through rate). Underlying these goals is a mission to drive revenue and maximize shareholder value and stock price. Typically, other metrics (user safety, etc.), if present, are a distant second in terms of importance.
2. To measure their effectiveness, these teams use an experimental system to launch their new changes to a small percentage of traffic (1% or 0.1% are typical values). I spearheaded the development of this system while at Twitter. A very similar approach is also in place at Google.[1] I believe other companies also operate related systems. This system logs a slew of data for every live experiment. Teams use this data to show per-experiment effects on various user and revenue metrics. Noticeably absent were any values tracking impacts on trust and safety metrics. For example, I never once saw any indication of how a given experiment affected any types of manually reviewed or actioned content, e.g., "Did a given experiment increase the reach of content later identified as hate speech?"
3. Executives will, typically once a week, as was the case at Twitter, hold an experimental review meeting and review all active experiments. After reviewing the data, they would shut down experiments, ramp them up to a more significant percentage of users, or fully launch them to 100%. Product and engineering ran these meetings. To the extent other functions like legal, finance, or trust and safety are present, they play a minor decision-making role compared to product, engineering, and sales leaders. Frequently they are not even present (and not viewed as mandatory attendees for a quorum).

Company Culture

It is essential to understand the culture of these companies in order to appreciate how the product development life cycle is applied. In terms of culture, they are hierarchies, with the "builders," namely, engineers, product managers, and designers, held in the highest regard. The strong norm in Silicon Valley is to not get in the way of "builders." This stems in part from the fact that the founders of these companies come from "builder" backgrounds. Other functions (e.g., legal, finance, compliance, etc.) are viewed very skeptically, and the bias is to make sure "corporate bureaucracy" doesn't slow down the building and releasing of new products or features. Perhaps no single illustration of this is more famous than Facebook's old motto, "Move fast and break things." A significant premium is placed on not slowing down the pace of development through reviews or bureaucracy.

---

[1] https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/36500.pdf

Review & Promotion System

The companies evaluate individuals via quarterly, semi-annually, or annual performance cycles. In this process, individuals typically write up self-reviews describing their accomplishments, reviews of their peers, and reviews up and down the management chain. Reviews heavily consider the individual's impact on their team and company. The way impacts are described is first and foremost in terms of what products or changes an employee built and what results they achieved. Benefits are measured most prominently by user engagement and financial metrics. Reviews determine the stack ranking of employees relative to one another, eligibility for cash or stock bonuses, and are inputs to the promotion process. Reviews are also considerations when companies make downsizing decisions in leaner times.

The promotion systems that I have observed highly emphasize impacts on key company metrics as well. Engineers highlight the effects that their work had on the company in as quantifiable a way as possible to promotion review committees. The most effective way to do this is to show an impact on financial or user engagement metrics. To the extent that you can get promoted for something on the trust and safety or privacy side of the house, a minority of builders get promoted this way, and the impact isn't valued as heavily as the core drivers of the company's growth. For example, the fastest way to get promoted as a front-end engineer on a core product feature team is to show that you made changes that drove more viral growth or increased revenue. There is an inherent tension between this progress and improving trust and safety. The former nearly always wins over the latter.

Company Behavior

If one understands these companies' incentives, product development processes, and internal culture, it should be no surprise to see them respond the way they do to external pressures. While attempting to convince the public that they are making serious investments in these areas, we see them continuing to be unaccountable for any measurable results against the questions of interest to this committee. They are incented to delay any oversight while they continue to build some of the most valuable companies in the world.

To change things, we should not accept certain types of answers as sufficient, and we must demand more transparency. Finally, we should be able to observe if their incentives have truly changed by pressing for transparency on their internal processes and testing to see if they are acting as one would expect against a new set of incentives.

Historical Data Sharing

Historically companies have released selective statements designed to convince people that they are taking the problem seriously in response to public or regulatory pressure. Examples of these statements from the companies include: "We are spending $X on this problem," "We are employing Y people," or "We remove Z pieces of content every day/week/month." These

numbers lack context (i.e., a denominator) and are shared because the numbers seem large in absolute terms. They are designed to convince external parties that they are tackling the problem. These numbers are significant because they come from some of the largest companies in human history. Alphabet/Google and Meta/Facebook are in the top ten companies by market capitalization globally. Any single spending or investment figure will look large, but the numbers are less impressive in the context of how much they spend on other initiatives (new product initiatives, stock buybacks, stock-based compensation, etc.).

Similarly, how many pieces of content are actioned or removed tells you nothing about what fraction of content that represents or the reach of the content (which is the crucial part). Finally, simply saying they are investing a lot is not the point. The metrics that the public receives should be measuring outcomes (e.g., as proposed by an external group), not simply about the size of the investment. If companies share only how much they are investing, it is too easy to hide a complete lack of measurable results. Even a significant investment, when made in the context of the current incentive system, should not be expected to make a meaningful improvement.

Transparency

Instead, regulators should demand transparency that gives a picture of how the system is behaving. What matters to understanding a network is not content creation but the distribution and engagement of individual pieces of content. It is impossible to understand the network without understanding the algorithms that govern the spread of content and drive interactions on that content. Understanding this requires more transparency than companies have historically shared.

If you want to understand what is happening more clearly, I recommend assembling an independent group of researchers and data scientists. Task them with coming up with the right questions to ask and a list of data they need to answer those questions. Then, fund a team of third-party individuals to constantly analyze this data, publish their findings openly, and recommend new questions to answer and the resulting datasets required.

The government already holds the private sector accountable in other technically demanding fields. Examples include building code inspections, NTSB crash investigations, FAA aircraft certification, and SEC financial crime enforcement. Consider also the role that third-party accounting firms play in generating audited financial statements of public companies. A functioning stock market is impossible without assurance that companies' financial statements are accurate. At the same time, companies need to keep proprietary information confidential. This is solved by third-party accounting firms that can audit and certify company financials. These firms successfully balance the public's need for accountability with the company's need to protect private and confidential information.

Due to their current incentives and how lucrative the status quo is to them, there will be pushback on any requests for more data. This pushback will come in the form of arguments

about user privacy, company confidentiality, the expense of these requests, and a myriad of other concerns. There are straightforward answers to these objections, and they can be overcome if someone is genuinely incented to overcome them.

The more complex the data requests, the more opportunities companies will have to obfuscate their answers or figure out how to make them unhelpful. There are many ways to do this, such as sharing numbers without sufficient context, sharing summary statistics such as averages when we look at highly skewed distributions where simple averages are not very informative, and many other obfuscating techniques. We should not underestimate the strength of the companies in figuring out how to avoid such requests for data. To understand their power, consider that just one team at one of these companies (the legal team at Google) has approximately the same number of lawyers as the total employee count of the entire Federal Trade Commission. This isn't even considering the amount of money they spend on external law firms.

So, as important as transparency is, it is necessary but not sufficient to tackle these problems. Other policy and legal experts have testified to various proposals that could change these companies' incentives, and I defer to those experts on that front. But until this is done, and until we change the fact that virality, attention, eyeballs, and clicks are the things the companies care about above all else, all the data sharing in the world will not address the problem.

Incentives

One way to measure whether incentives are changing is to compare companies' behavior when they believe content quality does matter to their financial performance against cases when they do not. One simple litmus test would be how companies handle advertisers' content vs. non-advertising (organic) content. If propagated by advertising messages, hate speech or inciting violence would be terrible for the company and likely to harm their core business. Having offensive ads, or ones that violate their ad policies could quickly and materially harm their financial performance. This is why most advertising systems I am aware of place ad review (automated and sometimes manual) as a step that has to occur before the new content ever makes its way into the live system for consumption by real users. Contrast this to how organic content goes live. Organic content is allowed to go live and potentially go viral instantly. Only after the fact, when and if the team gets around to it, is there any review or take-down action. In many cases, societal harm has already been done, while any interim set of impressions or engagement is good for the company. Today, if a company changes its behavior to prioritize content review before new content goes live, that would harm its financial and user growth performance. Of course, this would not be true with the right incentives.

To give an example of how incentives can change what content is recommended, let me share the following example. It is public information that the Chinese government is an investor and board member in Bytedance, Tiktok's corporate parent. Of course, the Chinese government has a series of incentives other than profit maximization and user engagement. As a result, I have come to understand the following: the Tiktok algorithm pushes educational science, engineering,

and math content on Chinese youth while pushing a feed containing twerking videos, misinformation, and other destructive content to US children. In addition, they enforce daily time limits on usage for Chinese students.[2] Even worse, US companies' profit maximization motives mean that they are essentially doing the work of the Chinese government for them.

Conclusion

Any suggestion for more data to share, or checks put in place before content can go live, will be met with strong objections. While this would certainly be an investment in computational power and software and have consequences that need to be considered, it is possible, given the resources at these companies' disposal, if they had the will to do so. Certainly, they have proved responsive to regulation by other governments (for example, in Europe). Many privacy and confidentiality concerns could also be addressed in a way that is not overly demanding. As seen with the advertisement content example, companies happily change their behaviors when it is in their best interest. That said, as long as their incentives are aligned to fight change, they will have endless reasons to oppose any such change.

Given what we know about companies' incentives, internal processes, and culture, we should not expect meaningful progress voluntarily. They will continue to benefit from the rampant spread of viral content online without feeling any significant downside from any harmful effects it may have on society. With the proper transparency, third-party oversight, and regulatory environment, I hope they will start earnestly tackling and making real measurable progress against these problems they have been so integral to exacerbating.

[2] https://www.cnn.com/2021/09/20/tech/china-tiktok-douyin-usage-limit-intl-hnk/index.html

Testimony of Brian Thomas Boland
Former Vice President Partnerships Product Marketing, Partner Engineering, Marketing, Strategic Operations, & Analytics, Facebook

Before the United States Senate Committee on Homeland Security and Governmental Affairs on "Social Media's Impact on Homeland Security: Part I"

Submitted on September 12th 2022

Good morning Mr Chairman and Members of the Committee. Thank you for holding these hearings that cover such important issues for our nation. And thank you for inviting me here today to provide testimony on my experiences as a senior executive at Facebook, now known as Meta. For the last few years I have grown increasingly concerned about the roles that Facebook, Instagram, YouTube, Twitter and TikTok play in driving the growth of misinformation, extremism and generally harmful content.

I worked at Facebook for 11 years in a variety of leadership roles helping to build products for organizations and businesses who use Facebook products and services. In my roles, I helped to shape product and marketing strategies for a broad range of products inducing advertising, news, media, games, Workplace, internet access and more. During my tenure at the company I worked with the most senior executives at the company and was deeply embedded in the product development process.

In the last two years of my time at the company, the CrowdTangle team and product was a part of my organization. CrowdTangle is a tool that provides limited, albeit industry leading, transparency into public News Feed content on Facebook. What finally convinced me that it was time to leave was that despite growing evidence that the News Feed may be causing harm globally, the focus on and investments in safety remained small and siloed.

The documents released by Frances Haugen, the Facebook whistleblower who last fall testified in Congress, highlight issues around polarization globally and the power of Facebook to lead people down a path to more extreme beliefs. These papers demonstrate thoughtful, well researched documentation of the harms that greatly concerned me. This research was done by highly skilled Facebook employees - experts in their field - and was extensive.

And rather than address the serious issues raised by its own research, Meta leadership chooses growing the company over keeping more people safe. While the company has made investments in safety, those investments are routinely abandoned if they will impact company growth. My experience at Facebook was that rather than seeking to find issues on the platform first they would rather reactively work to mitigate the PR risk for issues that came to light.

I have come to believe that several circumstances have put Americans at risk from the content on these platforms.

The first is the growth over safety incentive structures that lead to products that are designed and built without a focus on user safety. The next is the unprecedented lack of public transparency available from these platforms to analyze content and understand the impact from these tools. And finally, the lack of clear oversight for the business practices of these companies.

The reality is that for all of the debate about whether social media is good or bad - the truth is that we really don't know. The good news is that I believe with the right incentives in place and rules around transparency we can develop a better understanding of these issues and take steps to mitigate the harms.

Today I hope to shed light on the product development process, internal and external incentive structures for these organizations and the critical importance of transparency. I do so with an aim to guide us to improvements in how these businesses work - so that we can capture the many wonderful benefits that they bring to society while mitigating the harms that are broadly ignored today.

Prioritization and Resources

People criticize tech companies for not hiring more people in safety roles considering the incredible profits they report each quarter. Unfortunately, it isn't that simple. In any given year and in any given city there are only so many qualified engineers to hire at a company. And within that limited number even fewer will accept roles and join. This creates an internal scarcity where even though you may have the monetary budgets to hire anyone you want - you simply can't. The scarcest type of talent is engineering talent.

And so the CEO, Mark Zuckerberg at Meta, must define the priorities for the company and choose to assign engineers to the work he believes to be most important. Because of this, the best way to understand what truly matters for a tech company is to look at their relative staffing - whichever groups are allocated the most engineers are the highest priority.

Despite claims for a deep focus on safety and user integrity my experience was that these teams were often under-resourced. Further, safety and user integrity issues were isolated to a specific team focused on those issues rather than distributed among the various product teams. As such, safety and user integrity issues were not top of mind for nearly all development teams.

Business Incentives and User Safety

At Meta, employees are rewarded through a ratings system that measures their performance twice a year. Each employee is rated on where they stood with respect to the measurable goals that were set for them at the beginning of that six month period. For most product teams, those expectations are highly quantifiable product metrics. These goals focus on the growth of their product or feature - are more people using it, is revenue growing, etc.

One good way to understand what is important to these companies is to look at the standard metrics that they require for any product test. While user engagement and revenue metrics may be standard, user safety metrics are not. Until user safety is a core part of the development and testing process it will remain an afterthought.

As an illustration, when it became clear that the future for Facebook apps was on mobile devices, Mark Zuckerberg required that every product review meeting would first include designs of that product on a mobile device. In the first meeting without one, he sent the team away until they could come back with a mobile design. Everyone else came in with mobile designs after that.

Similarly, when the advertising business was slowing around the IPO, Mark mandated that advertising was a part of the goals for all teams - not just the ads team. This marked a shift at the company where it became clear that each team needed to think about, design for and measure their growth in ads revenue. This type of structural change for the teams led to a significant shift in focus and growth in advertising revenue.

In the same way, each team should have clear goals on user safety and reducing the conditions for misinformation. If teams are broadly measured and rewarded based on these metrics you would expect a significant improvement in understanding of the issues and steps to take to mitigate them.

To be fair, this focus on growth is how these companies are incentivized today. The stock market rewards growth with an increased share price while remaining unaware of the risks lurking below the surface. And there aren't rules today that would create an incentive to add safety or misinformation metrics to more teams at the company.

Defining Metrics Matters

In discussions of harmful content and misinformation the framing of the conversation and definition of specific metrics matters greatly.

It is important to avoid discussions on percentages instead of absolute numbers. The scale of these platforms is so massive that very small percentages of users turn into very large absolute numbers. If Facebook can reach 222 million people in the United States, 1% of that audience is 2.2 million people.

Another example is whether a company discusses "average prevalence" vs "p99 prevalence".

The issue with hate speech, misinformation and other harmful content isn't whether the average person is seeing more or less of it, but whether there are pockets of people receiving significantly more of it than others.

If the average person is seeing less hate speech but people at the fringe who are more susceptible to those ideas see an avalanche of it you can have a significant problem on your hands.

So rather than looking at averages, it is particularly helpful to look at p99 metrics. The p99 metric tracks the 99th percentile of most hate speech filled (or other content you define) News Feed. From this metric you can better understand the impact at the fringes and not focus on a watered down and much less useful average.

A strong discussion of this can be found at:
https://twitter.com/samidh/status/1450105897431175169

Understanding the Algorithms

I believe that we don't have a problem with an explosion of the creation of misinformation but rather a hyper amplification and bias problem. These platforms aren't creating a marketplace of ideas but rather skewed selections of reality. As more companies are creating algorithmic experiences, what we individually experience is highly curated.

There are studies that argue both sides of this curated experience - some stating that this algorithmic content brings us closer together and some that it is driving divisions in America. The truth is that nobody really knows.

Contrary to popular belief, technology companies cannot accurately predict what their algorithms will do. At Facebook, a robust A/B testing system was created so that engineers could observe and measure what happened when changes were made to the system. These A/B testing systems record a handful of important metrics.

During the product development process, teams will run a variety of tests on new features they would like to release. These experiments will track an array of metrics - some focused on user engagement with the features and other site metrics, and others focused on impacts to company revenue. People are rewarded for growth metrics, not safety metrics.

In many ways, this is similar to crash testing a car. Despite years of work to design and develop new cars, they are incredibly complex machines and when they encounter the real world they behave in ways that we don't expect. By crash testing cars, we can gather a significant amount of data on what happens and work to fix current issues and improve future designs. Further, the NHTSA gathers data on real world incidents involving cars for the same purpose - to uncover existing issues, mitigate them and improve future safety.

The challenge that we face here is that tech companies are deciding whether to crash test anything, what to test, when to test, whether to fix the product and whether to disclose it to the

public. There is no independent public ability to study these platforms and develop an understanding of their impacts.

These platforms function as a black box by design. As such researchers, academics and human rights organizations have an incredibly limited amount of data that they can use to protect Americans and people around the world. Once researchers and academics have a better view of what is happening on these platforms we can have a more informed public debate about the issues discovered and how to best mitigate them.

Lack of Transparency

One of the turning points for me was the summer of 2020. As political and race-related divisions in this country grew, I saw those same divisions reflected in the public data from CrowdTangle, a tool Meta had purchased to help companies understand how people engaged with their content. CrowdTangle was also utilized by researchers and journalists to find problematic content on Facebook and understand how people engaged with it.

As the public scrutiny intensified, Meta attempted to delegitimize the CrowdTangle-generated data. It rejected strong internal proposals to take industry-leading actions to increase transparency.

In some ways the internal resistance is understandable. There is no requirement for transparency. And often the limited data that is used can create significant PR headaches for the company - something that they aim to avoid. This leads us to the path where we are today - there is minimal investment in transparency and efforts like CrowdTangle are diminished. And the other tech platforms generally haven't taken steps to provide transparency.

Further, companies like Meta shift to a "Widely Viewed Content Report" which only provides the most seen content on Facebook in the US. The report delivers high level distribution which creates the same challenge as the average metrics discussed earlier. We don't get a view into what really is happening in pockets on Facebook - just a view of the highest level averages.

To solve this, we must require platforms to move beyond the black box with legislation like The Platform Transparency and Accountability Act.

Responsibility to Shareholders

I believe that these tech platforms have a fiduciary duty to their investors to better understand and mitigate the harms created by their platforms. Due to the lack of transparency in the platforms today, we understand little of what is actually happening on them and their impacts to homeland security and the depths of misinformation. Since the companies underinvest in investigating potential harms and fail to provide meaningful useful data to researchers and human rights organizations it is highly likely that severe vulnerabilities lurk in their systems.

These lurking risks could pop up at any time and cause a negative impact to investors in the company. We saw this type of price decline following the release of the stories and documents from Frances Haugen last fall.

Despite Meta's claims of significant investment, in context those investments are small. During the period of time where they claimed $13 billion in safety investments they spent $50 billion in stock buybacks. The commitment to short term shareholder value, I believe, is creating a significant future risk for shareholders.

Should pressure continue to grow from increased revelations and understanding of harms, ESG investors could divest from these platforms along with many other socially conscious investors. By failing to deeply investigate and understand their own platform, Meta is putting their shareholders at risk.

Conclusion

It is important to note that there are incredible benefits from these platforms. Social media has enabled families to stay in touch, allowed friends to remain connected across geographies, empowered people struggling with illness to find support, entertained us, given marginalized people voice and at times informed us.

Additionally, most of the people that I worked with are well intentioned, high integrity people who aspire for positive societal outcomes. Despite that however, the incentives, lack of transparency and the risks these platforms create have reached the point where regulation is greatly needed.

We can hope for a better future with these platforms where we see outsized benefit from them rather than living in a constant fear of the next unseen crisis building beneath the surface.

In order to get there we need to start with three steps:

> First, incentivize these companies to study, measure, prioritize and invest in innovations that can stem the tide of disinformation, misinformation and harmful content. Left alone these companies will continue to pursue investments and metrics that benefit their growth and the short term value of their stock. This is the rational approach with the incentives and rules that we have today.
>
> Second, legislate transparency for public content on tech platforms. The Platform Transparency and Accountability Act provides a framework for required transparency by tech companies. The framework is strong but I would further expand it to ensure access of public data to human rights organizations and journalists.
>
> Third, clarify which agency or agencies have oversight of these platforms and ensure

> that those agencies are funded to do the work. It isn't enough to simply open up the public data on these platforms - we need to ensure that as we better understand what is happening on these platforms we can change their incentives to better serve the public. If the incentives don't change, the results won't change despite improved transparency.

Our goal here is a common one - to ensure that we help American innovations like Facebook, Instagram, Google and Twitter to thrive and better incentivize them to serve their communities.

Thank you.

September 14, 2022

Geoffrey Cain, Senior Fellow for Critical Emerging Technologies, Lincoln Network

Written Testimony Before the Homeland Security and Governmental Affairs Committee
Social Media's Impact on Homeland Security

Chairman Peters, Ranking Member Portman, and Members of the Committee:

It is an honor to be invited to testify here on social media's impact on national security. Today, I will talk about one of the greatest technological threats facing our homeland security and democracy: TikTok, the social media app owned by the Chinese parent company ByteDance.

TikTok is the fastest-growing social media app ever and is expected to hit 1.8 billion users by the end of this year. Known for its fun and digestible video snippets, the app is enormously popular among celebrities and Generation Z users. It goes to great lengths to appeal to the sensibilities of the American market by loudly proclaiming progressive, democratic, egalitarian values. It posts messages on social media supporting inclusivity, diversity, LGBTQ+ rights, and pro-life causes.

All this is a distraction from the reality behind TikTok's parent company in China, called ByteDance. As an investigative journalist in China and East Asia for thirteen years, I have been detained, harassed, and threatened for my reporting on Chinese technology companies. ByteDance and its subsidiary TikTok have sought to distract us from well-documented ties to the Chinese Communist Party.

In internal meetings, ByteDance's leaders have extolled communist party virtues, pledging their absolute loyalty to a totalitarian government. The celebrity and cat videos are a distraction. TikTok is a major threat to our national security and freedom of discourse. Its parent company has censored Uyghur refugees who have suffered under a genocide now being carried out in China's western region of Xinjiang, as well as other heinous crimes.

TikTok claims that, despite reporting to executives from a company in the People's Republic of China, called ByteDance, it keeps the data of American and global users on TikTok separate from ByteDance's business operations in China. There, the leaders of the Chinese Communist Party (CCP) have repeatedly declared their hostility to our democracy and way of life.

Today, I will show you how TikTok has orchestrated a campaign of distraction and deflection to mask the alarming truth. Americans face the grave and unprecedented threat of software in our pockets that contains powerful surveillance and data-gathering capabilities, owned by private companies that must nevertheless comply with the dictates of the CCP, which has signaled its ambitions to assert global jurisdiction over private companies everywhere as a condition for doing business in China. TikTok is a disaster waiting to happen for our homeland security and the privacy of our citizens.

TikTok's Troubled Emergence in America

TikTok's explosive growth in America has been a troubling story of conflicting statements, broken promises, hollow reassurances, and profiteering complacency. We have TikTok executives here

today. According to their internal guidelines, if you ask them about the influence of their Chinese parent company ByteDance over the American product TikTok, executives must deceptively tell you that ByteDance is a separate parent company and that you should talk to ByteDance instead. They will attempt to confuse you, claiming that TikTok takes a localized approach, hiring local moderators, implementing local policies, and showing local content.[1]

TikTok executives will not tell you the real story about their ties to the world's most sophisticated and dystopian police state. They will not tell you about a Beijing-based engineer known as the "Master Admin" who had, according to leaked audio from internal company meetings, "access to everything" on the app.[2] Their employer does not give them the authority to tell the full truth. A leaked, 53-page public relations document that TikTok executives call their "Master Messages" tells employees to "Downplay the parent company ByteDance, downplay the China association, downplay AI."[3] They won't tell you that they report to ByteDance, and that ByteDance reports to the CCP.

The relationship between TikTok and ByteDance has been a problem from the start. Eight years ago, in 2014, the Chinese arm of the major Silicon Valley venture capital firm Sequoia Capital invested in TikTok's parent company ByteDance in China with a $500 million valuation, paving the way for its expansion into America. Sequoia Capital's China arm has been building ties to China's party elite—for example, by later hiring the daughter of a member of the CCP's powerful Standing Committee.[4]

TikTok's fast expansion into the American market was only possible because China has rigged the market, offering ByteDance vast market protection in China while banning competing American social media apps Facebook, Instagram, Twitter, and Google. In 2016, ByteDance initiated a $1 billion purchase of a Chinese-based music streaming company called Music.ly, popular among American teenagers. Nine months later, ByteDance merged Music.ly with its own software, cementing TikTok as the American version of its Chinese app, Douyin.

From the start, the acquisition was concerning. The *Financial Times* reported that ByteDance did not seek approval from the Committee on Foreign Investment in the United States (CFIUS), the government body that reviews foreign inflows into strategic and sensitive businesses in the U.S. According to the report, ByteDance executives believed they did not need to begin a CFIUS review because they were acquiring a Chinese company, not an American one.

This was a decision of questionable legality. Music.ly had an office in Los Angeles, placing this acquisition under the jurisdiction of CFIUS. CFIUS still has the authority to investigate and reverse the acquisition, which would force ByteDance to sell TikTok and terminate its American operations.

Other alarm bells sounded in the early days of TikTok in America. In 2018, ByteDance's and TikTok's founder and previous CEO, Zhang Yiming, wrote a letter promising Chinese regulators

---

[1] Chris Stokel-Walker, "Inside TikTok's Attempts to 'Downplay the China Association,'" *Gizmodo*, July 27, 2022, https://gizmodo.com/tiktok-master-messaging-pr-playbook-china-music-1849334736.

[2] Emily Baker-White, "Leaked Audio from 80 Internal TikTok Meetings Shows That U.S. User Data Has Been Repeatedly Access From China," *Buzzfeed*, June 17, 2022, https://www.buzzfeednews.com/article/emilybakerwhite/tiktok-tapes-us-user-data-china-bytedance-access.

[3] Chris Stokel-Walker, "Inside TikTok's Attempts to 'Downplay the China Association.'"

[4] Juro Osawa and Shai Oster, "Sequoia Capital's China Arm Employed Daughter of Politburo Member," *The Information*, September 9, 2022, https://www.theinformation.com/articles/sequoia-capitals-china-arm-employed-daughter-of-politburo-member.

that his company would follow "core socialist values," would introduce these "correct values into technology and products" and would ensure his products promoted the CCP's agenda. These "values," he wrote, included "strengthening the work of Party construction," "deepening cooperation with official Party media," and strengthening "content review" in line with these Party "values."[5]

ByteDance's public statements in China should be cause for alarm. American government employees, military personnel, and people in sensitive and strategic industries use TikTok. Because China has little separation between private business and the government's authoritarian ideology, ByteDance, like all Chinese companies, maintains an in-house Communist Party Committee mandated to enforce the political loyalty of employees in China. At a committee meeting in April 2018, ByteDance executives declared that their social media algorithm must be informed by the "correct political direction" and that content should "highlight socialist core values."

ByteDance engineers in China, not America, developed the algorithm that TikTok used in America. TikTok engineers employed in Mountain View, California reported to senior executives in China, where the company's Communist Party Committee set the course of ByteDance products.[6]

Researchers from the Citizen Lab, an internet research institute at the University of Toronto, found that the Chinese app Douyin and the American version TikTok use the same base code, but alter them for different markets.[7] Recent findings about the capabilities of TikTok code and data-gathering capabilities have been concerning. In August 2022, privacy researcher Felix Krause found that TikTok's browser contains code that can track users' keystrokes, including if they type in login information, passwords and credit card information. This is not a practice among major social media competitors. TikTok responded by claiming it uses this code for debugging and troubleshooting.[8]

We should take TikTok's claims with a grain of salt. Previously, in 2020, TikTok executives said they would end a similar feature that allowed TikTok to read users' Apple iOS clipboards, but never gave a clear date for the removal of the feature. Apple's clipboard allows users to save snippets of information on their phones which, for some users, could include sensitive military and government data, and could stay in TikTok's servers even if an iPhone user deletes it after a moment. Despite these promises to end the feature, an Apple software update later revealed TikTok was still snooping on the clipboard. It remains unclear if TikTok still has kept the feature, which it has not publicly clarified.[9]

TikTok and China's Human Rights Atrocities

---

[5] David Bandurski, "Tech Shame in the "New Era,"" China Media Project, April 11, 2018, https://chinamediaproject.org/2018/04/11/tech-shame-in-the-new-era/.

[6] Fergus Ryan, Audrey Fritz, and Daria Impiombato, "TikTok and WeChat," Austrian Strategic Policy Institute, September 8, 2020, https://www.aspi.org.au/report/tiktok-wechat.

[7] Pellaeon Lin, "TikTok vs Douyin: A Security and Privacy Analysis," Citizen Lab, March 22, 2021, https://citizenlab.ca/2021/03/tiktok-vs-douyin-security-privacy-analysis/.

[8] Paul Mozur, Ryan Mac, and Chang Che, "TikTok Browser Can Track Users' Keystrokes, According to New Research," *The New York Times*, August 19, 2022, https://www.nytimes.com/2022/08/19/technology/tiktok-browser-tracking.html.

[9] Joel Thayer, "On TikTok, It's All Fun and Games Until China Wants Your Info," *The Verge*, July 21, 2022, https://www.theverge.com/2020/6/26/21304228/tiktok-security-ios-clipboard-access-ios14-beta-feature.

When TikTok began seeing explosive growth in America, I was deeply worried as a foreign correspondent and investigative journalist in Xinjiang, China, where I was researching my second book, *The Perfect Police State: An Undercover Odyssey Into China's Terrifying Surveillance Dystopia of the Future*. This is a region where an estimated 1.8 million people from the ethnic Uyghur, Kazakh and other predominately Muslim minority groups have been held in a network of some 300 concentration camps—the largest internment of ethnic minorities since the Holocaust. The people of Xinjiang live under a total surveillance dystopia seemingly crafted out of a science fiction novel, erected with the help of Chinese and American technology companies. They are watched by China's surveillance network, SkyNet, which is powered by novel technologies in artificial intelligence, facial recognition, voice recognition, and biometric data collection. In December 2017, I made my final visit to Xinjiang. Within three days, I was detained by police and asked to leave.

I believed that ByteDance's and TikTok's expansion into the U.S. was ominous for our democracy, and I began following the story carefully, interviewing TikTok employees, users, and former Chinese government officials about their operations. A Uyghur technology worker from the regional capital, Urumqi, who helped establish the government's surveillance systems in Xinjiang, told me, "Of course ByteDance can spy for the CCP, and they do it all the time. Every Chinese app submitted the government's orders to send them all the data of sensitive users like Uyghurs and different ethnic groups. Why would TikTok be any different? It doesn't matter if those companies are operating in America or not."

His concerns were appropriate. A former employee claimed that ByteDance had an active role in trying to suppress news about the Uyghur genocide, attempting to build an algorithm that would suppress Uyghur livestreams that could potentially spread news of atrocities on the Chinese app.[10] In November 2020, TikTok public policy executive Elizabeth Kanter, testifying before the British parliament, said, "There was [*sic*] some incidents where content was not allowed on the platform, specifically with regard to the Uyghur situation."

The Uyghur genocide—declared a "genocide" by the State Department in January 2021 because of the erasure of an entire group, including through the forced sterilization of women—is the culmination of China's fascistic propaganda about the racial and cultural superiority of the dominant Han Chinese ethnic group. TikTok policies, implemented until 2019, have reflected these censorial party practices that uphold the myths about strength, power and purity.[11] Internal memos leaked to *The Intercept*, an investigative news website, instructed TikTok moderators globally to suppress video posts created by users whom they deemed too poor, ugly, or disabled, as well as to censor users who harmed "national honor."

---

[10] Isobel Asher Hamilton, "ByteDance Tried to Build an Algorithm to Censor Uighur Livestreams on TikTok's Chinese Sister App, a Former Employee Has Claimed," *Insider*, February 19, 2021, https://www.businessinsider.com/bytedance-uighur-livestreams-douyin-censorship-2021-2.

[11] TikTok says it has changed many moderation and content policies since 2019. Its internal public relations guidance tells employees to say: "We're a platform that's nearly 3 years old and we're operating in the scale of other big players. We take this responsibility seriously. In the early days, we made mistakes with our moderation policies and we take responsibility for them." Chris Stokel-Walker, "Inside TikTok's Attempts to 'Downplay the China Association.'" *Gizmodo*, July 27, 2022, https://gizmodo.com/tiktok-master-messaging-pr-playbook-china-music-1849334736.

Other guidelines penalized users for posting about the 1989 Tiananmen Square massacre and the Uyghur genocide.[12] It called these posts "violations," even if the users who posted them were not based in China. The memo instructed moderators to be on the lookout for videos with an "abnormal body shape," "ugly facial looks," "dwarfism," an "obvious beer belly," "too many wrinkles," "eye disorders," "dilapidated housing," "slums, rural fields" and many other "low quality" traits.[13]

As these revelations came to light, TikTok scrambled to repair its image for the U.S. market. It claimed it was implementing stronger privacy and content moderation policies and made the odd claim that these policies were in place to prevent online bullying, even though the leaked internal documents made no mention of anti-bullying.

TikTok also said data was stored in America and on a backup server in Singapore, not in Beijing, where the parent company is based. In August 2020, the CFIUS issued a divestment order to ByteDance, ordering it to sell TikTok to an American company. The order went unenforced and was later reversed in June 2021. TikTok continues to operate freely in America, under China's control.

The Oracle Failure

After these controversies, TikTok announced in September 2020 that it had selected American technology giant Oracle as a "technology partner," restructuring its operations with Oracle bidding to purchase part of TikTok's U.S. operations. Oracle didn't purchase TikTok in the end (no one did). Instead, TikTok struck an agreement with Oracle to migrate Americans' data to Oracle servers in the U.S. It was trying to convince the U.S. government that the personal data of Americans would not end up in the hands of China's government.

This plan has already failed on many counts. In June 2022, the news website *Buzzfeed* published material from leaked audio files from 80 internal TikTok meetings. The leaks revealed that Chinese engineers had already been accessing the data of Americans from September 2021 to January 2022, which could then be easily stored on Chinese servers, even by accident. The leaks contradicted the sworn Congressional testimony of a TikTok executive in October 2021, who claimed inaccurately that a "world-renowned, U.S.-based security team" decides who will have access to Americans' data. TikTok employees said on the recordings that they had to work through China-based teams to figure out the flows of American data.[14]

Second, TikTok announced it would maintain backup storage of Americans' data on its own servers. This would erase the benefits of storing the data on Oracle cloud servers. Third, Oracle, despite being an American company, is a dubious data protection partner for TikTok; there is strong reason to doubt the private data of Americans will be completely safe with Oracle as well. Mara

---

[12] Alex Hern, "Revealed: How TikTok Censors Videos that Do Not Please Beijing," *The Guardian*, September 25, 2019, https://www.theguardian.com/technology/2019/sep/25/revealed-how-tiktok-censors-videos-that-do-not-please-beijing.

[13] Sam Biddle, Paulo Victor Ribeiro, Tatiana Dias, "Invisible Censorship," *The Intercept*, March 16, 2020, https://theintercept.com/2020/03/16/tiktok-app-moderators-users-discrimination/.

[14] Emily Baker-White, "Leaked Audio from 80 Internal TikTok Meetings Shows that U.S. User Data Has Been Repeatedly Access from China," *Buzzfeed*, June 17, 2022, https://www.buzzfeednews.com/article/emilybakerwhite/tiktok-tapes-us-user-data-china-bytedance-access.

Hvistendahl, a longtime China journalist at *The Intercept*, has documented Oracle's egregious conflicts of interest selling data analytics software to Chinese police authorities for mass surveillance.[15]

These conflicts of interest and split loyalties between China's hostile authoritarianism and America's homeland security run deep. Oracle has inappropriately advertised its software services for the U.S. Department of Defense to potential Chinese police and security clients. Oracle has offered China's Ministry of Public Security, the powerful, rights-abusing policing body, the data analytics software that undergirds China's *1984*-style surveillance dystopia and crimes against humanity. This includes marketing software directly to Chinese police authorities in Xinjiang, where they are carrying out genocide against the minority Uyghur population.[16]

TikTok's Shadowy Corporate Structure

Even if TikTok stores the data on Oracle's servers in America, Oracle's and TikTok's deep exposure to China makes that data susceptible to the vague, powerful data collection laws that give the Chinese government sweeping powers. If China demands this data—which would happen in secret, if it hasn't happened already—both TikTok in America and its parent company ByteDance in China will have few ways of resisting through legitimate court hearings and court appeals in China. The sad reality is that ByteDance's and TikTok's corporate structure makes them accountable to the authoritarian demands of the Communist Party.

TikTok has claimed that its operations fall outside Chinese legal jurisdiction, so we do not need to worry about the privacy of Americans' data. This trite and deceptive answer does not address the inherent contradiction in ByteDance's corporate structure that makes it prone to CCP data meddling and legal orders.

In November 2021, ByteDance's co-founder and new CEO, Liang Rubo, announced that TikTok would be separated into a standalone business unit, allegedly separate from the six main business units of TikTok.[17] The goal was to appeal to American government regulators who were concerned about the lack of separation between the American TikTok app and the other Chinese business affiliates under ByteDance. The restructuring, however, was in name only. It does not represent a spin-off of TikTok.

What we know as "TikTok," with its main American office in Los Angeles, is really part of a shell company incorporated in the Cayman Islands. According to the Cayman's corporate registry, the director in charge of the ByteDance shell company is Liang Rubo, who is also listed on ByteDance's website as the CEO of the ByteDance corporation in China. Because both the Cayman and Chinese companies have the same person in charge, it is difficult to take TikTok executives seriously when they argue that these are in fact separate companies divided by an impenetrable wall.[18] The Cayman

---

[15] Mara Hvistendahl, "How Oracle Sells Repression in China," *The Intercept*, February 18, 2021, https://theintercept.com/2021/02/18/oracle-china-police-surveillance/.

[16] Mara Hvistendahl, "How Oracle Sells Repression in China," *The Intercept*, February 18, 2021, https://theintercept.com/2021/02/18/oracle-china-police-surveillance/.

[17] Coco Feng, "ByteDance Carves Out TikTok as World's Most Valuable Technology Unicorn Finds Way to Satisfy U.S.-China Regulatory Demands," *South China Morning Post*, November 2, 2021, https://www.scmp.com/tech/article/3154537/bytedance-carve-out-tiktok-worlds-sole-hectocorn-splits-six-units-delineating.

[18] Brooks Barnes and Jack Nicas, "Disney's Head of Streaming Is New TikTok CEO," *The New York Times*, May 18, 2020, https://www.nytimes.com/2020/05/18/business/media/tiktok-ceo-kevin-mayer.html.

Islands are a notorious offshore tax and regulatory haven with little transparency, where Chinese kleptocrats evade American regulatory pressure.

The fuzzy corporate structure has troubling implications for Americans' private data. TikTok's privacy policy states: "We may share all of the information we collect with a parent, subsidiary, or other affiliate of our corporate group." TikTok does not clarify the definition of "our corporate group." Worded this way, TikTok executives have given themselves enormous latitude to share data with whomever they want within their parent ByteDance company, whether in China or the Cayman Islands shell company, despite promising to keep that data out of China's hands.

TikTok executives might decide to share data with ByteDance's key subsidiary in China, called Beijing ByteDance Technology. Here's the danger: the Chinese government owns a 1 percent stake in Beijing ByteDance Technology and has installed its own director on the subsidiary's board.[19] Yet under the privacy policy, TikTok might be contractually clear if American users brought a legal claim against the company for allowing their private data to end up in the hands of Chinese authorities, through Beijing ByteDance Technology.

The Vast Intrusions of Chinese Data Law

TikTok's claims that its America-based data is not subject to Chinese law reveals an egregious misrepresentation of the Chinese legal system. Increasingly, China is asserting global legal jurisdiction and is using this self-proclaimed authority to pressure American and other foreign companies with ties in China. China does not operate under the principle of rule of law, but rule by the Party. The Party has the sweeping authority to enforce a collection of vague laws that criminalize the refusal to hand over the data of anyone, often anywhere in the world, it deems a threat.

One regulation, put in force in January 2021, allows China's Commerce Ministry to tell international companies to choose between complying with the extraterritorial regulations of China or the U.S., including the various sanctions or export controls now in force under U.S. law. Chinese courts can then hold companies liable for complying with American restrictions on Chinese commerce unless the Commerce Ministry grants them a waiver.[20] If ByteDance were to treat TikTok as a separate business unit and submit to American government orders to, say, divest and sell TikTok, ByteDance might find itself in legal trouble in China, and pressured to hand over Americans' data in Chinese court.

Another venue for harassment is the Data Security Law, passed in June 2021, giving China's government vast powers over the regulation and collection of "core data," a vague term that applies to any data that includes "national security, lifelines of the national economy, important aspects of people's lives, and the major public interest."[21] If any American TikTok data ends up on China-based servers, as the leaked audio files obtained by *BuzzFeed* show can easily happen, the CCP would have no trouble asserting the legal authority to obtain that data on "national security" grounds.

---

[19] Yingzhi Yang and Brenda Goh, "Beijing took stake and board seat in key ByteDance domestic entity this year," August 17, 2021, https://www.reuters.com/world/china/beijing-owns-stakes-bytedance-weibo-domestic-entities-records-show-2021-08-17/

[20] Amy Qin, "China's New Rules Could Hit U.S. Firms and Send a Message to Biden," Amy Qin, *The New York Times*, January 9, 2021, https://www.nytimes.com/2021/01/09/business/china-rules-trump-biden-sanctions.html.

[21] "Data Security Law of the People's Republic of China," June 10, 2021, http://www.npc.gov.cn/englishnpc/c23934/202112/1abd8829788946ecab270e469b13c39c.shtml

These laws only scratch the surface of China's broad and recent judicial expansion over the data of private citizens anywhere in the world. In June 2020, China passed the Hong Kong National Security Law, asserting extraterritorial jurisdiction over non-citizens of the People's Republic of China and Hong Kong, even if they live beyond China's borders, for collusion with "foreign forces."[22]

As of July 2022, China has charged at least 119 people under the law, which is usually targeted at political dissidents.[23] But the law's wording gives vast powers for the Chinese government to charge a person, anywhere in the world, including a TikTok or ByteDance executive traveling through Hong Kong, should that executive cooperate, for instance, with U.S. government requests to protect the data of American military personnel. TikTok closed its Hong Kong office in July 2020 and stopped offering the app in Hong Kong. This, however, does nothing to shield TikTok and its users from the ubiquitous and global territorial powers of the Hong Kong National Security Law.

Finally, two other sweeping laws, the 2015 National Security Law and the 2017 National Intelligence Law, assert similar government powers over private data in China and would apply to ByteDance and potentially its Caymans subsidiary TikTok.

The 2015 National Security Law states vaguely: "Citizens of the People's Republic of China, all state organs and armed forces, political parties and mass organization, enterprises, public institutions and other social organizations, all have the responsibility and obligations to preserve national security." This wording will compel ByteDance to fulfill any data obligation imposed by the Chinese government under the guise of "national security."[24]

The 2017 National Intelligence Law creates the obligation of "Chinese citizens to support national intelligence work," or face detention and possible criminal charges.[25] This law, of course, would apply to ByteDance and its executives in China, should Chinese intelligence agencies want to pressure them to hand over data on American government and military users gathered through TikTok.

Senators, I hope my testimony today will inform decisions you might be called upon to make about the TikTok threat. I hope that my summary of TikTok's connections to the CCP, data-gathering practices, broken promises, and pattern of deception has made a case to open a CFIUS review, once again. This review would potentially force ByteDance to sell TikTok to a more trustworthy company. It would be our best option moving forward. Thank you for having me.

---

[22] Human Rights Watch, "China: New Hong Kong Law a Roadmap for Repression," July 29, 2020, https://www.hrw.org/news/2020/07/29/china-new-hong-kong-law-roadmap-repression.

[23] Selina Cheng and Elliot Bentley, "How China's National Security Law Silences Hong Kong," July 1, 2022, https://www.wsj.com/articles/how-chinas-national-security-law-silences-hong-kong-11656673119.

[24] "National Security Law," July 1, 2015, https://www.chinalawtranslate.com/en/2015nsl/.

[25] Bonnie Girard, "The Real Danger of China's National Intelligence Law," February 23, 2019, https://thediplomat.com/2019/02/the-real-danger-of-chinas-national-intelligence-law/.

HEARING BEFORE
THE UNITED STATES SENATE COMMITTEE ON
HOMELAND SECURITY AND GOVERNMENTAL AFFAIRS

September 14, 2022

Testimony of Chris Cox
Chief Product Officer, Meta

I. Introduction

Chairman Peters, Ranking Member Portman, and distinguished members of the Committee, thank you for the opportunity to appear before you today. My name is Chris Cox. I am Meta's Chief Product Officer, and I oversee our apps and privacy teams.

I first joined Meta (then Facebook) in 2005 as one of the company's first 15 software engineers. I care deeply about the work we do to help people connect with the things and people they care the most about. It is important to us that we help people feel safe on our family of apps, and we stand firmly against the exploitation of social media by those committed to inciting violence and hate.

That's why, under the Facebook Community Standards, we prohibit hate speech, terrorism, and other harmful content. We employ tens of thousands of people and use industry-leading technology, including advanced artificial intelligence to enforce these rules. We regularly publish transparency reports so people can see how we're doing over time and how we compare to other internet platforms. We're proud of our work in this space—but we're always working to improve and to share our lessons learned with other companies in the US and around the world.

II. Combating Hate and Dangerous Organizations

Our Violence and Incitement policy prohibits content calling for or advocating violence, and we ban individuals and organizations that proclaim a violent mission under our Dangerous Individuals and Organizations policy. In August 2020, we expanded this policy further to address militarized social movements, such as certain militias, and violence-inducing conspiracy networks, like QAnon. We remove content that represents, praises, or supports those groups.

We work tirelessly to enforce these policies. We've designated more than 1,000 militarized social movements and 270 white supremacist organizations, and we removed 2.3 million pieces of content from Facebook tied to organized hate globally in the second quarter of 2022, nearly 97% of which we found before someone reported it.

We have a team of cross-functional experts focused on these issues at Meta, including more than 350 highly trained professionals who work exclusively or primarily to prevent terrorist and violent extremist content from appearing on our platform and to quickly identify and remove it if it does. These professionals possess expertise ranging from law enforcement and national security experience to counterterrorism intelligence and academic studies in radicalization.

We invested around $5 billion on safety and security last year alone—more than any other tech company, even adjusted for scale. We are proud that we have over 40,000 people working on safety and security issues, and our efforts are making a difference. For example, we've more than halved the amount of hate speech people see on Facebook over the last eighteen months. Hate speech now represents only about 0.02% of content views, or around 2 views per every 10,000. Our advanced artificial intelligence systems have also improved at keeping people safe on our platform by proactively removing content that violates our standards. We found more than 95% of the hate speech we removed before anyone reported it to us—up from just 23% a few years ago.

When it comes to advertisements that run on Facebook and Instagram, they must comply with the same Community Standards and Community Guidelines that apply to other content people post, and we place additional requirements on advertisers to help further protect people from poor experiences. For example, our Ads Policies prohibit predatory advertising practices and using ad targeting to discriminate against, harass, provoke, or disparage users. We have an ad review system that is designed to review all ads before they go live. If a violation is found at any point in the review process, the ad will be rejected.

Our efforts to combat extremism and hate don't end with our policies and enforcement efforts. We also have a number of programs to direct people to content or organizations to help them disengage from dangerous or violent movements. When people search for terms related to QAnon on Facebook and Instagram, we redirect them to credible resources from the Global Network on Extremism and Technology (GNET), the academic research network of the Global Internet Forum to Counter Terrorism (GIFCT).

We are proud to have co-launched GIFCT five years ago to help fight terrorism and extremism online alongside our peers at other technology companies, including Google, YouTube, Microsoft, and Twitter. GIFCT was the culmination of years of informal partnerships among our companies on these issues, and the group has created a shared industry database for violent terrorist imagery that we have removed from our services. Sharing information allows all of us to identify more quickly and more accurately potential extremist content on our respective platforms. Most importantly, it also allows smaller companies the ability to take advantage of our technology and tactics, even with fewer people and resources. We believe that terrorism and extremism are shared problems that require shared solutions, and we encourage all tech companies to continue to partner with us in our efforts to keep such violence and hate off of online communities.

III. Using Algorithms to Keep People Safe and Improve their Experience

Like most platforms, Facebook and Instagram use many different algorithms for various app features, including to enable our search function and to help enforce our policies. For example, we use algorithms to identify and help remove content that violates our policies, including hate speech, incitement, and terrorism. This often happens before anyone reports content to us, sometimes even at the point of creation.

In addition to safety and enforcement, we also have algorithms to help rank the content people follow and create their personalized Feed. Oftentimes when people refer to Facebook's "algorithm," they are talking about this sort of content ranking algorithm. This personalization is important because people often have thousands of posts in their Facebook Feed each day from their friends, Pages they choose to follow, and Groups they choose to join. Most people don't have time to look at all of this content every day, so the ranking process helps sort it and put the things they will find most valuable toward the top of their Feed.

I want to stress that the goal of ranking is to help people see what they find most valuable. It is not to keep people on the service for a particular length of time, and it's certainly not to give people the most provocative or enraging content. In fact, key parts of those systems are designed to do just the opposite. We reduce the distribution of many types of content—meaning that content appears lower in Feed—because they are sensational, misleading, gratuitously solicit engagement or are found to be false by our independent fact-checking partners.

In 2018, we publicly announced that we were making ranking changes that we thought would help people see more content that was meaningful to them—but that would nonetheless lead to people spending less total time on Facebook. The prediction proved correct; the change led to a decrease of 50 million hours' worth of time spent on Facebook per day, and we saw a loss of billions of dollars in the company's market cap. We view this as a success because it improved the experience of our users, and we think building good experiences is good for the business in the long term.

Transparency and choice are also important to giving people the best possible experience. We have made significant progress over the past several years in providing greater transparency into how the Feed ranking process works, what gets distributed and why. Notably, the "Why Am I Seeing This" feature lets people understand why a particular post shows up where it does—and to change their settings easily if something's not right. And we have published Content Distribution Guidelines, which explain why content might receive reduced distribution. Recently, we also made it easier for people to switch between, for example, seeing the newest content first or seeing only content from people they've chosen to be a "favorite."

IV. Incorporating Safety into Our Products by Design

Safety and integrity are key to the product experience, and we build and update our products with these values in mind. We embed teams focusing specifically on safety and integrity directly into product development teams across the company, allowing us to address potential issues during product development. And individual product teams use our integrity tools to build in preventative safeguards at the start.

Before launch, potential new products or product changes are reviewed against a set of integrity standards to help us provide a positive experience for users. The process also allows us to identify and anticipate potential abuses and build in mitigations from the start. This process helps us build-in effective privacy, security, and safety protections before a product launches.

After products launch, we continue to monitor their impact, including by looking at integrity

metrics, to ensure products are best serving our community. Some of these same integrity metrics are released quarterly to the public, as part of our Community Standards Enforcement Report, which provides data on how Meta enforces its policies.

Abuse of our products isn't static—and neither is the way we approach our integrity work. We're continuing to evolve how we approach integrity, embedding integrity teams with product teams across Facebook, Instagram, Messenger, WhatsApp, and the teams that are going to build the metaverse in the years to come. And we are committed to continuing to improve so that we can help keep people safe.

V. Conclusion

At the end of the day, our job is to build the best product for people, and that's a product that is reliable, fast, secure, relevant, and safe—a product that connects people to content relevant to their interests and connects them to their family and their friends. That's the product people want and the product we wake up every day trying to build.

We appreciate your attention to these important issues and look forward to continuing to work with your offices to find ways we can continue to improve our products, processes, and partnerships.

Thank you, and I look forward to your questions.

Written Testimony of Neal Mohan
Chief Product Officer, YouTube and SVP, Google

U.S. Senate Committee on Homeland Security and Governmental Affairs
"Social Media's Impact on Homeland Security: Part II"
September 14, 2022

Chairman Peters, Ranking Member Portman, and distinguished members of the Committee; thank you for the opportunity to appear before you today. My name is Neal Mohan, and I am the Chief Product Officer for YouTube. In my role, I am responsible for YouTube's products, user experience, and trust and safety globally. I appreciate the opportunity to provide testimony to the Senate Homeland Security and Governmental Affairs Committee.

At YouTube, our mission is to give everyone a voice and show them the world. Every day, we build and improve tools and systems that empower creators, viewers, and businesses to find and share information like never before. Billions of people come to YouTube to enjoy videos and music, upload original content, and learn new skills. Our work creates new economic opportunities for artists, creators, journalists, and small businesses to share their creativity and products in the United States and across the globe.

Openness has always been one of our guiding principles, but our commitment to openness works hand in hand with our responsibility to protect our viewers, creators, and partners from harmful content. Responsibility is our top priority at YouTube and informs every product and policy decision we make.

My testimony today will focus on (1) our approach to Responsibility at YouTube; (2) our policies against harmful content; and (3) our collaboration across industry to combat terrorist content online.

Youtube's Four Rs of Responsibility

At YouTube, we have four pillars of Responsibility, described in detail below, which we refer to as "the Four Rs." Used together, these pillars have enabled us to make substantial progress in tackling harmful content, as well as to respond quickly in the face of unprecedented events such as the COVID-19 pandemic and the war in Ukraine. Under the Four R approach, we

Remove content that violates our policies as quickly as possible; we *Raise up* authoritative sources when people are looking for news and information; we *Reduce* the spread of misinformation and content that brushes up against our policy lines; and we *Reward* trusted, eligible creators and artists.

Remove Violative Content: YouTube is an open platform but that has never meant anything goes. Our Community Guidelines provide public-facing rules of the road for content on YouTube and we remove content that violates these rules as quickly as possible. We work continuously to improve our efforts through enhanced detection and enforcement, relying on a combination of technology and people. Our automated detection systems are an area of significant investment for YouTube, and our engineering teams continuously evaluate their efficacy and make improvements. These systems help our human review teams remove content at scale, with the speed and volume that could not be achieved with people alone.

We are proud of the advancements we have made in our enforcement efforts as a result of our investment. In the first half of 2022, we removed close to 8.4 million videos for violating our Community Guidelines, more than 92% of which were first flagged by our automated systems. Approximately 67% of those first flagged by our systems received 10 or fewer views before we removed them. During these two quarters, we removed more than 130,000 videos specifically for promoting violence and violent extremism in violation of our policies. As noted above, our approach enables us to adapt quickly in the face of unprecedented events, including those that may lead to harmful disinformation on our platform. For example, since the war in Ukraine began, YouTube has removed more than 76,000 videos and 9,000 channels related to the ongoing war in Ukraine for violating our Community Guidelines and Terms of Service.

In addition to removing violative content, we track the percentage of views on YouTube that comes from this content. We refer to this metric as the Violative View Rate, or VVR. In the first half of 2022, the VVR was .09% - .11%. In other words, for every 10,000 views on YouTube, only 9-11 went to content that proved to be violative.

In order to hold ourselves accountable to our users and the public at large, we publish a quarterly Community Guidelines Enforcement Report that provides data on VVR and our enforcement across channels, videos, comments, user flags, and appeals and reinstatements.

Raise Up Authoritative Voices: YouTube is a source for news and information for people around the world—whether about events unfolding in local communities or more existential global issues like the COVID-19 pandemic. Not all queries are the same, however. For topics like music or entertainment, relevance, newness, and popularity are most helpful to understand what people are interested in. But for subjects such as news, science, and historical events, where accuracy and authoritativeness are key, the quality of information and context are paramount. Our search and recommendations systems are designed to raise up

authoritative voices in response to user queries that are "news-y" or related to topics prone to misinformation.

We also recognize that there are topics that may require additional context, which we provide through our information panels. These panels appear in the search results and video watch pages, with a link to an authoritative source. Topics include, for example, the holocaust, moon landing, and election-related information such as how to register to vote.

Reduce the Spread of Borderline Content: Equally important to raising up authoritative information is reducing the spread of borderline content, or content that does not quite cross the line of our policies for removal but that we don't necessarily want to recommend to people. We use machine learning to reduce the recommendations of this type of content, including potentially harmful misinformation, and as a result of our efforts it represents just a fraction of what is watched on YouTube in the United States. We are able to raise up authoritative information and reduce borderline content by using classifiers to identify whether a video is "authoritative" or "borderline." These classifications rely on human evaluators who assess the quality of information in each channel or video. These evaluators hail from around the world and are trained through a set of detailed, publicly available rating guidelines. We also rely on certified experts, such as medical doctors when content involves health information. Taken together, all of our responsibility work around recommendations has shown real impact. Watchtime of authoritative news is up dramatically and borderline viewing is down.

We use a number of signals to recommend videos to users, including a user's language, time of day, and user satisfaction. But responsibility outweighs these considerations: if a video is identified as borderline content, it will be demoted in recommendations. As a result of our efforts, we saw a 70% drop in watchtime on this type of content in the US in 2019, and continue to keep views below 1% of total views of content on YT.

Reward Trusted Creators: Finally, we are proud to be a place where creative entrepreneurs can build thriving businesses, and we reward trusted creators. However only creators that meet a high bar can make money on YouTube.

More than 2 million creators around the world are part of our YouTube Partner Program (YPP). Through YPP, we share the majority of our advertising revenue with our partners, and offer a variety of ways to make money, including ads and channel memberships. Over the past three years we've paid out more than $30 billion to creators, artists and media companies (as of November 2020).

In order to join YPP and earn money on YouTube, creators must meet specific eligibility requirements and comply with our monetization policies. Individual videos must follow our advertiser-friendly content guidelines to earn money.

Channels that repeatedly brush up against our monetization policies or our Community Guidelines—including hate speech, harassment, and misinformation—will be suspended from the YouTube Partner program, meaning they can't run ads on their channel or use other monetization features like Super Chat. To protect our ecosystem of creators, advertisers, and viewers, we also tightened our advertising criteria in 2017. After careful analysis and conversations with creators, we changed certain eligibility requirements for monetization, which significantly improved our ability to identify creators who contribute positively to the community, while also preventing potentially inappropriate videos from monetizing content.

As a testament to our responsibility, we were the first digital platform to receive accreditation for content level brand safety from the Media Rating Council (MRC). This means that MRC found us to be 99% effective at ensuring ads only show up where they should. We also provide metrics in all four categories of the Global Alliance for Responsible Media's cross-industry Aggregated Measurement Report, which evaluates platform safety for brands and consumers.

YouTube's Robust Policies To Address Violent Extremist Content

As noted above, our Community Guidelines set forth what content is not allowed on YouTube. With respect to violent extremist content, we have a network of robust policies in place that work together to combat it.

First, our Community Guidelines prohibit terrorist organizations from using YouTube for any purpose, including recruitment, as well as content that promotes terrorism, glorifies terrorist acts, or incites violence. To complement the policies aimed at terrorist organizations, the Guidelines also include policies that address violent extremism, violent or graphic content, harmful or dangerous content, election misinformation, and hate speech. Together, these policies work to prohibit key categories of harmful content regardless of who posts it. More details about these policies is below.

- **Harmful or Dangerous Content:** YouTube prohibits content that encourages dangerous or illegal activities that risk serious physical harm or death. This policy applies to various types of conduct, including content that provides instructions to build a bomb meant to injure or kill others, and content that promotes or glorifies violent tragedies, such as school shootings.
- **Firearms:** YouTube prohibits content intended to sell firearms, instruct viewers on how to make firearms, ammunition, and certain accessories, or instruct viewers on how to install those accessories. We also do not allow live streams that show someone holding, handling, or transporting a firearm, unless they are uniformed police or military.

- **Harassment & Cyberbullying:** YouTube prohibits content that threatens individuals or targets an individual with prolonged or malicious insults based on intrinsic attributes. This policy includes, for example, direct or implied threats, or sharing someone's personally identifiable information online, known as doxxing.

- **Hate Speech:** YouTube removes content promoting violence or hatred against individuals or groups based on age, caste, disability, ethnicity, gender identity and expression, nationality, race, immigration status, religion, gender, sexual orientation, veteran status, or victims of a major violent event and their kin. This policy also prohibits videos alleging that a group is superior in order to justify discrimination, segregation or exclusion based on qualities like age, gender, race, caste, religion, sexual orientation or veteran status. We also prohibit content denying that well-documented violent events, like the Holocaust or the shooting at Sandy Hook Elementary, took place.

- **Violent Criminal Organizations:** YouTube prohibits content intended to praise, promote, or aid violent criminal organizations. Examples of content that violates this policy would be videos or comments directing users to sites hosting manifestos from the perpetrators of well-documented violent events or content that is aimed at recruiting new members to violent criminal or terrorist organizations designated by the U.S. government.

- **Violent or Graphic Content:** YouTube prohibits violent or gory content intended to shock or disgust viewers, or content encouraging others to commit violent acts. This includes, but is not limited to, content that incites others to commit violent acts against individuals or a defined group of people; fights involving minors; and content where animals are encouraged or coerced to fight by humans. It also applies to footage, audio, or imagery involving road accidents, natural disasters, war aftermath, terrorist attack aftermath, street fights, physical attacks, sexual assaults, immolation, torture, corpses, protests or riots, robberies, medical procedures, or other such scenarios with the intent to shock or disgust viewers. Further, footage or imagery showing bodily fluids such as blood or vomit with the intent to shock or disgust viewers is prohibited by this policy.

Content that violates our policies against violent extremism, detailed above, includes material produced by organizations designated by the U.S. government as 'foreign terrorist organizations'. We do not permit these terrorist organizations to use YouTube for any purpose. Content produced by violent extremist groups that are not government-listed foreign terrorist organizations is subject to our policies, including those described above, regardless of who posts the video, or the group to which they belong.

We regularly review and update our policies to address new and emerging threats. For example, in June 2019, we strengthened our hate speech policy to specifically prohibit videos alleging that a group is superior to justify discrimination, segregation or exclusion based on

qualities like age, gender, race, caste, religion, sexual orientation or veteran status. In December 2019, we updated our harassment policy to take a stronger stance against threats and personal attacks, and introduced new penalties for patterns of harassing behavior. In October 2020, we expanded our hate and harassment policies to prohibit content that targets an individual or group with conspiracy theories that have been used to justify real-world violence, like QAnon.

In addition to our efforts to remove violative content produced by individuals and organizations, Google deploys efforts to counter government-based influence operations. We take the security of our users very seriously, and we have dedicated teams in place to protect against attacks from a wide range of sources. YouTube works closely with Google's Threat Analysis Group (TAG), which is dedicated to protecting users from threats posed by state-sponsored malware attacks and other advanced persistent threats. On any given day, TAG is tracking more than 270 attacker groups from more than 50 countries. When we detect attempts to conduct coordinated influence operations on our platforms, whether state-backed or otherwise, we swiftly remove offending content from our platforms and terminate these actors' accounts. We take steps to prevent possible future attempts by the same actors, and share intelligence to prevent attacks elsewhere.

This ongoing work includes the removal of Chinese threat actors seeking to conduct operations on our platforms and target our users. In the first 6 months of 2022, we terminated more than 24,000 of these as a result of investigations into coordinated influence operations linked to China. Data about YouTube channel terminations made as a result of these investigations is included in TAG's quarterly Bulletin.

Due to the evolving nature and shifting tactics of groups promoting violative content, we continuously review and adapt our policies to stay ahead of bad actors. Combating these threats is an area where we constantly look to build stronger defenses, including technological mechanisms to defend our platform at scale.

YouTube's Collaborative Work to Curtail Terrorist or Extremist Content

In light of the increasingly interconnected nature of the information ecosystem, we collaborate across the industry to combat terrorist and violent extremist content on the Internet.

As noted above, in 2016, we collaborated with industry partners to develop a hash-sharing database where we share "digital fingerprints" of terrorist content to stop its spread. The shared database includes more than 320,000 unique hashes of both video and still image

media. both videos and images. Since 2017, the number of companies contributing to and benefiting from this database has grown from 4 to 13.

We also share best practices on counterterrorism with our industry peers through the Global Internet Forum to Counter Terrorism (GIFCT), which is dedicated to disrupting terrorist abuse of digital platforms. With GIFCT partners, we developed a content incident protocol to enable a swift and coordinated response to perpetrator-created content across platforms. GIFCT also collaborates with the Tech Against Terrorism initiative to hold workshops with more than 100 smaller tech companies around the world.

* * *

Thank you, Mr. Chairman, for convening this important hearing. Responsibility is and will continue to be YouTube's number one priority—our business literally depends on it. We look forward to continuing to work with you to address these challenges. Thank you.

Testimony Before the U.S. Senate Committee on Homeland Security and Governmental Affairs

Written Statement of Testimony

Social Media's Impact on Homeland Security: Part II

Testimony of Vanessa Pappas

Vanessa Pappas, Chief Operating Officer, TikTok Inc.

September 14, 2022

Chairman Peters, Ranking Member Portman, and Members of the Committee:

Thank you for the opportunity to appear before you today to discuss how internet companies are working to prevent online extremism to ensure our platforms do not pose a threat to homeland security.

My name is Vanessa Pappas, and I am the Chief Operating Officer at TikTok, where I oversee TikTok's global operations, including content, marketing, Trust and Safety, and user operations. I live with my family in Los Angeles, where I work in one of more than a dozen offices we have across the U.S.

I've been in the United States for more than twenty years, and have spent my career working in entertainment and media. Prior to joining TikTok, I worked at YouTube where I served as the Global Head of Creative Insights, overseeing the company's strategic growth initiatives to drive its daily active user base across key markets and user segments. Prior to that, I was the Global Head of Audience Development at YouTube, focusing on developing audience growth strategies for creators, media publishers, labels and artists.

TikTok's mission is to inspire creativity and bring joy. Over the past couple of years, we have seen tremendous growth and now more than 1 billion people come to TikTok for an authentic, entertaining experience. Every day, individuals come to TikTok to be entertained, to express themselves, and to learn, and small businesses come to reach new customers and build their brands. I came to TikTok because I believe in this mission and see examples every day about how it is lifting people up and helping drive mutual understanding from people in different parts of the world and from different walks of life. In an increasingly complex world, I see TikTok as a contributor to bringing us together and helping us understand each other better.

We are keenly aware that with this success and growth come greater accountability and responsibility. TikTok is committed to being a trusted industry leader in safety and transparency, and we appreciate the Committee's interest in these efforts.

Safety

TikTok strives to create a safe environment where creative, joyful content can flourish, and we have made decisions that prioritize safety over short-term commercial success. For example, we do not allow political ads on the platform, even though they could be a source of significant revenue. Similarly, we do not accept advertisements for categories of content that may hurt our efforts to support the safety of our community, such as advertisements associated with violence or threats, including guns, knives, explosives, pepper spray, and other weapons; ammunition; and tactical gear such as police or military uniforms, armored vests, handcuffs, or batons.

At TikTok, the focus on safety starts at the top, and as an executive, there's no responsibility more important to me than protecting the people on our platform. One of our leadership team's top goals is to "strengthen safety and build trust." TikTok has Terms of Service and Community Guidelines to help ensure that content on TikTok is safe. Safety reviews are a standard part of our product launch process. We also apply our Community Guidelines to everyone and to everything on TikTok. They define a set of norms and common code of conduct, and they provide guidance on what is and is not allowed on TikTok in order to create a welcoming and safe space for entertainment. Specifically, these Community Guidelines include policies that prohibit harmful misinformation, coordinated inauthentic behavior, and promotion of violence.

We educate our community about these guidelines in a variety of ways in the app. For example, we produce in app videos called @tiktoktips, that help people understand how to produce safe content and avoid content that might get their videos removed from the platform. People are notified when they have violated the guidelines and of the consequence(s) of their violation(s). TikTok also educates creators through a dedicated portal that includes detailed information about our policies.

Strong policies are insufficient if they are not enforced through constant attention by human moderators using modern tools. Content moderation policy and implementation for the United States is led by our U.S. Safety Team in Los Angeles, which reports to me. Moderators will remove content—including video, audio, livestream, images, and comments—that violates our Community Guidelines. We use a mixture of automated and human review for content moderation, with automation helping to scale moderation and human reviewers focused on making decisions that are more nuanced relative to our guidelines. Trust and safety represents our largest labor expense for TikTok's U.S. operations. There are thousands of people working across safety, privacy, and security on a daily basis. TikTok invests heavily in these teams, as well as in technology to detect potential violations and suspicious accounts at scale.

We understand that verifying certain information during dynamic and fast moving events can be challenging and so, to minimize risk, we work with independent fact checkers to evaluate content for false or misleading claims. We currently have 13 fact-checking partners, including Agence France-Presse (AFP), Australian Associated Press (AAP), Animal Político, Deutsche Presse-Agentur (Dpa), Facta, Estadão Verifica, Lead Stories, Logically, Newtral, Newschecker, PolitiFact, Science Feedback, and Teyit. These fact-checking partners support 33 languages and assess content in 64 markets around the world. All of our fact-checking partners are accredited

by the International Fact-Checking Network as verified signatories of the International Fact-Checking Network's code of principles.

TikTok also encourages people who identify violations of our Community Guidelines to report such violations through the app. People can report violative videos by pressing on the video, selecting "Report", and following the instructions provided when prompted. Systems are also in place to allow trusted flaggers to escalate content for human review.

Violent Extremism and Hateful Behavior

In addition to removing content that violates our policies, we also employ proactive measures, informed by the analysis U.S. intelligence agencies and trusted organizations in civil society, to prevent the spread of hateful ideologies and violent extremist groups on our platform. For example, we have a zero tolerance stance on content that promotes white supremacy or nationalism and remove accounts with repeated content violations. In addition, we remove race-based harassment and the denial of well-documented and violent events. We work with a range of experts, including the Anti-Defamation League (ADL), to continually expand our database of hateful terms and symbols, and incorporate them into our models, moderation systems, and training materials.

We take into account publicly available information from experts, including the United Nations Security Council and Southern Poverty Law Center, to designate dangerous or hateful individuals and organizations. In appropriate cases, we work to ban all content associated with violent extremist groups and individuals from appearing on the platform. Examples include foreign terrorist organizations, drug cartels, and groups such as QAnon, Three Percenters, and Oath Keepers. Users who search for this content, including related hashtags or keywords, are redirected to our Community Guidelines.

An example of our approach to violent extremism can be seen in content related to the attack on the Capitol on January 6, 2021. TikTok is an entertainment-first platform, which is why it was not the platform of choice for those who organized the violence at the Capitol. From our review of 850 sets of Department of Justice charging documents related to the January 6 attack, there were 686 references to social media companies. TikTok was mentioned in only 18 of those 686 cases.

Transparency

In addition to keeping our platform safe, we are committed to being transparent about how we accomplish that goal. Every quarter, we release a Community Guidelines Enforcement Report. These reports contain detailed information about the type and volume of content we remove. For instance, in the first quarter of 2022, our proactive removal rate for content violating our violent extremism policy was 91.4 percent. That means that more than 9 times out of 10, we discovered and removed the content on our own before receiving any reports from users or third parties.

Additionally, 83.9 percent of the removals under our violent extremism policy happened before the content received any views, and in 88.4 percent of the cases, the removals occurred within 24

hours of the content being posted. Our proactive detection rates have improved each year, and we are dedicated to continuing this trend.

Twice a year, we also disclose data about requests we receive from law enforcement or governments. TikTok is committed to responding to law enforcement requests for user information in a manner that respects the privacy and other rights of our community members. Each request we receive is carefully reviewed. In our semiannual reports, we include a detailed breakdown of these requests, including the countries from which they originate.

We are also building Transparency and Accountability Centers in Los Angeles and Washington, DC, that will allow us to share information about how we moderate and recommend content. We have brought parts of the tour into an online experience during the pandemic and would be happy to arrange a virtual tour for Members and Committee staff at your convenience, or invite you to join us in the physical centers when construction is complete. We have also confirmed that Oracle will be vetting validating our content recommendation and moderation models.

Algorithm

TikTok's For You feed is part of what enables connection and discovery. It is central to the TikTok experience and where most people spend their time. When you are watching the For You feed, you are presented with a stream of videos curated to your interests, making it easy to find content and creators you love. This feed is powered by a recommendation system that delivers content to each user that is likely to be of interest to that particular user. Part of the magic of TikTok is that there's no single For You feed—while different people may come upon some of the same standout videos, each person's feed is unique and tailored to that specific individual.

Our uniquely powerful yet easy-to-use tools democratize video creation, enabling everyday people to express themselves creatively and find their community on the platform. This approach has resulted in more authentic content and has helped launch new cultural trends from feta pasta to the resurgence of Fleetwood Mac's "Dreams." It has allowed small businesses to find their voice, and to expand their reach and customer base, and it has been a bright spot for American families during the Covid-19 pandemic. In an October 2021 study conducted by Nielsen, TikTok was the only app where a top reason for usage was "to lift my spirits" and found that people around the globe find TikTok content to be authentic, genuine, unfiltered, and trendsetting.

For people who don't select categories, we start by offering them a generalized feed of popular videos. Their first set of likes, comments, and replays will initiate an early round of recommendations as the system begins to learn more about their content tastes. When people decide to follow new accounts, for example, that action will help refine their recommendations, as will exploring hashtags, sounds, effects, and trending topics on the Discover tab. All of these are ways to tailor the experience and invite new categories of content into the feed.

TikTok's recommendation system ranks videos for people based on a combination of factors, including their own interactions with other videos and device and account settings. All these factors are processed by our recommendation system and weighted based on their value. A strong indicator of interest, such as whether a person finishes watching a longer video from

beginning to end, would receive greater weight than a weak indicator, such as whether the video's viewer and creator are both in the same country. Videos are then ranked to determine the likelihood of a user's interest in a piece of content and are delivered to each unique For You feed.

TikTok is home to creators with many different interests and perspectives, and sometimes users may come across a video that isn't quite to their taste. Users can long-press to add a video to their favorites, and can long-press on a video and tap "Not Interested" to indicate that they don't care for a particular video. People can also choose to hide videos from a given creator or made with a certain sound, or report a video that seems out of line with our guidelines. All these actions contribute to future recommendations in the For You feed.

To keep the For You feed interesting and varied, our recommendation system works to offer diverse and new areas of discovery and delight. For example, the For You feed generally will not show two videos in a row made with the same sound or by the same creator. We also do not recommend duplicated content, content a user has already seen before, or content that is considered spam. However, a person might be recommended a video that has been well received by other people who share similar interests.

Diversity is essential to maintaining a thriving global community, and it brings the many corners of TikTok closer together. What this means is that sometimes people may come across a video in their feed that does not appear to be relevant to their expressed interests or have amassed a huge number of likes. This is an important and intentional component of our approach to recommendation: bringing a diversity of videos into the For You feed gives people additional opportunities to stumble upon new content categories, discover new creators, and experience new perspectives and ideas.

By offering different videos from time to time, the system is also able to get a better sense of what is popular among a wider range of audiences to help provide other people using TikTok a great experience, too. The goal is to find balance between suggesting content that is relevant while also helping them find content and creators that encourage them to explore experiences they might not otherwise see.

The recommendation system is also designed with safety as a consideration. We recently announced the development of a new content classification system that will help ensure that people are getting relevant, interesting, and appropriate content. We are focusing initially on safeguarding the teen experience with this new system, and in the coming months will add content filtering options for our entire community so they can enjoy more of what they love.

Consistent with our safety-first approach, content that is being fact checked and reviewed content that cannot be substantiated will be ineligible for recommendation into the "For You" feed (FYF). Similarly, videos that have just been uploaded, are under review, or are flagged as spam-like content seeking to artificially increase traffic may also be ineligible for recommendation into a user's FYF.

API for Researchers

To provide greater transparency, we are in the process of developing a research API that would allow selected researchers to gain access to certain data about content and activity on our platform, and plan to make it available later this year.

In addition, we have developed a moderation system API that we plan to make available at our Transparency and Accountability Centers. This moderation system API will provide selected researchers an effective way to evaluate our content moderation systems and examine existing content available on our platform. In our Transparency and Accountability Centers, selected researchers will also be able to upload their own content to see how different types of content are either permitted, rejected, or passed to moderators for further evaluation.

The independent experts on our U.S. Content Advisory Council and regional Safety Advisory Councils will also be granted API access as well as access to confidential information, such as our keyword lists (which are used to help detect and flag potentially violative content) for deeper analysis. We do not make keyword lists available publicly in order to avoid providing a roadmap for bad actors who attempt to subvert our safeguards. While we have dedicated teams regularly stress-testing our processes and tools to ensure they're robust and effective, we know that perspectives and insights from experts can strengthen our approach.

Data Security

Since 2020, we have been clear and transparent about our broader objective to limit the number of employees who have access to user data and the circumstances under which data access is enabled. As we noted in an April 2020 blog post, "[o]ur goal is to minimize data access across regions so that, for example, employees in the APAC region, including China, would have very minimal access to user data from the EU and U.S." Separately, in a September 2020 sworn declaration, TikTok acknowledged that ByteDance personnel in China can have access to TikTok user data based on demonstrated need to perform their roles.

Employees outside the U.S., including China-based employees, can have access to TikTok U.S. user data subject to a series of robust cybersecurity controls and authorization approval protocols overseenby our U.S.-based security team. In addition, TikTok has an internal data classification system and approval process in place that assigns levels of access based on the data's classification and requires approvals for access to U.S. user data. The level of approval required is based on the sensitivity of the data according to the classification system. We monitor and review access to U.S. user data on an ongoing basis.

While we have been transparent about our data access policies and protocols, we also have been working assiduously to address national security concerns identified by U.S. policymakers and regulators, including the Committee on Foreign Investment in the United States (CFIUS). Although I will not be commenting on the CFIUS process during my remarks today because of the confidentiality of the process, I can tell you that we've made very significant progress in that process, some of which has been disclosed in the media. Specifically, as was reported in the

press earlier this year, for more than a year we have been pursuing a multi-pronged initiative called "Project Texas" to strengthen TikTok's U.S. data security program.

We recently reached a significant milestone by changing the default storage location of U.S. user data to the Oracle Cloud Infrastructure. TikTok now stores 100% of U.S. user data by default in the Oracle cloud environment, and we are working with Oracle on new, advanced data security controls that we hope to finalize in the near future. We still use our U.S. and Singapore data centers for backup, but as we continue our work, we expect to delete U.S. users' private data from our own data centers. This work puts us closer to the day when we will be able to pivot toward our industry-leading system for protecting the data of our users in the United States, with robust, independent oversight to ensure compliance.

Additionally, we are making operational changes in line with these protocols. In May 2022, TikTok announced the creation of a new division—U.S. Data Security ("USDS")—to bring heightened focus and governance to our ongoing efforts to strengthen our data protection policies and protocols, further protect our U.S. users, and build confidence in our systems and controls in the United States. This division has U.S.-based leadership. Access to U.S. user data by anyone outside of our new U.S. Data Security team will be limited by, and subject to, robust data access protocols that are being developed in close collaboration with Oracle and the U.S. government.

In order to facilitate a global platform, our goal is to ensure non U.S.-based employees, including China-based employees, will only have access to a narrow set of TikTok U.S. user data, such as public videos and comments available to anyone on the TikTok platform, to ensure global interoperability.

Privacy Policy

TikTok's privacy policy describes the information we collect and how we use that information. In some cases, our privacy policy describes information that we do not currently collect, but may collect in the future. We appreciate the Committee's interest in further clarity about what certain terms in our privacy policy mean and whether we are currently collecting certain data elements that are referenced in our privacy policy.

Like other apps, we collect information about what people search for within our app. We do this to recommend more relevant content on TikTok. We do not collect people's Google searches or queries on other search engines. Similarly, if users browse content within our app, we collect that information to serve more relevant videos in their For You feed. In our Privacy and Security Center, we explain that this includes "[b]rowsing history in the TikTok in-app browser to help make platform improvements, such as optimizing page load times and ad measurement." Like other platforms, we may also collect a limited amount of data related to user activity on advertiser's apps and websites when those advertisers elect to share such data with us.

TikTok does not currently collect precise geolocation in the United States. TikTok collects coarse (approximate) location information based on things like the subscriber identity module (SIM) card and IP address associated with the user's device. Such location information is much

less precise than GPS coordinates and is used for things like recommending locally relevant content and ads.

We do not use facial, voice, or other physical features to identify U.S. users. We offer a variety of special effects that people can use when creating videos. When a user creates a video, they can choose to apply an effect to their face (e.g., add sunglasses) or to change the tone of their voice (e.g., like a chipmunk). To enable such effects, we collect and process information about the images and audio that are part of the user's video, such as the location of facial features within an image (e.g., detecting the location of eyes for a sunglasses effect) and aspects of the audio (e.g., to raise the pitch to sound like a chipmunk).
We use keystroke patterns and rhythms as one of the signals to identify, detect, and thwart inauthentic or spammy behavior. There have been media reports suggesting that when someone accesses a website by using TikTok's in-app browser, TikTok captures everything that the user is typing (including credit card information). Such reports are inaccurate. We do not collect precise keystroke or text input through the code at issue, but instead solely use this information for debugging, troubleshooting, and performance monitoring.

As discussed earlier in my testimony, we recently announced the creation of a new division—USDS—to help strengthen and improve our systems and controls in the United States. Access to U.S. user data by anyone outside of USDS will be limited by, and subject to, robust data access protocols that will be developed in close collaboration with Oracle and the U.S. government.

These operating protocols will necessarily and significantly constrain our ability to share U.S. user data within our broader group of corporate entities. TikTok, which is not available in China, is subject to the laws of the jurisdictions where it operates. As we have stated before, we have not been asked for US user data by the Chinese government. We have not provided such data to the Chinese government, nor would we if asked.

Relationship with ByteDance

TikTok is a flagship product of ByteDance Ltd., a global technology company operating a range of content platforms that inform, educate, entertain, and inspire people across languages, cultures, and geographies.

As a global entertainment platform, TikTok spans most major markets except China, where ByteDance offers a different short-form video app called Douyin. TikTok is provided in the United States by TikTok Inc., which is incorporated in California and subject to U.S. laws and regulations. Like many global technology companies, we have product development and engineering teams all over the world collaborating to deliver the best product experience for our community. TikTok is led by an executive team in the United States and Singapore and has global offices including Los Angeles, Silicon Valley, New York, Dublin, London, Paris, Berlin, Dubai, Singapore, Jakarta, Seoul, and Tokyo.

ByteDance Ltd. is the ultimate parent entity that is incorporated outside of China. The board is comprised of our CEO Rubo Liang, Bill Ford of General Atlantic, Arthur Dantchick of Susquehanna International Group, Philippe Laffont of Coatue, and Neil Shen of Sequoia.

ByteDance's investors include global institutional funds such as Blackrock, Coatue, Fidelity, General Atlantic, KKR, Sequoia, Softbank, and Susquehanna International Group. Today, over sixty percent of ByteDance Ltd. is owned by Western investment firms, with most of the rest of the company owned by the founders and employees.

Conclusion

We appreciate the challenges that we face as an industry, and TikTok will remain steadfast and vigilant in promoting safety, transparency, and trust on our platform. Thank you for the opportunity to discuss these important issues before the Committee.

EXHIBIT A TO SUBMISSION

TikTok

June 30, 2022

The Honorable Marsha Blackburn
United States Senate
357 Dirksen Senate Office Building
Washington, DC 20510

The Honorable Roger Wicker
United States Senate
555 Dirksen Senate Office Building
Washington, DC 20510

The Honorable John Thune
United States Senate
511 Dirksen Senate Office Building
Washington, DC 20510

The Honorable Roy Blunt
United States Senate
260 Russell Senate Office Building
Washington, DC 20510

The Honorable Ted Cruz
United States Senate
127A Russell Senate Office Building
Washington, DC 20510

The Honorable Jerry Moran
United States Senate
521 Dirksen Senate Office Building
Washington, DC 20510

The Honorable Shelley Moore Capito
United States Senate
172 Russell Senate Office Building
Washington, DC 20510

The Honorable Cynthia Lummis
United States Senate
124 Russell Senate Office Building
Washington, DC 20510

The Honorable Steve Daines
United States Senate
320 Hart Senate Office Building
Washington, DC 20510

Dear Senators Blackburn, Wicker, Thune, Blunt, Cruz, Moran, Capito, Lummis, and Daines,

Thank you for your letter dated June 27, 2022. We appreciate the opportunity to address the concerns you set forth. Many of your questions appear to stem from a recent BuzzFeed article, which contains allegations and insinuations that are incorrect and are not supported by facts. We appreciate the opportunity to set the record straight by answering your questions.

Before doing so, we would like to contextualize what many of the people quoted in the article were talking about and what the company has been broadly working to achieve. For well over a year, we've been pursuing a multi-pronged initiative called "Project Texas" to strengthen the company's data security program. Security experts can confirm that these initiatives are often painstaking and complex, even with expert assistance from world-class companies like Oracle and Booz Allen. Some people working on these projects do not have visibility into the full picture, working on a task

TikTok

without realizing that it's a single step in a much bigger project or a test to validate an assumption.

That's critical context for the recordings leaked to Buzzfeed, and one thing their reporting got right: the meetings "*were in service of Project Texas's aim to halt this data access.*"

The broad goal for Project Texas is to help build trust with users and key stakeholders by improving our systems and controls, but it is also to make substantive progress toward compliance with a final agreement with the U.S. Government that will fully safeguard user data and U.S. national security interests. We have not spoken publicly about these plans out of respect for the confidentiality of the engagement with the U.S. Government, but circumstances now require that we share some of that information publicly to clear up the errors and misconceptions in the article and some ongoing concerns related to other aspects of our business.

While we are disappointed that leaks have put us in this position, we are pleased to share the substantial progress on our objectives. As we recently reported, we now store 100% of U.S. user data by default in the Oracle cloud environment, and we are working with Oracle on new, advanced data security controls that we hope to finalize in the near future. That work puts us closer to the day when we will be able to pivot toward a novel and industry-leading system for protecting the data of our users in the United States, with robust, independent oversight to ensure compliance.

We are taking additional measures beyond data security, which we will briefly touch on in our responses below.

We have been clear dating back to an early 2020 blog post that we are working on a broad set of objectives: "*Similar to industry peers, we will continue to drive our goal of limiting the number of employees who have access to user data and the scenarios where data access is enabled. Although we already have controls in place to protect user data, we will continue to focus on adding new technologies and programs focused on global data residency, data movement, and data storage access protections worldwide.*" (https://newsroom.tiktok.com/en-us/our-approach-to-security). There is a distinction between data storage and data access, but they are both—together—important components of our efforts to earn trust and improve security for TikTok; our solution will now ensure both the storage of all U.S. user data in the United States and all data sharing outside of the protected enclave in the United States will be pursuant to protocols and terms approved by the U.S. Government.

In light of the context above, we are confident that when you review our responses, you will see that TikTok has not, at any point, misled Congress about our data and security controls and practices. We understand, respect, and appreciate the incredibly important work of your Committee and Congress, and we have always approached our engagements with Members and staff, both in public and in private, with candor and

TikTok

integrity. We stand by the statements Michael Beckerman made before Congress and are grateful for his leadership.

As we continue our productive conversations with the Administration and continue to explore commercial partnerships with companies like Oracle, we look forward to keeping you and the full Committee apprised of our work to further ensure the security of U.S. user data.

Please see below for TikTok's responses to your questions.

1. **Is it true that TikTok employees located in China currently have, or had in the past, access to U.S. user data? This could include programmers, product developers, data teams, as well as trust and safety and content moderation professionals.**
    a. **If yes, please explain in detail which employees have or had such access and for what purposes.**
    b. **If the employees had this access in the past but no longer do, please identify the applicable date ranges.**

Employees outside the U.S., including China-based employees, can have access to TikTok U.S. user data subject to a series of robust cybersecurity controls and authorization approval protocols overseen by our U.S.-based security team. In addition, TikTok has an internal data classification system and approval process in place that assigns levels of access based on the data's classification and requires approvals for access to U.S. user data. The level of approval required is based on the sensitivity of the data according to the classification system.

The solution that TikTok is implementing pursuant to Project Texas has focused on evaluating and revising TikTok's internal policies and operational controls in relation to U.S. user data access, to take steps to strengthen data security around U.S. user data and, ultimately, to make the organizational, process, and technical changes to help ensure compliance and enhance protection of U.S. user data defined as "protected" through engagement with CFIUS. As we are in the process of undergoing CFIUS national security review, we have kept CFIUS informed of these efforts. This protected user data will be stored in Oracle Cloud Infrastructure with access limited only to authorized personnel, pursuant to protocols being developed with the U.S. Government.

2. **TikTok's privacy policy says you share data you collect with your parent companies and affiliates and that you transmit user information to servers and data centers overseas.**
    a. **Have any ByteDance employees—located in China or elsewhere—had access to U.S. User data, either currently or in the past?**

Please see our response to question 1.

TikTok

b. What are the locations of the servers and data centers overseas where TikTok transmits U.S. user data?

TikTok has long stored U.S. user data in data centers in the U.S. and Singapore, as well as in cloud-based services offered by AWS, the Google Cloud Platform, and Azure. Our Virginia data center includes physical and logical safety controls such as gated entry points, firewalls, and intrusion detection technologies. It is also important to maintain backup data storage locations to guard against catastrophic scenarios where user data could be lost, and our data center in Singapore serves as the backup data storage location for our U.S. user data.

100% of U.S. user traffic is now being routed to Oracle Cloud Infrastructure. We are still using our U.S. and Singapore data centers for backup, but as we continue our work to deliver on U.S. data governance, we expect to delete U.S. users' protected data from our own systems and fully pivot to Oracle cloud servers located in the U.S.

3. Do any ByteDance employees have a role in shaping TikTok's algorithm?

Subject to the controls described in our response to question 1, ByteDance engineers around the world may assist in developing those algorithms, however our solution with Oracle will ensure that training of the TikTok algorithm only occurs in the Oracle Cloud Infrastructure and will also ensure appropriate third-party security vetting and validation of the algorithm. For more information about how TikTok's algorithm recommends content, please see our Newsroom post: https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you.

4. Do any Douyin employees have any access to American user data or a role in shaping TikTok's algorithm?

ByteDance developed the algorithms for both Douyin and TikTok, and therefore some of the same underlying basic technology building blocks are utilized by both products, but TikTok's business logic, algorithm, integration, and deployment of systems is specific to the TikTok application and separate from Douyin.

Under Project Texas and as a result of our work with the U.S. Government, going forward our solution with Oracle will ensure the TikTok application and platform, including the algorithm, is deployed through the Oracle Cloud Infrastructure in the United States with third-party security vetting and validation of the software for the application and platform, including the TikTok algorithm.

TikTok

5. **In the past, TikTok has said that it has never—nor would it ever—provide user data to the Chinese government, even if asked. Yet your privacy policy says you can disclose data collected to respond to government inquiries.**
    a. **Has TikTok ever disclosed any U.S. user data to respond to government inquiries from the Chinese Communist Party?**
    b. **If the Chinese Communist Party asked you for U.S. user data, what is to stop you from providing it? Can the CCP compel you to provide this data, regardless of response? Can they access it, regardless of response?**
    c. **Has ByteDance ever responded to CCP inquiries on TikTok's behalf?**
    d. **Has TikTok ever shared U.S. user data with ByteDance for the purpose of responding to a CCP inquiry?**

We have not been asked for such data from the CCP. We have not provided U.S. user data to the CCP, nor would we if asked.

More information about government requests for user data that we receive across the world is available in our Information Request Reports, available at https://www.tiktok.com/transparency/en-us/information-requests-2021-1/.

6. **Do TikTok employees in the U.S. use software developed by ByteDance, such as Lark?**

Yes.

7. **Does ByteDance have any role—either in the past or in the present—in hiring TikTok employees in the U.S.?**

As would be expected of any global company with subsidiaries, ByteDance plays a role in the hiring of key personnel at TikTok. However, as we have described before, TikTok is led by its own global CEO, Shou Zi Chew, a Singaporean based in Singapore.

8. **Does TikTok own or lease its own office space in the U.S., and does ByteDance have any ownership or lease stake in those facilities?**

TikTok leases office space in cities across the U.S., including Los Angeles, Austin, Chicago, New York, Detroit, Seattle, DC, and Nashville. These leases are through U.S. entities, TikTok Inc. (a California corporation) and ByteDance Inc. (a Delaware corporation).

9. **Does the Chinese government have an ownership stake or seat on the Board of Directors, or provide personnel in any other leadership position, of the Beijing ByteDance Technology Company?**

TikTok

a. **What role does this seat play in impacting decisions made at ByteDance or TikTok?**
b. **Does this position afford an opportunity for the board member to determine whether and how TikTok or ByteDance will respond to CCP inquiries?**
c. **Does this position afford an opportunity for the board member to view TikTok user data?**
d. **Would you be informed, as a matter of policy, if a board member did view the data? If the board member did share the data, in any capacity, with the CCP?**

As multiple corporate entities share the "ByteDance" name, several China-based ByteDance entities were renamed earlier this year to keep the names of businesses and entities more consistent. Beijing Bytedance Technology Co. Ltd is now called Beijing Douyin Information Service Limited. We will refer to it here using its new name for avoidance of confusion.

ByteDance Ltd., the ultimate parent entity that is incorporated in the Cayman Islands, has a global board, including Bill Ford of General Atlantic, Arthur Dantchik of Susquehanna International Group, Philippe Laffont of Coatue, Neil Shen of Sequoia, and the company's CEO Rubo Liang. The majority of ByteDance's investors are global institutional funds such as Coatue, General Atlantic, KKR, Sequoia, Softbank, and Susquehanna International Group.

Beijing Douyin Information Service Limited is a separately held subsidiary of ByteDance Ltd. Beijing Douyin Information Service Limited does not have any direct or indirect ownership interest in or control over any TikTok entity. Further, employees of Beijing Douyin Information Service Limited are restricted from U.S. user database access. The Chinese state-owned enterprise's acquisition of 1% of Beijing Douyin Information Service Limited was necessary for the purpose of obtaining a news license in China for several China-based content applications, such as Douyin and Toutiao.

The Chinese government does not directly or indirectly have the right to appoint board members or otherwise have specific rights with respect to any ByteDance entity within the chain of ownership or control over the TikTok entity.

10. **How will TikTok's new cloud service arrangement be structured, and how will the company determine which data is "protected" such that it is not shared with employees or others in China?**

TikTok recently published a Newsroom post outlining our U.S. data governance practices and announcing a commercial relationship with Oracle in support of these practices (https://newsroom.tiktok.com/en-us/delivering-on-our-us-data-governance).

TikTok

As described in question 1, U.S. user data at issue is being defined as "protected" through engagement with CFIUS, and will be stored in the Oracle Cloud Infrastructure with access limited only to certain personnel in USDS. Under the contemplated arrangement, access to U.S. user data by anyone outside of USDS will be limited by, and subject to, robust data access protocols, with further monitoring and oversight mechanisms by Oracle to validate compliance.

In order to facilitate a global platform, non U.S.-based employees, including China-based employees, will have access to a narrow set of non-sensitive TikTok U.S. user data, such as public videos and comments available to anyone anywhere in the world, to ensure global interoperability so our U.S. users, creators, brands, and merchants are afforded the same rich and safe TikTok experience as global users.

11. Why is TikTok not planning to ensure that all U.S. user data is blocked from view of employees or others in China?

As described in our response to question 10, access to U.S. user data by anyone outside of our new USDS team will be limited by, and subject to, robust data access protocol that will be developed in close collaboration with Oracle and CFIUS.

We're proud to be able to serve a global community of more than a billion people who use TikTok to creatively express themselves and be entertained, and we're dedicated to giving them a platform that builds opportunity and fosters connections worldwide. We also work hard to safeguard our community, both in how we address potentially harmful content and how we protect against unauthorized access to user data.

Consistent with the operation of this global platform, and as described in our response to question 10, certain China-based employees will have access to a narrow, non-sensitive set of TikTok U.S. user data, such as the public videos and comments available to anyone, to ensure global interoperability so our U.S. users, creators, brands, and merchants are afforded the same rich and safe TikTok experience as global users. But this access will be very limited, it will not include private TikTok U.S. user information, and it will only occur pursuant to protocols being developed with the U.S. Government.

*****

TikTok

We thank you for your questions and appreciate the opportunity to provide additional details and clarification. We know we are among the most scrutinized platforms from a security standpoint, and we aim to remove any doubt about the security of U.S. user data. We're dedicated to earning and maintaining the trust of our community and of policymakers, and will continue to work every day to protect our platform and provide a safe, welcoming, and enjoyable experience for our community.

Sincerely,

Shou Zi Chew
CEO, TikTok

cc:
The Honorable Maria Cantwell
The Honorable Richard Blumenthal

United States Senate Committee on Homeland Security & Governmental Affairs
"Social Media's Impact on Homeland Security: Part II" Hearing

Testimony of Jay Sullivan
General Manager of Bluebird (Consumer Products) &
Interim General Manager of Goldbird (Revenue Products)
Twitter, Inc.

September 14, 2022

Chairman Peters, Ranking Member Portman, and Members of the Committee:

Thank you for the opportunity to speak to you today on Social Media's Impact on Homeland Security. My name is Jay Sullivan. I joined Twitter last November as a Vice President on the Consumer Product team. Five months ago, I was promoted to the General Manager of Bluebird, Twitter's Consumer Product team that is responsible for the main features people use on Twitter's mobile app and website. I am also currently the interim General Manager of Goldbird, Twitter's Revenue Product team.

Twitter's purpose as a company is to serve the public conversation. The open nature of our service gives a voice to a world of diverse people, perspectives, ideas, and information. We foster free and global conversations that allow all people to consume, create, distribute, and discover information about the topics and events they care about most. At Twitter, we operate with the belief that together, we are and will continue to be a force for good in the world.

For example, in the past year we have seen people come to Twitter to get on-the-ground information about the conflict in Ukraine, including ways people can help those in need. Countless individuals have used our service to access potentially life-saving information during natural disasters, and to exchange ideas about diverse topics ranging from news, to culture, to sports.

It is our fundamental belief that public conversation should be healthy and safe. Providing our valuable service comes with challenges. As with any tool, some people and organizations will try to abuse it for their own gain or to harm others. We take our responsibility to address these issues seriously. My role at Twitter is to lead its product vision, strategy, and execution. Today, I want to focus my testimony on how keeping our service healthy and safe is an integral part of our product strategy and is essential for how we grow Twitter.

I look forward to sharing with the Committee some of the important work we are doing through product design and interventions, policies, and external engagements to make sure that Twitter is enjoyed by everyone in safe and healthy ways, and in ways that further the values of freedom of speech and expression.

Incentivizing Health & Safety

I want to make clear at the outset that Twitter as a company is incentivized to keep our platform healthy and safe. Indeed, my top two objectives as the General Manager of Bluebird are to develop products to grow the number of users on Twitter and to prioritize health and safety. These two priorities go hand-in-hand because if people aren't protected from hate, abuse, and harassment, they will leave the service. Toxic behavior therefore impacts not only the health and safety of Twitter, but also harms long-term user growth. We therefore build health and safety into the design of new features, but if we are not satisfied, we will pause, delay, or stop a product rollout because of health and safety concerns.

A healthy and safe public conversation is also essential for our advertisers, who want to ensure that their brand, products, services, and activity are not depicted alongside harassment, vitriol, extremism, or false and misleading information. If they do not have these assurances, they will withdraw their ads from our service. Focusing on a healthy and safe public conversation also allows Twitter to further our core value of defending and respecting the rights of people using our service while promoting the core tenets of free expression.

In sum, mitigating risks and prioritizing a healthy and safe Twitter is good for our users, our business, and society. And, it allows us to better achieve our growth and financial goals..

Product Design To Promote Healthy and Safe Public Conversation

Since health is an integral part of our product strategy and development process, Twitter collectively works to prioritize health and safety every step of the way, from ideation to launch of a product's lifecycle. At the outset, during the ideation and design phase before we begin development,, we assess the impact on health and safety. This process includes a comprehensive assessment of potential risks and unintended consequences. It also includes developing mitigation strategies, which are integrated into product development and planning. To do this work effectively, the product team works collaboratively internally and includes individuals with a range of expertise, ranging from research to human rights.

Before any major product or policy launch, a cross-functional group of people will work together to consider potential risks, unintended consequences, responses from bad actors, and risk of abuse. This work includes an analysis of the potential risk posed by product features to Twitter users, the platform, and society. Our Trust & Safety team includes subject matter experts on different issues, as well as experts focused on considering the consequences of product and policy decisions and how best to remediate them. Furthermore, our research team uses internal and external studies to inform our work, and we also have a number of employees publishing academic papers on their work to share insights.

We also build in outside perspectives in a range of ways — as part of a formal public feedback process, direct engagement with experts, academics and civil society groups, or through our research work. These perspectives help us understand risks, mitigations, and trade-offs that inform our wider product and policy strategy. We act on these risks and build mitigations as we develop products.

We know we won't get everything right the first time and that people might react differently to how we expect when a new feature or policy is launched. So we use experiments to test new features, sometimes based on geography, other times with a random sample of people around the world. We pause, delay, or cancel initiatives if the risks can't be mitigated.

At Twitter, we also have intentionally prioritized openness and real, meaningful transparency. Transparency is central to how we build and ship products, as well as how we work to improve experiences on Twitter. When we develop a new feature, as much as possible, we do so in the open, incorporating feedback from the people who use Twitter, and ensuring we create a safe, accessible end-product for everyone.

Transparency also means accountability — owning our mistakes and correcting them. When we get something wrong, we communicate transparently about it and hold ourselves accountable for fixing it. Our commitment to transparency is likewise embodied in other key ways, like our open application programming interfaces — also known as APIs — including our free academic research track access, our ongoing industry-leading disclosures of state-linked information operations, and the information regularly shared in the Twitter Transparency Center.

In addition to building health into product design, we also use product interventions to promote health and safety. We have been proactively developing a new set of products and features that give users more control over their experience and help them feel safe and we are seeing promising data and outcomes. Here are a few that are emblematic of this focus:

- *Labels on Tweets:* This allows us to label Tweets that are misleading with clear warnings, accessible context, and de-amplify and limit engagements on certain Tweets through Likes, Retweets, and Replies.

- *Prompts before certain actions like Retweets and Replies are taken:* We've found that simple prompts that encourage people to read articles — past the headline alone — or consider a potentially abusive response before sharing have a demonstrated impact. These are speed bumps that essentially slow down content creation or viral sharing, with the intended effect of encouraging people on Twitter to *consider* what they're reading or saying before sharing it.

- *Conversation Controls:* We allow people on the service to control who can respond to their Tweets. We also allow them to hide unwanted replies to their Tweets or unmention themselves from a conversation to help people have more control over their experience.

- *Birdwatch*: Birdwatch allows people to identify information in Tweets they believe is misleading and write notes that provide informative context. We believe this approach has the potential to respond quickly when misleading information spreads, adding context that people trust and find valuable. Eventually, we aim to make notes visible directly on Tweets for the global Twitter audience, when there is consensus from a broad and diverse set of contributors.

- *Disabling Algorithmic Ranking*: The Sparkle button — which has been a feature of our service since 2018 — allows people on Twitter to view Tweets in reverse chronological order, rather than in an order suggested by our technology. This is a simple tool, and gives people control.

As we develop and expand our product roadmap moving forward we will continue to build on these and introduce new capabilities to keep our platform and customers safe. These capabilities include expanding our systems and processes to ensure that we are de-amplifying objectionable content on the platform, removing content that violates our policies, and making it easier for customers to report problematic content to us.

Policies Designed to Mitigate Harm and Promote Safety

While our Trust & Safety team is responsible for developing the policies and governance frameworks that prevent and mitigate harm to the people who use Twitter, I want to briefly touch upon our Twitter Rules — the policies we have in place to make sure people can participate in the public conversation freely and safely. These policies make clear that violence, hateful conduct, harassment, and other types of threatening behavior are not permitted on Twitter.

Our policies are built around the promotion and protection of fundamental human rights, including freedom of expression, safety, and privacy. These rights, among others, are enshrined in the Universal Declaration of Human Rights, which is an international document adopted by the U.N. and numerous countries around the world.

We believe deeply in and advocate for freedom of expression and open dialogue. We know that people do not always agree. The ability to dissent, to share information and opinions freely, even when unpopular, provocative, or questioned, is a value that makes up the foundation of free expression, but that means little as an underlying philosophy if voices are silenced because people are afraid to speak up due to threats to their physical safety or privacy. That is why we have policies that make clear that we will not allow for the promotion of violence, disinformation, or hateful conduct on Twitter as they undermine our ability to serve the public conversation, our customer experience, our business, and our ability to promote the open internet.

Platform Integrity & Authenticity Policies

Our platform integrity and authenticity policies promote the health of the public conversation by addressing, among other things, efforts to spread misinformation relating to civic integrity, moments of crisis, COVID, and synthetic and manipulated media. We are constantly reviewing and evaluating misinformation efforts and focusing on those that are most harmful. For example, we added our crisis misinformation policy in May 2022 as we determined that in times of crisis — such as situations of armed conflict, public health emergencies, and large-scale natural disasters — false and misleading information has a special capacity to bring harm to vulnerable populations and shape crisis dynamics.

Our coordinated harmful activity policy addresses those situations where we find groups, movements, or campaigns that are engaged in coordinated activity resulting in harm on and off of Twitter and take enforcement action on any accounts that we identify as associated with those entities. In order to take action under this policy, we evaluate these groups, movements, or campaigns against an analytical framework, with specific on-Twitter consequences if we determine that they are harmful. You can read more about this approach here.

Safety Policies

Our safety policies are built to prohibit abuse, harassment, violence, and criminal actions on Twitter. Among the policies included in this category are: non-consensual nudity, suicide and self-harm, perpetrators of violent attacks, private information, hateful conduct, sensitive media, abusive behavior, violent organizations, violent threats, glorification of violence, abusive profile information, child sexual exploitation, and illegal or certain regulated goods or services. You can read more about these policies here.

Our violent threats, wishes of harm, and glorification of violence policies prohibit content on Twitter that promotes, incites, or threatens violence off of the platform. All forms of incitement of violence — whether veiled, coded, or opaque — fall squarely under this prohibition.

Brand Safety Policy

Our Brand Safety policy, which is led by our Customers team, builds upon the foundation laid by the Twitter Rules to promote a safe advertising experience for all users and brands. In addition to our Brand Safety efforts, which help inform ad placement on Twitter, we also have Advertising Policies that determine permissible content in ads and conduct of advertisers on Twitter. You can learn more about our Ads Policies here.

External Engagements

We know at Twitter that our efforts to create healthy conversation require engagement with industry, academia, the public, governments, and civil society, among others, to be successful and to address the most serious online threats and develop products and policies that further our efforts. I want to highlight a few of our external engagements today.

Twitter Trust and Safety Council

The Twitter Trust and Safety Council is a group of independent expert organizations from around the world. Together, they advocate for safety and advise us as we develop our products, programs, and rules. At the end of 2019, we expanded the Council to include even more global experts and diverse perspectives. The Council is made up of several advisory groups, each dedicated to issues critical to the health of the public conversation.

This year we've engaged with the Trust and Safety Council on 6 projects, following the 13 projects we engaged with the Council on in 2021, early in the development process. We distilled and put to use their feedback on ways we can offer a better and safer experience for people using

Twitter. Their feedback directly informed our approach on several products. You can read more about the Twitter Trust and Safety Council here.

Global Internet Forum to Counter Terrorism ("GIFCT")

Twitter co-founded GIFCT with YouTube, Microsoft, and Facebook in 2017. GIFCT helps technology companies, government, civil society, and academia share information to counter terrorist and violent extremist activity online. GIFCT evolved with the Christchurch Call to Action, an initiative that governments, tech platforms, and civil society organizations committed to after the March 2019 mosque shootings in Christchurch, New Zealand and viral spread of the perpetrator's live-streamed video of the attack. Among GIFCT's work:

- A real-time content incident protocol (CIP) that allows us to respond to a violent act quickly to ensure that we share valuable information across industry to limit the spread of terrorist and violent extremist content.
- A shared, safe and secure industry database of "perceptual hashes" of known images and videos — produced by terrorist entities on the United Nations designated terrorist groups lists.
- Establishing the Global Network on Extremism and Technology (GNET), an independent, industry-funded initiative for better understanding, and counteracting, terrorist use of technology.

Since the attack in Christchurch, GIFCT members have shared alerts relating to hundreds of incidents around the world and activated its content incident protocol three times in response to violent attacks in Halle, Germany (2019), Glendale, Arizona (2020) and Buffalo, New York (2022). Supported by table-top exercises and post-incident reviews, we continue to strive to limit the spread of perpetrator-produced content, whether through video, audio or manifestos. You can read more about GIFCT here.

Digital Trust and Safety Partnership

The Digital Trust and Safety Partnership (DTSP) is an initiative focused on promoting a safer and more trustworthy internet. Twitter joined as an inaugural member of the group because we supported its efforts to develop, use and promote industry best practices, reviewed through internal and independent third-party assessments, to ensure consumer trust and safety when using digital services. We have committed to DTSP's five fundamental areas of best practices, including: product development, governance, enforcement, improvement, and transparency. You can read more about the commitments and self-assessments of the DTSP here.

Global Alliance for Responsible Media (GARM)

GARM is a cross-industry initiative established by the World Federation of Advertisers to address the challenge of harmful content on digital media platforms and its monetization via advertising. Through our engagement with the GARM, for example, we have been able to contribute to the creation of the industry-standard Brand Safety Floor and Suitability Framework.

These are but a few examples of our external engagements where we are learning, discussing, and taking action that not only promote the public conversation on Twitter but working across industry to create sharing-mechanisms and standards for keeping content safe for people across the digital ecosystem.

Conclusion

We have invested, and will continue to invest heavily, in building the technologies, policies, and procedures necessary to offer informative and safe experiences for the millions of people on Twitter.

Our task is not an easy one and what I described today in terms of our products, policies, and external engagements will change as new challenges, risks, and threats develop. In order to mitigate the harms this Committee is examining, we must constantly change and evolve. We are committed to doing what is necessary to continuously foster a healthy service while upholding the tenets of free expression, which is, in our view, the best way for Twitter to help protect democracy in the United States and abroad.

Thank you, and I look forward to answering your questions.

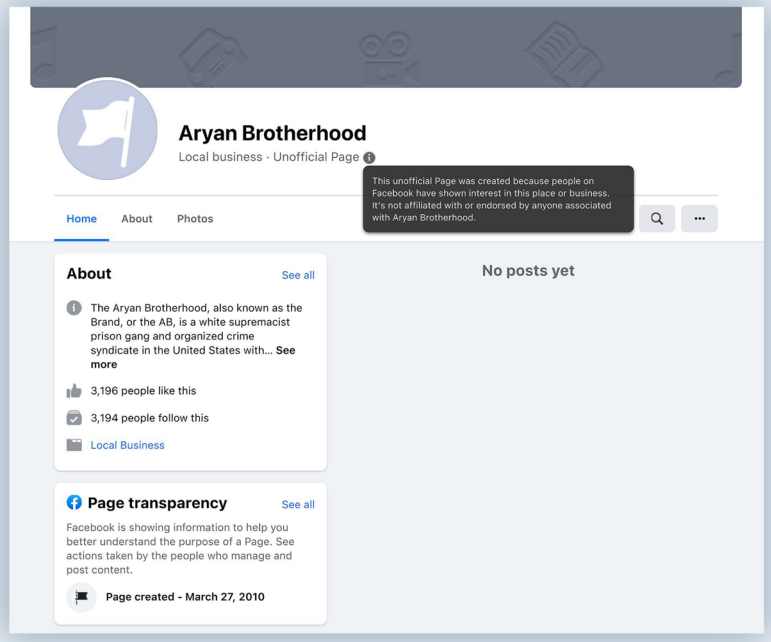
TECHNOLOGY

Racists and Taliban supporters have flocked to Twitter's new audio service after executives ignored warnings

"We did not prioritize identifying and mitigating against health and safety risks before launching Spaces... We have launched products and features without adequate exploration of potential health implications" – *internal Twitter report*

Source: The Verge

Facebook Auto Generates Pages for Extremist Groups



The page was on Facebook for 12 years and was removed in August 2022 following public pressure.

Source: Tech Transparency Project

The Verge

POLICY

How Twitter's child porn problem ruined its plans for an OnlyFans competitor

Internal documents and Twitter employees reveal the need for massive investment to remove illegal content — but executives haven't listened

By **ZOE SCHIFFER** *and* **CASEY NEWTON**
Aug 30, 2022, 10:00 AM EDT | 0 Comments / 0 New

Kristen Radtke / The Verge; Getty Images

In the spring of 2022, Twitter considered making a radical change to the platform. After years of quietly allowing adult content on the service, the company would monetize it. The proposal: give adult content creators the ability to begin selling OnlyFans-style paid subscriptions, with Twitter keeping a share of the revenue.

Had the project been approved, Twitter would have risked a massive backlash from advertisers, who generate the vast majority of the company's revenues. But the service could have generated more than enough to compensate for losses. OnlyFans, the most popular by far of the adult creator sites, is projecting $2.5 billion in revenue this year — about half of Twitter's 2021 revenue — and is already a profitable company.

Some executives thought Twitter could easily begin capturing a share of that money since the service is already the primary marketing channel for most OnlyFans creators. And so resources were pushed to a new project called ACM: Adult Content Monetization.

Before the final go-ahead to launch, though, Twitter convened 84 employees to form what it called a "Red Team." The goal was "to pressure-test the decision to allow adult creators to monetize on the platform, by specifically focusing on what it would look like for Twitter to do this safely and responsibly," according to documents obtained by *The Verge* and interviews with current and former Twitter employees.

Executives are apparently well-informed about the issue, and the company is doing little to fix it

What the Red Team discovered derailed the project: Twitter could not safely allow adult creators to sell subscriptions because the company was not — and still is not — effectively policing harmful sexual content on the platform.

"Twitter cannot accurately detect child sexual exploitation and non-consensual nudity at scale," the Red Team concluded in April 2022. The company also lacked tools to verify that creators and consumers of adult content were of legal age, the team found. As a result, in May — weeks after Elon Musk agreed to purchase the company for $44 billion — the company delayed the project indefinitely. If Twitter couldn't consistently remove child sexual exploitative content on the platform today, how would it even begin to monetize porn?

Launching ACM would worsen the problem, the team found. Allowing creators to begin putting their content behind a paywall would mean that even more illegal material would make its way to Twitter — and more of it would slip out of view. Twitter had few effective tools available to find it.

Taking the Red Team report seriously, leadership decided it would not launch Adult Content Monetization until Twitter put more health and safety measures in place.

Twitter has not committed sufficient resources to detect, remove, and prevent harmful content from the platform

The Red Team report "was part of a discussion, which ultimately led us to pause the workstream for the right reasons," said Twitter spokeswoman Katie Rosborough.

But that did little to change the problem at hand — one that employees from across the company have been warning about for over a year. According to interviews with current and former staffers, as well as 58 pages of internal documents obtained by

The Verge, Twitter still has a problem with content that sexually exploits children. Executives are apparently well-informed about the issue, and the company is doing little to fix it.

"Twitter has zero tolerance for child sexual exploitation," Twitter's Rosborough said. "We aggressively fight online child sexual abuse and have invested significantly in technology and tools to enforce our policy. Our dedicated teams work to stay ahead of bad-faith actors and to help ensure we're protecting minors from harm — both on and offline."

While the Red Team's work succeeded in delaying the Adult Content Monetization project, nothing the team discovered should have come as a surprise to Twitter's executives. Fifteen months earlier, researchers working on the team tasked with making Twitter more civil and safe sounded the alarm about the weak state of Twitter's tools for detecting child sexual exploitation (CSE) and implored executives to add more resources to fix it.

The system that Twitter heavily relied on to discover CSE had begun to break

"While the amount of CSE online has grown exponentially, Twitter's investment in technologies to detect and manage the growth has not," begins a February 2021 report from the company's Health team. "Teams are managing the workload using legacy tools with known broken windows. In short (and outlined at length below), [content moderators] are keeping the ship afloat with limited-to-no-support from Health."

Employees we spoke to reiterated that despite executives knowing about the company's CSE problems, Twitter has not committed sufficient resources to detect, remove, and prevent harmful content from the platform.

Part of the problem is scale. Every platform struggles to manage the illegal materials users upload to the site, and in that regard, Twitter is no different. The platform, a critical medium for global communication with 229 million daily users, has the content moderation challenges that come with operating any large space on the internet and the added struggle of outsized scrutiny from politicians and the media.

But unlike larger peers, including Google and Facebook, Twitter has suffered from a history of mismanagement and a generally weak business that has failed to turn a profit for eight of the past 10 years. As a result, the company has invested far less in content moderation and user safety than its rivals. In 2019, Mark Zuckerberg boasted that the amount Facebook spends on safety features exceeds Twitter's entire annual revenue.

Meanwhile, the system that Twitter heavily relied on to discover CSE had begun to break.

For years, tech platforms have collaborated to find known CSE material by matching images against a widely deployed database called PhotoDNA. Microsoft created the service in 2009, and though it is accurate in identifying CSE, PhotoDNA can only flag known images. By law, platforms that search for CSE are required to report what they find to the National Center for Missing and Exploited Children (NCMEC), a government-funded nonprofit that tracks the problem and shares information with law enforcement. An NCMEC analysis cited by Twitter's working group found that of the 1 million reports submitted each month, 84 percent contain

newly-discovered CSE — none of which would be flagged by PhotoDNA. In practice, this means Twitter is likely failing to detect a significant amount of illegal content on the platform.

Twitter failed to remove the videos, "allowing them to be viewed by hundreds of thousands of the platform's users"

The 2021 report found that the processes Twitter uses to identify and remove CSE are woefully inadequate — largely manual at a time when larger companies have increasingly turned to automated systems that can catch material that isn't flagged by PhotoDNA. Twitter's primary enforcement software is "a legacy, unsupported tool" called RedPanda, according to the report. "RedPanda is by far one of the most fragile, inefficient, and under-supported tools we have on offer," one engineer quoted in the report said.

Twitter devised a manual system to submit reports to NCMEC. But the February report found that because it is so labor-intensive, this created a backlog of cases to review, delaying many instances of CSE from being reported to law enforcement.

The machine learning tools Twitter does have are mostly unable to identify new instances of CSE in tweets or live video, the report found. Until February 2022, there was no way for users to flag content as anything more specific than "sensitive media" — a broad category that meant some of the worst material on the platform often wasn't prioritized for moderation. In one case, an illegal video was viewable on the platform for more than 23 hours, even after it had been widely reported as abusive.

"These gaps also put Twitter at legal and reputation risk," Twitter's working group wrote in its report.

Rosborough said that since February 2021, the company has increased its investment in CSE detection significantly. She noted that it currently has four open

positions for child safety roles at a time when Twitter has slowed down its pace of hiring.

Earlier this year, NCMEC accused Twitter of leaving up videos containing "obvious" and "graphic" child sexual abuse material in an amicus brief submitted to the ninth circuit in *John Doe #1 et al. v. Twitter*. "The children informed the company that they were minors, that they had been 'baited, harassed, and threatened' into making the videos, that they were victims of 'sex abuse' under investigation by law enforcement," the brief read. Yet, Twitter failed to remove the videos, "allowing them to be viewed by hundreds of thousands of the platform's users."

This echoed a concern of Twitter's own employees, who wrote in a February report that the company, along with other tech platforms, has "accelerated the pace of CSE content creation and distribution to a breaking point where manual detection, review, and investigations no longer scale" by allowing adult content and failing to invest in systems that could effectively monitor it.

The years-long struggle to address CSE ran into a competing priority at Twitter: greatly increasing its user and revenue numbers

To address the issue, the working group called on Twitter executives to work on a series of projects. The group recommended that the company finally build a single tool to process CSE reports, collect and analyze related data, and submit reports to NCMEC. It should create unique fingerprints (called hashes) of the CSE it finds and share those fingerprints with other tech platforms. And it should build features to protect the mental health of content moderators, most of whom work for third-party vendors, by blurring the faces of abuse victims or de-saturating the images.

But even in 2021, before the company's tumultuous acquisition by Musk began, the working group acknowledged that mustering the necessary resources would be a challenge.

"The task of 'fixing' CSE tooling is daunting," they wrote. "[The Health team]'s strategy should be to chip away at these needs over time starting with the highest priority features to avoid the too-big-to-prioritize trap."

The project may have been too big to prioritize after all. Aside from enabling in-app reporting of CSE, there appears to have been little progress on the group's other recommendations. One of the research teams that had been most vocal about fixing Twitter's CSE detection systems has been disbanded. (Twitter's Rosborough says the team has been "refocused to reflect its core purpose of child safety" and has had dedicated engineers added to it.) Employees say that Twitter's executives know about the problem, but the company has repeatedly failed to act.

The years-long struggle to address CSE ran into a competing priority at Twitter: greatly increasing its user and revenue numbers. In 2020, the activist investor Elliott Management took a large position in Twitter in an effort to force out then-CEO Jack Dorsey. He survived the attempt, but to remain as CEO, Dorsey made three hard-to-keep promises: that Twitter would increase its user base by 100 million people, speed up revenue growth, and gain market share in digital advertising.

Dorsey quit as CEO in November 2021, having made little progress toward reaching those milestones. It was left to his hand-picked successor, former chief technology officer Parag Agrawal, to fulfill Elliott's demands.

Under its former head of product, Kayvon Beykpour, Twitter had spent the past few years adding products for creators. Last summer, it began rolling out "ticketed Spaces," letting users charge for access to its Clubhouse-like live audio product. The company added "Super Follows," a way for users to offer subscriptions for non-sexually explicit content, last September. In both cases, the company takes a percentage of the user's revenue, allowing the company to make money outside its core ad business.

"Adult content was a huge differentiator for Twitter, and for those [working] on revenue, it was an untapped resource."

While all of that unfolded, Twitter had become a major destination for another type of content: porn. In the nearly four years since Tumblr banned adult content, Twitter had become one of the only mainstream sites that allows users to upload sexually explicit photos and videos. It also attracted a significant number of performers who use Twitter to market and grow their businesses, using photos and short video clips as advertisements for paywalled services like OnlyFans.

"Adult content was a huge differentiator for Twitter, and for those [working] on revenue, it was an untapped resource," a former employee says.

Twitter is so important to the porn world that fears the company will eventually cave to external pressures and shut it down have regularly occasionally roiled the world of adult creators. In fact, though, by this spring, the company was considering a move that would make porn even *more* important to the platform — by placing it at the center of a new revenue plan.

Twitter already had Super Follows for non-explicit content, the thinking went. Why not add the feature for creators of adult content, too? The timing felt right, especially after OnlyFans alienated users by saying last year that it would ban adult content, only to reverse its stance a few days later.

Executives rarely discuss its popularity as a destination for adult content. (One document obtained by *The Verge* suggests the company has a strategy "to minimize focus and press" related to the subject.) But over the past two years, the company got very serious about adult content and began actively exploring an OnlyFans-like service for its users.

By this spring, the company was nearing a final decision. On April 21st and 22nd, Twitter convened another Red Team, this time for a project called Adult Creator Monetization, or ACM.

Twitter would have several strengths if it decided to compete with OnlyFans, the Red Team found. Adult creators have a generally favorable attitude toward the company, thanks to how easy Twitter makes it for them to distribute their content. The project was also "consistent with Twitter's principles in free speech and freedom of expression," they said. Finally, the company was planning to obtain a money transmitter license so it could legally handle payments.

Given the size of the opportunity, the Red Team wrote, "ACM can help fund infrastructure engineering improvements to the rest of the platform."

But the team found several key risks as well. "We stand to lose significant revenue from our top advertisers," the team wrote. It speculated that it could also alienate customers and attract significant scrutiny from Congress.

The biggest concerns, though, had to deal with the company's systems for detecting CSE and non-consensual nudity: "Today we cannot proactively identify violative content and have inconsistent adult content [policies] and enforcement," the team wrote. "We have weak security capabilities to keep the products secure."

Twitter has had several high-profile data breaches. Eventually, Twitter abandoned the project.

Fixing that would be costly, and the company would be likely to make enforcement errors. Non-consensual nudes, they wrote, "can ruin lives when posted and

monetized."

Moreover, the report said, "There are several challenges to maintaining this as a top priority. ... We're thinking about health as a parallel to monetization, instead of as a prerequisite."

Beykpour, Twitter's former head of product, had pushed Twitter to roll out Real ID — a feature that would require users to upload government documents to prove their identity. If Twitter wanted to monetize adult content, it would need to verify the ages of the people creating that content, as well as the people watching it. But employees had already determined that Real ID presented serious problems. Matching IDs with government databases was expensive and required a secure network. Twitter has had several high-profile data breaches. Eventually, Twitter abandoned the project.

Soon, the group's priorities would change completely. On August 23rd, Twitter announced that the health team would be reorganized and combined with a team tasked with identifying spam accounts. The move came amid increasing pressure from Elon Musk, who claimed the company was lying about the number of bots on the platform.

"It was a gut punch," says a former researcher on the team. "For Elon Musk to declare that spam was the single most important question that needed to be answered in order for him to buy the company is ludicrous."

But Twitter's troubles with Musk — and the internal chaos they would cause — were just beginning.

JOIN THE CONVERSATION 0

FEATURED VIDEOS FROM THE VERGE

Sony made a gaming monitor for PC and PS5

More from **Policy**

Bannon surrenders to New York prosecutors over border wall crowdfunding campaign

Here's how much energy crypto mining gobbles up in the US

Kiwi Farms has been scrubbed from the Internet Archive

Former Conti ransomware gang members helped target Ukraine, Google says

1 We finally got our hands and eyes on the PlayStation VR2
JAY PETERS SEP 14

2 Microsoft was right all along
MONICA CHIN SEP 14

3 Listen to our podcast about our website redesign
ANDREW MARINO SEP 14

4 Apple iPhone 14 Pro review: early adopter island
NILAY PATEL SEP 14

5 Welcome to the new Verge
NILAY PATEL SEP 13

Today's Storystream

FEED REFRESHED 59 MINUTES AGO • WELCOME TO THE NEW VERGE

K

KEVIN NGUYEN 59 MINUTES AGO

You don't have to be a billionaire. Patagonia founder Yvon Chouinard is giving ownership of his company away to "a specially designed set of trusts and nonprofit organizations," ensuring its roughly $100 million a year in profits go to fighting climate change.

Billionaire No More: Patagonia Founder Gives Away the Company

[THE NEW YORK TIMES]

D

DAN SEIFERT TWO HOURS AGO

"Maybe it's Photonic Engine; maybe it's just good old-fashioned physics." That's my favorite line from Allison Johnson's iPhone 14 review — it's really hard to tell what's driving the improvements to the camera more, the better lens and sensor or Apple's ridiculously-named "Photonic Engine."

Don't sleep on the video either, which features a delightful intro.

Apple iPhone 14 review: meet the iPhone 13S
ALLISON JOHNSON 9:00 AM EDT

APPLE

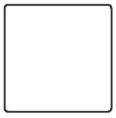
Apple iPhone 14 Pro review: early adopter island

NILAY PATEL 9:00 AM EDT

APPLE

How to use iPhone's Safety Check and Emergency SOS features

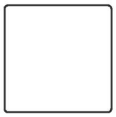
BARBARA KRASNOFF
43 MINUTES AGO



GAMING

Intel's DLSS rival, XeSS, seems to be a succeSS

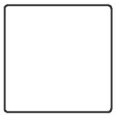
MITCHELL CLARK AN HOUR AGO



APPLE

iOS 16.1 beta adds Apple's ugly new battery percentage indicator to the iPhone Mini

JAY PETERS TWO HOURS AGO



Everyone knows what YouTube is – few know how it really works

NILAY PATEL SEP 13

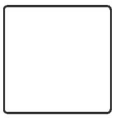
- **The HP Elite Dragonfly Chromebook is incredible – with one big problem**
  MONICA CHIN SEP 13
- **How to use iPhone's Safety Check and Emergency SOS features**
  BARBARA KRASNOFF SEP 14

SEE ALL TECH

APPLE

The best laptop deals you can get right now

SHEENA VASANI TWO HOURS AGO



Welcome to the new Verge

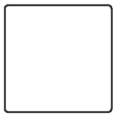
Revolutionizing the media with blog posts

NILAY PATEL SEP 13

MICROSOFT

Microsoft Teams now has a remixed ringtone from TikTok

TOM WARREN TWO HOURS AGO



R

RICHARD LAWLER TWO HOURS AGO

Google's getting wary of experimenting. 7 of the 14 projects at Google's startup incubator Area 120 — where employees spent 100 percent of their time working on 20 percent projects — have been cut, according to *TechCrunch*, sending workers looking for spots on other projects to stay employed.

Google's change of heart isn't that surprising, as the 20 percent rule Area 120 was named for — which spawned projects like Gmail, AdSense, and Google News — has been dead since 2013.

Google cancels half the projects at its internal R&D group Area 120

[TECHCRUNCH]

A

ANDREW J. HAWKINS TWO HOURS AGO

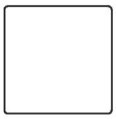
Tesla is scrambling to figure out how to make more EV batteries in the US. *The Wall Street Journal* reports that the electric automaker is pausing its plans to make batteries at

its Gigafactory in Berlin. The shift in strategy came about after the passage of the Inflation Reduction Act, which includes $10 billion in tax credits for companies who build clean-tech facilities in the US. The company is also exploring building a lithium processing plant on the Gulf Coast. Meanwhile, Tesla's vehicles will re-qualify for the $7,500 EV tax credit starting January 1st. (The company was the first to lose eligibility back in 2019.)
Tesla Shifts Battery Strategy as It Seeks U.S. Tax Credits
[WSJ]

SMART HOME

Lockly Vision Elite review: two become one

JENNIFER PATTISON TUOHY
TWO HOURS AGO



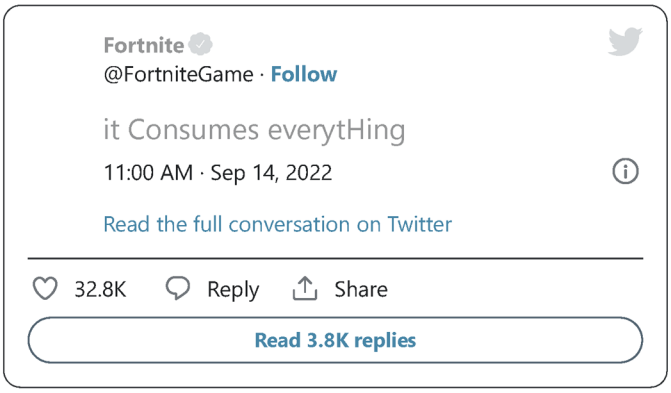
DAVID PIERCE SEP 3

SEE ALL PODCASTS

J

JAY PETERS 2:40 PM EDT

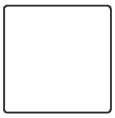
Fortnite is teasing its next season, which will seemingly feature a weird chrome goop. Capital letters in a Twitter thread spell out the word CHROME. Some teaser images show hands reaching out from under the weird goop, including one that could be Gwen Stacy from Spider-Man. (Perhaps she'll be a battle pass skin?) Epic even made a strange TikTok video where the goop consumes a cereal box. The new season, titled Paradise, kicks off on September 18th.



MICROSOFT

Canva's new Visual Worksuite has its sights set on Google and Microsoft

JESS WEATHERBED 2:37 PM EDT

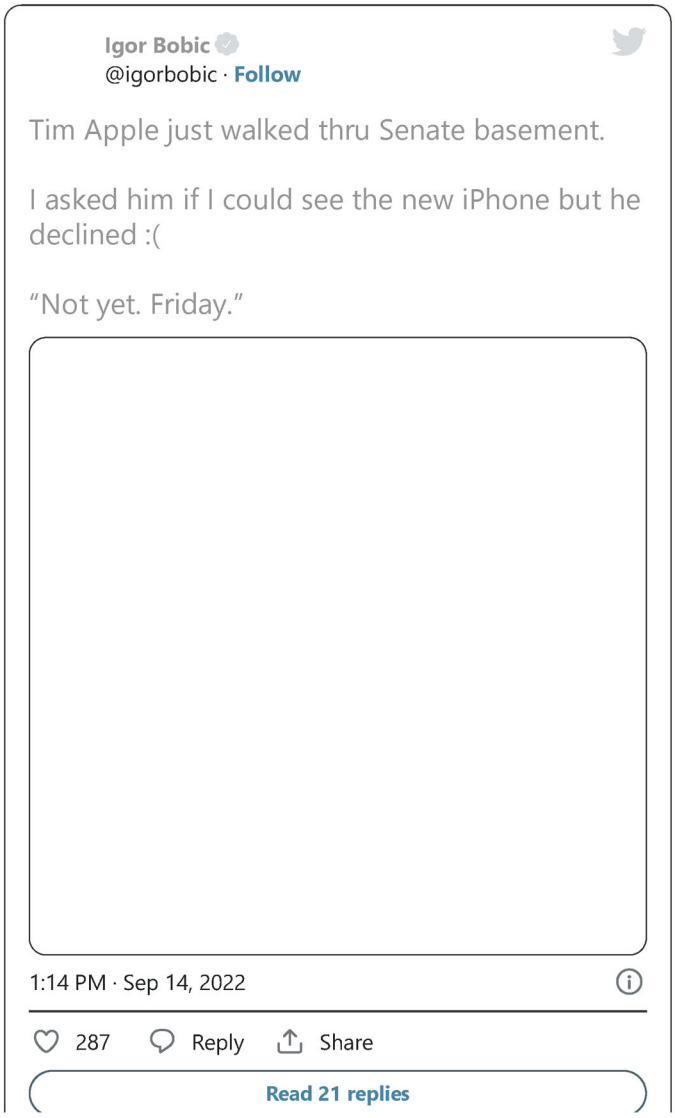


M

MAKENA KELLY 2:32 PM EDT

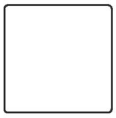
Tim Cook was at the US Capitol today. It appears to be an unannounced visit — unrelated to the Homeland Security

hearing going on at the same time. Do you know why Mr. Apple went to Washington? Let me know!



POLICY

California governor signs law requiring social networks to post moderation rules



ARI ROBERTSON, 9:00 PM EDT

ADI ROBERTSON 2:20 PM EDT

DAVID PIERCE 2:00 PM EDT

Still wondering why this site works the way it does? Here's the inside story. On today's Vergecast, we talked all about the new site, the future of The Verge, the future of news, the future of the internet, the future of everything, and our feelings about all of it.

Websites are back: inside Th

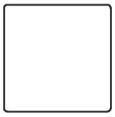
Sep 14 · The Vergecast

1:06:56

TECH

WhatsApp made a movie with Giannis Antetokounmpo

JAY PETERS 1:32 PM EDT



The Verge Most Popular

1 **Apple iPhone 14 Pro review: early adopter island**
NILAY PATEL SEP 14

2 **Welcome to the new Verge**
NILAY PATEL SEP 13

3 **iOS 16 review: unlocking the lock screen**
DAVID PIERCE SEP 12

4 **We finally got our hands and eyes on the PlayStation VR2**
JAY PETERS SEP 14

5 **Apple iPhone 14 review: meet the iPhone 13S**
ALLISON JOHNSON SEP 14

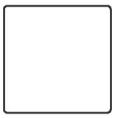
TOM WARREN 1:30 PM EDT

Yo dawg, Microsoft heard you like widgets. So it put some widgets in your widgets so you can see weather while you check your stocks and widget around. (Microsoft is testing a fullscreen widgets board for Windows 11).

Look at all those widgets! Image: Microsoft

MICROSOFT

Xbox app on PC now launches faster and has HowLongToBeat integration
TOM WARREN 1:26 PM EDT



ALEX CRANZ 1:11 PM EDT

Animation continues to be a casualty of the streaming wars. First, Netflix canceled a number of animation projects and laid off 70 employees, then HBO Max pulled multiple animated shows leaving creators and fans shocked and with

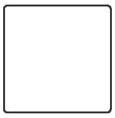
little recourse.

Now *Deadline* has confirmed that 30 more employees are being laid off from Netflix Animation. Animation is historically a pricier form of entertainment than reality TV [illegible] Netflix Animation Lays Off 30 As Overhaul Continues [DEADLINE] [illegible]ely, a prime focus for companies looking to balance their books.

TECH

Microsoft was right all along

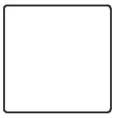
MONICA CHIN 1:04 PM EDT



DEALS

Today, for three hours only, you can get Samsung's M8 Smart Monitor for $499

CAMERON FAULKNER 12:56 PM EDT



REVIEWS

- **iOS 16 review: unlocking the lock screen**
  DAVID PIERCE SEP 13
- **EcoFlow Delta Pro battery review: maximum solar power for an uncertain world**
  THOMAS RICKER SEP 10

SEE ALL REVIEWS

MITCHELL CLARK 12:50 PM EDT

The next Amazon union election is coming up. On October 12th, workers at the ALB1 warehouse in Albany, New York

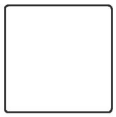
will start voting on whether to unionize with the Amazon Labor Union — the same group that successfully organized the JFK8 warehouse in Staten Island. Votes will be counted on October 18th, according to the National Labor Relations Board.

Amazon warehouse workers in Albany have filed to unionize
JAY PETERS AUG 16

GOOGLE

Memories in Google Photos are getting a new look and fresh features

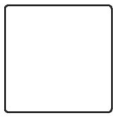
JESS WEATHERBED 12:30 PM EDT



DEALS

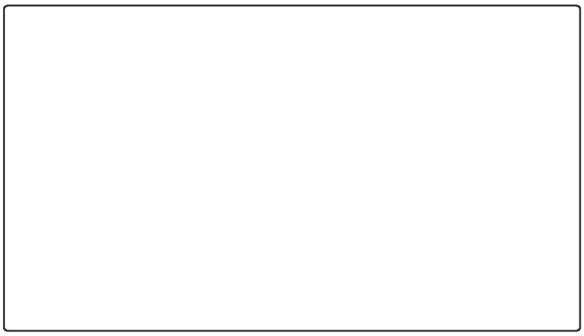
Newegg has so many extra GPUs that it's bundling them with free monitors

ALICE NEWCOME-BEILL 12:26 PM EDT



JAY PETERS 12:20 PM EDT

Supposed RTX 4090 pic shows some big GPUs on the way. Zotac's upcoming GPUs look like they're going to have some curves, based a photo posted to Chinese social network Baidu (via *PC Gamer*). Nvidia CEO Jensen Huang is keynoting the company's GTC 2022 conference on September 20th, so maybe we'll hear official details about RTX 40-series cards then.



I don't hate the curves. Image: Baidu (via *PC Gamer*)

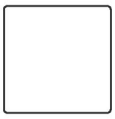
A **ANDREW WEBSTER** 12:15 PM EDT

The Toronto International Film Festival 2022 is underway and we have reviews. TIFF kicked off last week, and I've been braving the crowds to check out an unhealthy amount of movies since then. Watching three movies a day for a week is hard work, I swear. You can read my thoughts on some of my favorites so far: *Glass Onion: A Knives Out Mystery*, *Weird: The Al Yankovic Story*, and *Pearl*.

APPLE

How to quickly update all of your iPhone apps on iOS 16

SHEENA VASANI 12:03 PM EDT



SCIENCE

Flo period tracker launches 'Anonymous Mode' to fight abortion privacy concerns

NICOLE WETSMAN SEP 14

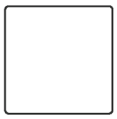
- **NASA has a new launch date for its Artemis I megarocket**
  EMMA ROTH SEP 13
- **Elon Musk's SpaceX and Tesla emails are for his eyes only**
  ELIZABETH LOPATTO SEP 13

SEE ALL SCIENCE

MOVIE REVIEW

Pearl is a slasher prequel that makes the original even better

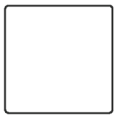
ANDREW WEBSTER 12:00 PM EDT



GAMING

Discord starts rolling out new Forum Channels to help organize conversations

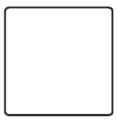
TOM WARREN 12:00 PM EDT



TV SHOWS

HBO Max and Discovery Plus' reinvention might come with some new costs

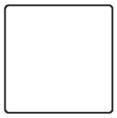
CHARLES PULLIAM-MOORE
11:50 AM EDT



APPLE

Google's Pixel Buds Pro are 10 percent off exclusively for Verge readers

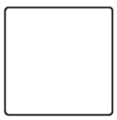
ANTONIO G. DI BENEDETTO
11:29 AM EDT



TECH

Zoom is reportedly working on calendar and email tools to take on Office and Google

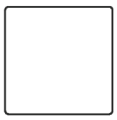
DAVID PIERCE 11:16 AM EDT



GAMING

Sega announces Yakuza 8 and a slew of other Yakuza games

ASH PARRISH 11:14 AM EDT



FACTORS

Microsoft Teams now has a remixed ringtone from TikTok

TOM WARREN SEP 14

- **The GoPro Hero 11's big upgrade is a smaller version**
  MITCHELL CLARK SEP 14
- **Patreon is laying off 17 percent of its workforce and closing offices**
  MITCHELL CLARK SEP 13

SEE ALL CREATORS

ADI ROBERTSON 11:05 AM EDT

South Korea fined Meta and Google for using personal info without consent. Meta was fined around $22 million and Google around $50 million — a fraction of their revenue, but part of a protracted global crackdown on lax privacy policies. Meta says it's considering fighting the decision in court.

S.Korea fines Google, Meta billions of won for privacy violations

[REUTERS]

GAMING

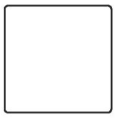
The Sims 4 base game is going to be free next month

ANDREW WEBSTER 11:00 AM EDT

PODCASTS

Listen to our podcast about our website redesign

ANDREW MARINO 10:50 AM EDT



ELIZABETH LOPATTO 10:50 AM EDT

Crypto villain Do Kwon is wanted by authorities in South Korea. Kwon, the founder of Terraform Labs, was at the center of crypto's biggest scandal this year: the cataclysmic collapse of Luna/Terra coins from $60 billion to zero. South Korean authorities issued the warrant for his arrest on "allegations that include violations of the nation's capital markets law," *Bloomberg* reports. Kwon is in Singapore.

Terra Co-Founder Do Kwon Faces Arrest Warrant in South Korea

[BLOOMBERG.COM]

MORE STORIES

“We have the most robust vaccine safety system we have ever had in this country.”

CDC Director Rochelle Walensky
Testimony to the Senate Committee on Health, Education, Labor, & Pensions (11/04/2021)

DRUG ADVERSE EVENT COMPARISON

FDA AND CDC DATA: WORLDWIDE

	Adverse events	Deaths	Deaths/year
1/1/1996 – 9/30/2021:			
Ivermectin	3,756	[illegible]	15
HCQ	[illegible]	[illegible]	69
Flu vaccines	[illegible]	2,001	77
Dexamethasone	[illegible]	15,910	618
Tylenol	112,244	26,356	1,024
Since 20[illegible]			
Remdesivir	6,504	1,612	921
In 12 months:			
Covid vaccines	1,000,229	21,002	21,002

CENSORED

FDA FAERS system, CDC VAERS system. Reports from all locations worldwide. Data as of Dec. 24, 2021; downloaded Jan. 1, 2022.

RON JOHNSON U.S. SENATE

DRUG ADVERSE EVENT COMPARISON

FDA AND CDC DATA: WORLDWIDE

	Adverse events	Deaths	Deaths/year
1/1/1996 – 6/30/2022:			
Ivermectin	**4,032**	**415**	**16**
HCQ	**26,294**	**1,936**	**73**
Flu vaccines	**197,936**	**2,034**	**77**
Dexamethasone	**93,880**	**17,627**	**665**
Tylenol	**121,903**	**28,276**	**1,067**
Since 2020:			
Remdesivir	**7,846**	**1,906**	**953**
In 20 months:			
Covid vaccines	**1,400,350**	**30,796**	**18,115**

FDA FAERS system, CDC VAERS system. Reports from all locations worldwide. Data as of Sept. 2, 2022; downloaded Sept. 9, 2022.

RON JOHNSON U.S. SENATE

Public Health England

Protecting and improving the nation's health

SARS-CoV-2 variants of concern and variants under investigation in England

Technical briefing 23

17 September 2021

Table 5. Attendance to emergency care and deaths of sequenced and genotyped Delta cases in England by vaccination status (1 February 2021 to 12 September 2021)

Variant	Age group (years)**	Total	Cases with specimen date in past 28 days	Unlinked	<21 days post dose 1	≥21 days post dose 1	≥14 days post dose 2	Un-vaccinated
Delta cases	<50	497,105	119,611	49,527	30,359	83,009	85,407	248,803
	≥50	95,587	35,596	7,602	314	7,129	71,991	8,551
	All cases	593,572	155,252	58,003	30,674	90,138	157,400	257,357
Cases with an emergency care visit§ (exclusion‡)	<50	16,709	N/A	167	1,051	2,494	2,518	10,479
	≥50	5,445	N/A	21	30	448	3,747	1,199
	All cases	22,162	N/A	196	1,081	2,942	6,265	11,678
Cases with an emergency care visit§ (inclusion#)	<50	22,719	N/A	273	1,364	3,060	3,162	14,860
	≥50	10,102	N/A	50	64	755	6,532	2,701
	All cases	32,834	N/A	336	1,428	3,815	9,694	17,561
Cases where presentation to emergency care resulted in overnight inpatient admission§ ((exclusion‡)	<50	3,490	N/A	95	174	352	453	2,416
	≥50	2,784	N/A	10	18	184	1,908	664
	All cases	6,280	N/A	111	192	536	2,361	3,080
Cases where presentation to emergency care resulted in overnight inpatient admission§ (inclusion#)	<50	6,230	N/A	144	283	565	721	4,517
	≥50	6,167	N/A	33	42	393	3,913	1,786
	All cases	12,407	N/A	187	325	958	4,634	6,303
Deaths within 28 days of positive specimen date	<50	204	N/A	7	6	11	48	132
	≥50	2,336	N/A	32	11	138	1,565	590
	All cases	2,542	N/A	41	17	149	1,613	722

63.5 %

This platform has repeatedly censored and manipulated the normal flow of public discourse and information-sharing with false warnings about foreign interference and disinformation.
Big tech has a clear and alarming pattern of speech suppression carried out at the direction of agencies and officials in the Federal government. In other words, the Federal government is using our platform to violate your First Amendment rights.

Google

Mark Isakowitz
VP, Government Affairs & Public Policy, US and Canada
Google LLC
25 Massachusetts Ave., NW - 9th Floor
Washington, DC 20001

February 3, 2022

The Honorable Rick Scott
United States Senate
502 Hart Senate Office Building
Washington, D.C. 20510

The Honorable Gus Bilirakis
United States House of Representatives
2354 Rayburn House Office Building
Washington, D.C. 20515

Dear Senator Scott and Representative Bilirakis:

I write in response to your letter of January 19, 2022 regarding Google's efforts to combat the sale of illicit drugs online. We appreciate the opportunity to respond directly to your concerns on this important topic. Google is combining our products and technology with government and non-profit efforts to empower families and individuals to prevent, treat, and recover from substance use disorders related to opioids, fentanyl, and other prescription and illicit drugs.

Partnering with the U.S. Government. Our policies and procedures are designed to be dynamic and responsive to emerging trends in substance abuse disorders and the online sale of drugs. We are proud that these policies and processes have benefitted from feedback and close partnerships with law enforcement and regulatory agencies in the United States, including the Department of Health & Human Services (HHS), the Substance Abuse and Mental Health Services Administration (SAMHSA), the Office of the Surgeon General (OSG), and Drug Enforcement Administration (DEA), and the Food and Drug Administration (FDA). These partnerships have resulted in innovative and creative techniques to combat the sale of illicit drugs, and to make the Internet safer and more useful for users seeking treatment and recovery.

Enforcing Laws and Regulations. Google is proud to work with the FDA, the DEA, and other regulatory and law enforcement agencies involved in enforcing laws and regulations concerning the sale of drugs online. For many years, Google has made numerous referrals to law enforcement, including the FDA's Office of Criminal Investigations and the DEA's Special Operations Division.

We provide proactive referrals to the FDA and DEA in different ways, including phone calls, meetings, and emails. Although we typically provide notice to our users prior to disclosing their information in response to a request from a government agency, we will not give notice when we are legally prohibited under the terms of the request, such as a non-disclosure order.

Policies and Removals. Google does not allow advertising for prescription opioid painkillers except for those that are intended for use as medication-assisted treatment (MAT) for opioid use disorder and meet all other requirements pertaining to prescription drugs under our healthcare and medicines policy. Google allows advertisements for opioid drug terms by treatment services to assist with substance abuse disorders. Substance abuse disorder treatment services that want to run ads using drug terms related to opioids must be certified under our Addiction Services policy.

In other limited cases, we allow exceptions to this policy for public health and safety awareness campaigns from governmental or well-established non-profit health advocacy organizations. Advertisers that qualify for these exceptions can request certification with Google through the Healthcare Certification Form.

Google contracts with LegitScript, an independent company, to significantly increase the breadth and scope of non-pharmaceutical products prohibited by Google's policies from being surfaced on its products, including in ads, shopping, and YouTube. LegitScript keeps track of thousands of internet pharmacies and dangerous health products, and tracks merchants of dangerous health products such as "legal highs" and tainted supplements.

LegitScript also runs sweeps which it reports to Google for enforcement. These notices are received from LegitScript twice a week and have resulted in the removal of tens of thousands of ads from Google's systems.

We have also worked with LegitScript to monitor and prohibit thousands of new products, ingredients, and psychoactive substances, regardless of legal status, based on legitimate concerns about the safety of these products and substances from health professionals, regulators, and law enforcement around the world. LegitScript also runs weekly sweeps on YouTube for videos of purported offers to sell illicit opioids which are reported for removal.

Empowering Families and Individuals Through Products and Technology. Google works hard to keep people safe both online and offline, and our products and technology reflect our commitment to combat the opioid and fentanyl epidemic. Beginning in 2018, we partnered with the DEA on National Prescription Drug Take Back Day, promoting the DEA's initiative across Google properties twice a year. We developed, for example, a Google Maps API locator tool to help people dispose of their prescription drugs at temporary event locations. With the help of this tool, the DEA and its local partners have collected 7,000 tons of unused medication since the program's inception.

In addition to National Prescription Drug Take Back Day occurring twice a year, we make it easier for Americans to quickly find disposal locations on Google Maps and Search year round. Queries on Google Search for "drug drop off near me" and "medication disposal near me" display permanent disposal locations at pharmacies, hospitals and government buildings.

We prominently surface a display box on Google Search containing SAMHSA's national helpline number and a link to the federal website FindTreatment.gov whenever someone queries for treatment related to substance abuse disorders. Google has expanded this functionality to similar queries on YouTube. We also are

committed to helping all people lead better lives with our Recover Together website, which includes a searchable map to find nearby recovery groups and support resources for people in recovery and their families.

Google has also created knowledge panels that surface in response to search queries relating to opioid addiction. The panels provide information on prevalence and dangers of opioid abuse and dependence, but also on the symptoms of substance abuse disorder and treatment options. We also surface a text box in Google and YouTube search results containing information about the non-profit Partnership to End Addiction with options to call, text, or email a helpline specialist in response to search queries relating to teen drug addiction. Across Search and YouTube, we provide financial support of more than one million dollars for the Partnership's public service announcements, including the organization's "Start the Connection" campaign, aimed at families dealing with the opioid epidemic.

In addition, Google is a founding member of the Center for Safe Internet Pharmacies (CSIP), a nonprofit organization committed to raising public awareness about the dangers associated with illicit online pharmacies, such as identity theft and counterfeit medications. Google sponsors CSIP's online advertisements that lead to their consumer education website, www.verifybeforeyoubuy.org, to raise awareness to those shopping online for medications.

Finally, Google is proud to host a variety of free mobile apps by the FDA on Google Play, including those designed to educate consumers by providing them with product information. By having increased access to information through their mobile devices, we hope families can make more informed decisions concerning drugs and products they buy and use.

* * *

We appreciate your attention to this issue, and we hope our response has provided you a more fulsome picture of Google's efforts to combat the online sales of opioids and fentanyl. We are committed to constantly improving, innovating, and collaborating to combat this public health crisis.

Sincerely,

Mark Isakowitz
Vice President
Government Affairs and Public Policy, US and Canada

Post-Hearing Questions for the Record
Submitted to Mr. Geoffrey Cain
From Senator Rob Portman

"Social Media's Impact on Homeland Security"
September 14, 2022

1. In June, it was announced that TikTok will migrate its U.S. user data to Oracle servers located in the United States. What is the relationship between Oracle and TikTok?

Geoffrey Cain: In September 2020, TikTok announced that it had selected American technology giant Oracle as a "technology partner," restructuring its operations with Oracle bidding to purchase part of TikTok's U.S. operations, as part of a CFIUS (Committee on Foreign Investment in the United States) agreement being negotiated under the Trump administration.

Oracle didn't purchase TikTok in the end. Instead, TikTok struck an agreement with Oracle to migrate Americans' data to Oracle servers in the U.S. It was trying to convince the U.S. government that the personal data of Americans would not end up in the hands of China's government. According to a report in *The New York Times*, the Biden administration has been negotiating a new CFIUS deal with TikTok that would require TikTok to complete the migration of data to Oracle's servers.

a. What is the relationship between Oracle and the Chinese Communist Party?

Oracle, despite being an American company, is a dubious data protection partner for TikTok because of Oracle's inappropriate business dealings with Chinese authorities. There is strong reason to doubt the private data of Americans on TikTok will be completely safe under the watch of Oracle's auditors. Mara Hvistendahl, a longtime China journalist at the investigative news website *The Intercept*, has documented Oracle's egregious conflicts of interest selling data analytics software to Chinese police authorities for mass surveillance.

These conflicts of interest and split loyalties between China's hostile authoritarianism and America's homeland security run deep. Oracle has inappropriately advertised its software services for the U.S. Department of Defense to potential Chinese police and security clients. Oracle has offered China's Ministry of Public Security, the powerful, rights-abusing policing body, the data analytics software that undergirds China's 1984-style surveillance dystopia and crimes against humanity. This includes marketing software directly to Chinese police authorities in Xinjiang, where they are carrying out genocide against the minority Uyghur population

b. Will the migration of data from TikTok's servers to Oracle's servers alleviate any concerns that the Chinese Communist Party can access Americans' TikTok data?

Even if TikTok stores the data on Oracle's servers in America, TikTok's deep exposure to China makes that data susceptible to the vague, powerful data collection laws that give the Chinese government sweeping powers. If the Party or its intelligence bodies demand this data in secret,

employees at both TikTok and its parent company ByteDance in China will have few ways of resisting through legitimate court hearings and court appeals in China. The sad reality is that since ByteDance and TikTok have employees in China, those employees will be required to conform to the demands of the Chinese Communist Party to hand over data. This risk is not present with large American social media giants such as Facebook, which are blocked in China and are less likely to submit to Chinese legal demands as a result.

2. TikTok's Chief Product Officer, Vanessa Pappas testified that TikTok does not abide by Chinese law, because they do not "operate" within China. Do you believe that is misleading? Why or why not?

Ms. Pappas's testimony was deceptive and self-contradictory to a puzzling degree. Ms. Pappas also stated in her testimony that TikTok has employees in China. These employees must abide by Chinese law, including China's National Intelligence Law, which requires citizens to partake in intelligence operations at the request of the government, along with a host of similar draconian laws.

Furthermore, a report in *BuzzFeed* revealed that Chinese software engineers based in China, including an unnamed individual referred to in leaked audio tapes as the "Master Admin," have had access to Americans' data, despite TikTok's repeated claims to the contrary. Because this data was accessed in China, it could easily be stored in China and subject to Chinese law. In October 2022, *Forbes* reported that a ByteDance team in Beijing planned to use TikTok to monitor the locations of specific US citizens, a planned act by individuals who fell directly under Chinese legal jurisdiction.

On November 28, 2022, *Forbes* further reported that a leaked internal audit from ByteDance, from 2021, urged TikTok and ByteDance executives to change their practices or risk being criminally charged for making potentially false representations to the US government. I believe this explains TikTok's contradictory statements about the applications of Chinese law to their employees based in China. The auditors found it impossible to verify the accuracy of information that ByteDance reported to government agencies because of the risks of Chinese ownership. This means that TikTok executives in America could not verify their statements made to the US government, too, but they went ahead and knowingly deceived the US government.

3. Ms. Pappas also twice denied requests to cut off U.S. user data and metadata flows to China or any other parties associated with Chinese-based entities. If data continues to be shared with Chinese-based entities (regardless of the outcome of the CFIUS process), do you believe it will remain vulnerable to exploitation or access by the Chinese Communist Party?

I believe that Ms. Pappas denied these requests because TikTok realizes it is impossible to cut off U.S. user data and metadata flows to China, despite earlier promises to the contrary. For instance, the internal audit revealed in *Forbes* concluded that it was impossible to keep sensitive data from being stored in China. There is no reason to conclude that this has changed, because *BuzzFeed* reported that China-based engineers had access to Americans' data as recently as

January 2022. All this data will remain vulnerable to exploitation and access by the Chinese Communist Party, regardless of the outcome of the CFIUS deal, and regardless of the migration of data to US-based servers audited by Oracle.

1 Hacker Way
Menlo Park, CA 94025
United States

November 4, 2022

Chairman Gary Peters
Ranking Member Rob Portman
US Senate Committee on Homeland Security and Governmental Affairs
340 Dirksen Senate Office Building
Washington DC, 20510

Dear Chairman Peters, Ranking Member Portman, and Members of the Committee:

Thank you for your questions for the record from the Committee on Homeland Security & Governmental Affairs' September 14, 2022 hearing entitled "Social Media's Impact on Homeland Security." Per your request, attached are the answers for the record to your questions, including those provided in Meta's initial submission on October 28.

Sincerely,

Meta Platforms, Inc.

Questions from Chairman Peters

Question 1. **Please provide the following information regarding Meta's current employees as of September 1, 2022: (i) the total number of full-time engineers at your company, (ii) the total number of full-time engineers working full time on ensuring trust and safety or integrity of your platforms, and (iii) the total number of full-time engineers working full time on product development.**

We estimate that, as of the third quarter of 2022, there were approximately 40,000 full-time engineers at the company. There are a wide variety of engineering roles at Meta, including roles not directly related to traditional software engineering.

Of those approximately 40,000 full-time engineers, we estimate that, as of the third quarter of 2022, over 3,000 work full-time on ensuring the trust and safety or integrity of our platforms. However, as Mr. Cox testified during the hearing, our trust and safety and integrity-related work is not limited to the engineers and other employees (including product managers, designers, researchers, data scientists, and others) who work full-time on those issues. We approach this work comprehensively, relying on numerous teams across apps and the company. Safety and integrity are key components of the user experience, and we both build our apps and continually update them with safety and integrity in mind. This approach is built into the DNA of the company.

We estimate that approximately half of the full-time engineers at Meta as of the third quarter of 2022 work on developing our family of apps and building the metaverse. This estimate does not include, for example, engineers who focus on building and improving our internal infrastructure. It is important to note that engineers working on trust and safety overlap with those developing our apps. For purposes of this response, this estimate excludes the over 3,000 engineers who work full-time on trust and safety, the vast majority of whom do so in the context of developing our family of apps. Regardless of how they are categorized, we expect all of our engineers to understand the impact of their work on safety and integrity.

As we have previously mentioned, because we embed integrity teams into various app groups, our integrity teams take safety and integrity considerations into account during app development, and they advise app teams on best practices. Many changes also go through an Integrity Review process, a cross-functional process by which proposed app changes are evaluated on integrity criteria prior to launch. This process allows us to identify and anticipate potential abuses and build in ways to mitigate them from the start. As discussed in the response to Question 2 below, engineers also look at a variety of metrics, including integrity metrics, as they do their work. New apps or features are reviewed against a set of integrity standards to help ensure protection and a positive experience for our users.

Question 2. **Please provide all metrics evaluated in your A/B testing for each of the last 2 years, including a description of each metric and the percentage of A/B tests in which the metric was evaluated.**

A/B testing is a process of showing multiple variants of the same app randomly to users and comparing their responses across different variants. These tests consider a number of factors that help us understand whether the change is likely to improve people's experience on our platform.

Generally, before any code is put into production, either to update existing functionality or build a new feature, we undertake various review processes depending on the change or feature. For example, development teams may go through an Integrity Review, a process by which we evaluate potential app changes on relevant integrity criteria prior to launch.

After that, our standard development practice is to set up an A/B experiment. At any given time, we have over ten thousand such experiments running. A/B experiments include certain core metrics, which include integrity metrics such as the prevalence of certain types of violating content, but each experiment considers hundreds of additional metrics on average chosen by the development team, out of the hundreds of thousands of metrics that have been used before (or new metrics the developer chooses to use). Therefore, it is not practicable to provide every metric for every A/B test over the last two years.

While the experiment is running, alerts may be sent if an experiment significantly impacts a metric that may harm our users—for example, integrity teams may be alerted if an experiment had a significant impact on integrity metrics. Some alerts may automatically trigger the experiment to stop.

After testing, there is generally an additional review process to evaluate the metrics that were observed and determine whether or not the change should be launched to all users as a permanent feature. Many changes for which we run experiments end up not launching, for a variety of reasons. For changes that do move forward, the review and testing processes allow us to identify and anticipate potential abuses and build in mitigations from the start. And as Mr. Cox stated in his testimony, if we felt that safety was a concern with an app or feature, we would not launch it.

Question 3. **Does Meta use metrics to measure whether certain demographics see a disproportionate amount of harmful content on your platforms, such as, for example, the number of users who have repeated exposures to harmful content, or the amount of harmful content seen by the user in the 99th percentile of exposure to harmful content? If yes, please share a list of these metrics and their current values and values from last year.**

As a general matter, we do not track data in the manner described in the question.

Rather, we focus on prevalence. Prevalence tells us how often content that violates our standards is seen relative to the total amount of times any content is seen on Facebook. It is how we objectively evaluate our progress and reflects our efforts to reduce hate speech distribution. Some pieces of content get many more views than others, so removing that content likely has a greater effect on how much hate speech is actually seen by users. Meta's enforcement teams typically prioritize review of content that is likely to accumulate a high number of views, and therefore the removals Meta makes may have an outsized positive effect on reducing user encounters of hate speech. We talk about and report on prevalence in our Community Standards Enforcement Report every quarter and describe it in our Transparency Center.

While we know our work will never be done in this space, the fact that prevalence for hate speech has been reduced from 0.11% when we first began reporting it in the third quarter of 2020 to 0.02% for the second quarter of 2022 shows that taken together, our efforts are having an impact. As reported in our Community Standards Enforcement Report, we can attribute a significant portion of the drop to our improved and expanded AI systems.

Question 4. **What is the total number of views that violative content has received across your platforms over the last year, month, and day? Please break down by category of violative content and platform.**

We do not track information in the manner requested. In order to estimate the views that violative content receives, we track prevalence, which tells us how often content that violates our standards is seen relative to the total amount of times any content is seen on Facebook. It is how we objectively evaluate our progress. We talk about and report on prevalence in our Community Standards Enforcement Report every quarter and describe it in our Transparency Center.

In the second quarter of 2022, the prevalence of hate speech was 0.02% (or 2 out of every 10,000 views) on Facebook and Instagram, the prevalence of violence and incitement was 0.03% on Facebook (or 3 out of every 10,000 views) and 0.01% to 0.02% on Instagram (or 1 to 2 out of every 10,000 views), and content violating our policies against terrorism had a prevalence rate of no more than 0.05% on Facebook and Instagram (or no more than 5 out of every 10,000 views).

Additionally, no company should grade its own homework, which is why we undertook an independent assessment of the metrics we report in our Community Standards Enforcement Report last year. EY conducted the assessment and concluded that the metrics were fairly stated, and our internal controls are suitably designed and operating effectively. To date, we are the only one of our peers who have undertaken this type of assessment.

Question 5. **Are trust and safety metrics used to determine compensation of all product development employees in their individual goals? If no, why not? If yes, please provide which metrics are used and how they affect compensation.**

We have dozens of teams and thousands of employees who work on developing our apps, and our goaling and incentives process varies significantly by organization, team, level, and role as a result. Broadly speaking, our goaling and incentives process is set up with the goal of building the best app, long term. The best app is also a safe app, so safety and security are an important part of our assessment of overall app health. We report on safety and security publicly in our quarterly Community Standards Enforcement Report, and we monitor safety and performance internally. Teams and individual employees are evaluated on a number of organization-specific factors, as well as the employee's expectations and goals set that year. During performance reviews, managers assess and calibrate the performance of the people on their team and assign them a performance rating, which drives compensation.

The entire process is holistic and does not turn on any single goal or metric. The employees who work on developing our apps have many goals related to their app's overall health, including

safety, security, engagement, and the prevalence of objectionable content. We believe that all of these goals are important to consider because they help us evaluate whether our users are having a positive experience with our apps. For instance, if changes made not only increased engagement but also problematic content, we would not consider that organization successful.

Question 6. **Are growth, engagement, or revenue metrics used to determine compensation of all product development employees in their individual goals? If yes, please provide which metrics are used and how specifically they affect compensation.**

Please see the response to your previous question.

Question 7. **How much has Meta spent in the last year on trust and safety?**

We invested about $5 billion last year on trust and safety issues, and we have over 40,000 people at the company working on these issues.

Question 8. **How much has Meta spent in the last year on product development?**

Meta spent $9.17 billion on research and development in the third quarter of 2022. For the first three quarters of 2022 combined, Meta spent $25.567 billion on research and development. Research and development expenses consist primarily of payroll and related expenses, which include share-based compensation; facilities-related costs for employees on our engineering and technical teams who are responsible for developing new products, as well as improving existing products; and professional services.

Question 9. **Please provide the number of employees who specifically research extreme content on your platforms and whether that research includes the extent to which your recommendation algorithms promote extreme content.**

We've tripled the size of our teams working on safety and security since 2016, and we now have over 40,000 people working on these issues. We have more than 350 dedicated specialists who work exclusively or primarily on preventing content that violates our policies against dangerous individuals and organizations from appearing on our family of apps and quickly identifying and removing it if it does. These professionals possess expertise ranging from law enforcement and national security to counterterrorism intelligence and academic studies in radicalization.

Our researchers work on a wide range of issues to help us understand activity on our apps and improve them, including research related to various aspects of our algorithm. We engage with governments and inter-governmental agencies around the world, and partner with several organizations with expertise in terrorism, violent extremism, cyber intelligence, and adversarial shifts in behavior online. We also routinely commission independent research from think-tanks, academics, and NGOs on various topics related to violent extremist and terrorist internet use in order to help our industry understand and make progress on these important issues.

Our internal research has informed important changes to our platforms, such as our decision in 2019 to ban white nationalism and white supremacism under our Dangerous Individuals and Organizations policy.

Question 10. **What restrictions, if any, does Meta place on the type of research employees can perform or the type of data they can access?**

Meta looks to build apps and services that improve the lives of the people that use our technologies. Research helps us understand not just how, but why people and businesses use our services the way they do, as well as provide new perspectives on how those services can become more useful or valuable. Researchers at Meta work on problem-solving that could include economic, demographic, or social issues, and they employ a range of methods to do so, including comprehension surveys, surveys on apps or existing behavior, focus groups, field studies, one-on-one interviews, prototype testing, and usability testing. Researchers also partner with independent academic organizations to conduct their own research either based on Meta's data or datasets we make available to researchers, and they publish and co-author academic papers with many researchers outside of Meta.

Researchers identify research questions (often in concert with their business partners), and then design and execute the most appropriate methodological approaches to address those questions. The work that our researchers do is often a response to feedback we hear from external stakeholders (e.g., civil society experts, academics) that in turn spurs a particular research project or collaboration. Whenever possible, our researchers seek to publish and share their findings with the broader academic community outside of Meta.

Our research teams are trained on robust ethics and compliance guidelines to ensure that they understand how to conduct research fairly, appropriately, and in support of Meta's mission. Research projects may be reviewed by or discussed with cross-functional teams to make sure that these guidelines are being followed and that the research is structured to help us improve our services or our users' experience, but as a general matter, we do not have policies restricting the topics that researchers may investigate to achieve that goal. Different teams of researchers focus on different areas of our services and company mission.

In regard to data access, for reasons including user privacy and security, Meta has safeguards in place to limit employee access to user data to situations in which an employee has a business need. Researchers therefore need to request access to certain types of data when their work requires it, but as a general matter, researchers are not prevented from accessing data that is appropriate and necessary for their research.

Question 11. **What is the takedown rate of (i) hate speech, and (ii) all violative content over the last year, month, and day?**

To the extent that "takedown rate" is defined here in the same manner as in Question 14 below, we do not track this information. We report the number of pieces of content we take down and the prevalence rate for a particular type of violating content, such as hate speech. We do not aggregate takedown and prevalence numbers across categories of violative content for several

reasons, one of which is that content may violate more than one of our policies. Different types of content are also associated with different types of potential harms, and we believe sharing this granular information is critical to understanding how these issues exist and are combatted on the platform.

We report this information quarterly in our Community Standards Enforcement Report. In the second quarter of 2022, we removed 13.5 million pieces of hate speech content from Facebook and 3.8 million pieces of hate speech content from Instagram. The prevalence of hate speech on Facebook and Instagram was 0.02%, or approximately 2 views per 10,000 views of content While we do not report an overall yearly prevalence rate, over the past four quarters for which we have reported data (the third quarter of 2021 through the second quarter of 2022), the prevalence of hate speech on Facebook was between 0.02% and 0.03%, meaning that overall views of hate speech content were between 2 and 3 views per 10,000 pieces of viewed content. Over the same period, the prevalence of hate speech on Instagram was between 0.01% and 0.03%, meaning that overall views of hate speech content were between 1 and 3 views per 10,000 pieces of viewed content. Over the same period, we removed a total of 68.3 million and 17 million pieces of content for violating our rules on hate speech from Facebook and Instagram, respectively.

Question 12. **What was the total number of QAnon content views and shares on Facebook and Instagram (i) before you began removing content in August 2020, and (ii) since you began to take down and down-rank content in August 2020?**

We did not track the number of QAnon Pages, Groups, Events, or profiles before the change in our policy, and we are unable to provide the information requested. From the time the policy changed in August 2020 to August 15, 2022, we removed about 4,200 Pages, 12,000 groups, 840 events, 67,200 Facebook profiles and 38,800 Instagram accounts for violating our policy against QAnon. We do not track views or shares for Pages, Groups, Events, or profiles.

Question 13. **How many of the views and shares of QAnon content between 2017 and 2020 were generated as a result of algorithmic recommendation?**

We do not track the information requested.

Question 14. **Frances Haugen released documents that contain internal estimates that the takedown rate, defined as the percent of views of hate speech on your platform that were taken down, was 3-5%. Was this number accurate at the time those estimates were made?**

The estimate referenced was a rough estimate by one data scientist of how much hate speech content Meta removed that incorrectly compared to prevalence (which represents a number of views, not of pieces of content), and was also an incomplete representation of Meta's actions against hate speech. It was not the product of or verified by rigorous analysis, and it did not address the amount of hate speech that is actioned and prevented from being viewed (for example, content that is downranked or demoted in Feed) through a number of means, as discussed here: https://about.fb.com/news/2021/10/hate-speech-prevalence-dropped-facebook/.

The rough numbers in the document misinterpreted the hate speech prevalence metric, which we publicly report every quarter in our Community Standards Enforcement Report, to represent the number of unique pieces of hate speech content on the platform, when the prevalence metric actually represents the estimated number of views of hate speech on the platform. Some pieces of content get many more views than others, so removing that content likely has a greater effect on how much hate speech is actually seen by users. Meta's enforcement teams typically prioritize review of content that is likely to accumulate a high number of views, and therefore the removals Meta makes may have an outsized positive effect on reducing user encounters of hate speech.

The figure also does not account for the actions that Meta takes in addition to simply removing content. This includes, but is not limited to, actions that effectively reduce the chance that hate speech is actually seen by a user. For example, content that is identified by Meta's algorithm as potential hate speech (but where the confidence level is not high enough to justify removal) may be demoted, making it less likely to appear in Feed, and thus less likely to be viewed by users. Nor does the figure account for the steps we take to prevent problematic content from being posted in the first place, such as removing accounts, Pages, and Groups; or the warning screens and checks we use when someone is posting something that may be hateful.

These are some of the reasons why Meta believes that the best measure of its hate speech enforcement efforts is to look at prevalence, which measures the amount of hate speech that users might actually see on the platform and reflects Meta's efforts to reduce hate speech distribution. We publicly report hate speech metrics, including prevalence, every quarter in our Community Standards Enforcement Report. Meta recently reported that prevalence for hate speech was about 0.02% (or about 2 views of violating content per 10,000 views) for the second quarter of 2022. This is a reduction from 0.11% (or about 11 views of violating content per 10,000 views) since we began reporting it for the third quarter of 2020, which is about an 80% decrease.

Question 15. **Why does Facebook automatically generate pages, including pages for hate groups that are banned from the platform, as reported by the Tech Transparency Project and others? Why did it take Facebook twelve years to take down the automatically generated Aryan Brotherhood page and why has Facebook not taken down this automatically generated page for the Proud Boys created on January 6, 2021?**

The more than 270 groups that Meta has designated as white supremacist organizations (as well as the many other groups we have designated as militarized social movements, terrorist organizations, or other dangerous organizations and individuals) are banned from our platform. We invest extensively in technology, people, and research to keep our platforms safe. We are also working to fix an auto generation issue, which incorrectly impacted a small number of pages. We removed the violating Pages flagged by the report, and we will continue to work with outside experts and organizations in an effort to stay ahead of violent, hateful, and terrorism-related content and remove such content from our platforms.

As we make clear on the platform, Unofficial Pages are created automatically. They are not affiliated with or endorsed or maintained by anyone associated with that group, and these groups and pages often have little to no interaction on Facebook.

Questions from Ranking Member Portman

Question 1. **The New York Post recently accused Facebook of "silencing" content and accounts that have discussed FBI whistleblower Steve Friend. Did Facebook remove content or accounts for speaking about this whistleblower or his accusations, yes or no?**

a. If yes, how did Facebook determine the content or accounts to violate Facebook's community guidelines?

No. With respect to the specific allegation you reference, the suggestion that we seek out peoples' private messages for anti-government language is plainly inaccurate. Accounts are permitted to post content and send messages about this individual and his allegations, consistent with our policies. For the avoidance of doubt, we have not removed content or accounts for speaking about this whistleblower or his accusations.

Question 2. **It has been reported that Meta's content moderators have been instructed to use the "Tanner scale" for identifying potential child sexual abuse material (CSAM). The Tanner scale has limited use for identifying CSAM, because it used to measure stages of puberty – not age. This is concerning, because it surely fails to identify certain cases of CSAM. Additionally, Meta's policy of "erring on the side of an adult" for moderating CSAM undoubtedly results in underreporting to the National Center for Missing and Exploited Children. Will Meta commit to halting its use of the Tanner scale and immediately cancel its policy of "erring on the side of an adult" for its CSAM moderation, yes or no?**

Content or behavior that exploits children is abhorrent and is prohibited on our services. We follow a three-pronged approach to combat child exploitation: (1) prevent this abhorrent harm in the first place; (2) detect, remove, and report exploitative activity that escapes these efforts; and (3) work with experts and authorities to keep children safe. We report all apparent instances of child exploitation appearing on our site from anywhere in the world to NCMEC, and in turn, NCMEC coordinates with law enforcement authorities from around the world.

Companies are governed by law as to what content they may report. If a provider does not have a good faith belief that the content depicted is that of an apparent minor, making a report (i.e., disclosure to NCMEC and law enforcement around the world) of that content may be in breach of the federal Stored Communications Act, which requires a provider to keep the content confidential unless an exception under the law applies. If there aren't clear indicators that the content depicts a minor, there is not a legal basis to make a report.

In the overwhelming majority of circumstances, determining whether a person is an adult or minor in violating content is clear. There is a small subset of reviews on content where the age of the person requires closer review. There is unfortunately no standardized aging system to use when deciphering the age of minors in photo or video media. This is why we continually refine our internal guidelines for deciphering the age of people in content on our platforms. The Tanner scale is only one indicator when assessing this content. We also rely on additional context that

might be present at the account/profile level and other physical indicators in an image or video. Other tech platforms, companies, law enforcement, and legal systems also use the Tanner scale, along with other biological indicators and contextual factors to determine the age of a person within an image or video.

Our training shows our clear intent to identify age accurately and report all content pertaining to an apparent minor. In our effort to constantly refine our analysis and work, we have developed new internal aging guidelines and also revised existing policies and training materials to ensure they align with expert recommendations and best practices. Based on academic sources, there have been efforts from governments, law enforcement, and researchers to develop such a system, but there are multiple limitations that need to be factored in, including aging differences across ethnicity, socioeconomic status, and environments. If a better resource that can be operationalized based on the four corners of an image or other limited context is developed and uniformly accepted by forensic pediatricians and other relevant stakeholders, we would welcome the opportunity to have access to it.

Across our family of apps, we take a comprehensive approach to child safety that includes zero-tolerance policies prohibiting child exploitation, cutting-edge technology to prevent, detect, remove, and report policy violations, and victim resources and support. We use industry-leading technology to find and remove violating CSAM content so it does not spread on our platform. At Meta, our work on child safety has spanned over a decade, and we will continue our industry-leading efforts to combat child exploitation.

Questions from Senator Carper

Question 1. **Tragically, over the last few years we have seen numerous deadly and disturbing hate crimes and domestic terrorism attacks. Many of the attacks were shared on social media while they occurred or after they occurred. One devastating example was the 2017 Christchurch mosque attack in New Zealand that was live-streamed on Facebook.**

And I recognize that one of the difficulties of the current threat landscape is its constant evolution. When it comes to hateful rhetoric that is often used by both foreign and domestic terrorists, bad actors are learning how to evolve or change rhetoric so that it is more difficult to detect.

a. As such, how do current safety features on your platforms evolve with the changes and tactics used by individuals or extremist organizations so that dangerous content is promptly caught and removed?

At Meta, we stand firmly against the exploitation of social media by those committed to inciting violence and hate. That's why, under the Facebook Community Standards, we prohibit hate speech, terrorism, and other harmful content. We employ tens of thousands of people and use industry-leading technology, including advanced artificial intelligence to enforce these rules—finding the majority of the content we remove before users even see it. We regularly publish transparency reports so people can see how we're doing over time and how we compare to other internet platforms. We are proud of our work in this space—but we are always working to improve and to share our lessons learned with other companies in the US and around the world.

Our Violence and Incitement policy prohibits content calling for or advocating violence, and we ban individuals and organizations that proclaim a violent mission under our Dangerous Individuals and Organizations policy. In August 2020, we expanded this policy further to address militarized social movements, such as certain militias, and violence-inducing conspiracy networks, like QAnon. We remove content that represents, praises, or supports those groups. We work tirelessly to enforce these policies. We have designated more than 1,000 militarized social movements and more than 270 white supremacist organizations, and we removed 2.3 million pieces of content from Facebook tied to organized hate globally in the second quarter of 2022, nearly 97% of which we found before someone reported it.

We have a team of cross-functional experts focused on these issues at Meta, including more than 350 highly trained professionals who work exclusively or primarily to prevent terrorist and violent extremist content from appearing on our platform, to quickly identify and remove it if it does, and to track it as it evolves. These professionals possess expertise ranging from law enforcement and national security experience to counterterrorism intelligence and academic studies in radicalization.

We invested around $5 billion on safety and security last year alone—more than any other tech company, even adjusted for scale. We are proud that we have over 40,000 people working on safety and security issues, and our efforts are making a difference. For example, hate speech now represents only about 0.02% of content views, or around 2 views per every 10,000. Our

advanced artificial intelligence systems have also improved at keeping people safe on our platform by proactively removing content that violates our standards. We found more than 95% of the hate speech we removed before anyone reported it to us—up from just 23% a few years ago.

Our efforts to combat extremism and hate do not end with our policies and enforcement efforts. We also have a number of programs to direct people to content or organizations to help them disengage from dangerous or violent movements. When people search for terms related to QAnon on Facebook and Instagram, we redirect them to credible resources from the Global Network on Extremism and Technology ("GNET"), the academic research network of the Global Internet Forum to Counter Terrorism ("GIFCT").

We are proud to have co-launched GIFCT five years ago to help fight terrorism and extremism online alongside our peers at other technology companies, including Google, YouTube, Microsoft, and Twitter. GIFCT was the culmination of years of informal partnerships among our companies on these issues, and the group has created a shared industry database for violent terrorist imagery that we have removed from our services. Sharing information allows all of us to identify more quickly and more accurately potential extremist content on our respective platforms. Most importantly, it also allows smaller companies the ability to take advantage of our technology and tactics, even with fewer people and resources. We believe that terrorism and extremism are shared problems that require shared solutions, and we encourage all tech companies to continue to partner with us in our efforts to keep such violence and hate off of online communities.

Abuse of our services is not static—and neither is the way we approach our integrity work. We are continuing to evolve how we approach integrity, and we are committed to continuing to improve so that we can help keep people safe.

We have made considerable progress here. Meta has one of the most comprehensive policies in the industry targeting terrorist organizations, hate groups, and large-scale criminal organizations, such as cartels, and violence inducing conspiracy theories. Our Dangerous Individuals and Organizations policy and definitions are part of our Community Standards and are public. Under this policy, we remove content that praises, substantively supports, or represents ideologies that promote hate, such as nazism and white supremacy. In the wake of the Christchurch Call to Action, we announced updates on our policies and approach to enforcement. Since then, we have invested heavily in both people and tools, allowing us to scale up our enforcement, global scope, and precision. We approach this space through a combination of AI and human intelligence—and also invest in research and work with outside experts and organizations to stay on top of this evolving space.

Working with Law Enforcement (for Mr. Cox)

Question 2. **Social media has fundamentally transformed the way we stay in touch with loved ones, create connections, and the way businesses large and small reach customers around the world. Unfortunately, it is also used to recruit, influence, and mobilize individuals to commit violent attacks.**

The rate at which information is shared using social media has significantly impacted the speed at which individuals may be radicalized and inspired to violence, which can narrow the window of opportunity for law enforcement to stop them before they commit violent attacks.

a. **What information do your platforms proactively share with law enforcement to prevent or flag the spread of violent or hateful content on social media before an act of violence is committed? What more needs to be done?**

Law enforcement plays a critical role in keeping people safe, and we have a long history of working successfully with law enforcement to address a wide variety of threats. If we identify serious violations that constitute a credible threat for real-world harm, we may contact law enforcement. We handle disclosures to law enforcement on a case-by-case basis to ensure consistency with our data privacy practices and legal requirements. As explained in our Privacy Policy, formerly known as the Data Policy, consistent with federal law, Meta may share user information with law enforcement when it has a good-faith belief it is necessary to detect, prevent and address fraud, unauthorized use of our tools, violations of our terms or policies, or other harmful or illegal activity; to protect ourselves (including our rights, property, or tools), people who use our apps, or others, including as part of investigations or regulatory inquiries; or to prevent death or imminent bodily harm.

Moreover, in responding to an emergency, a law enforcement official may submit a request through the Law Enforcement Online Request System, where any law enforcement agent or emergency responder authorized to gather evidence for an official investigation or investigate an emergency involving the danger of serious physical injury or death may request records from Meta. We disclose account records in accordance with our terms of service and applicable law. Our online records request system allows us to scale support to the over 18,000 law enforcement agencies and nearly 700,000 law enforcement officers in the U.S. alone.

For many years, we have had investigative and subject-matter expert teams working across the company to look for potential threats to public safety. These teams include former federal, state, and local law enforcement agents, and officers with decades of experience in safety and security fields. We also work with partners and experts in government, law enforcement, other companies (e.g., GIFCT), and civil society to help identify credible, real-world threats. We've invested more than $16 billion in safety and security since 2016, and we continue to prioritize it.

Questions from Senator Lankford

Question 1. **In Feb. 2022, I and Sens. Tillis, Cotton, Cruz, and Cramer sent a letter to Meta regarding the contradiction between Meta's Terms of Service and their allowance of information on Facebook about human smuggling. Meta responded to our letter by stating: "Meta's policy is to remove content that offers to provide or facilitate human smuggling. However, we do not remove content seeking information on how to cross borders."**

How does Meta distinguish between "content that offers to provide or facilitate human smuggling" and "content seeking information on how to cross borders"?

How do Meta's algorithms make this distinction? What keywords, images, hashes, or other content and indicators do Meta's algorithms recognize as "content that offers to provide or facilitate human smuggling"? In answering this question, please provide all keywords, images, hashes, or other content Meta's algorithm recognizes as "content that offers to provide or facilitate human smuggling."

What keywords, images, hashes, or other content and indicators do Meta's algorithms recognize as "content seeking information on how to cross borders"? In answering this question, please provide all keywords, images, hashes, or other content Meta's algorithm recognizes as "content seeking information on how to cross borders."

How do Meta's human content reviewers make this distinction? What keywords, images, hashes, or other content and indicators do Meta's human content reviewers recognize as "content that offers to provide or facilitate human smuggling"? In answering this question, please provide all keywords, images, hashes, or other content Meta's algorithm recognizes as "content that offers to provide or facilitate human smuggling."

What keywords, images, hashes, or other content and indicators do Meta's human content reviewers recognize as "content seeking information on how to cross borders"? In answering this question, please provide all keywords, images, hashes, or other content Meta's algorithm recognizes as "content seeking information on how to cross borders."

Are Meta's human content reviewers trained to identify content that offers to provide and facilitate human trafficking? Yes or no. If yes, please provide the committee with the training materials offered to human reviewers.

As a general matter, content is allowed on Meta's platforms unless the content violates a specific policy. Facebook's Community Standards govern what content may not be posted on the Facebook platform across the globe. In May 2019, after consulting with third-party experts, academics, and practitioners from around the world, Meta consolidated multiple related policies into a Human Exploitation ("HEx") Policy. The policy consolidation was consistent with the advice Meta received from these individuals and groups, who encouraged Meta to address a broad range of harmful and exploitative activities through one comprehensive human exploitation policy.

We prohibit content that facilitates or coordinates the exploitation of humans on Facebook and Instagram and remove it when detected. Meta defines human trafficking in our Community Standards as the "exploitation of humans in order to force them to engage in commercial sex, labor, or other activities against their will. It relies on deception, force and coercion, and degrades humans by depriving them of their freedom while economically or materially benefiting others." Meta adapted the definitions of human trafficking and human smuggling from the Protocol to Prevent, Suppress, and Punish Trafficking in Persons ("Palermo Protocol") and Protocol Against Smuggling of Migrants.

Meta uses a variety of tools to disrupt criminal organizations, including designation under our dangerous organizations policies, conducting human review, and employing a range of artificial intelligence ("AI") and network disruptions. Meta relies on people and technology to remove this content and works with NGOs and other stakeholders to combat ways our platforms may be used by those who want to harm people. We are constantly evaluating ways to improve our enforcement so we can most effectively find and remove content that breaks our rules.

We support our ability to detect violating content related to human exploitation through major investments by our technical and operational teams. We are always looking for ways to do more, which is exactly why we hire specialists in key fields to help us research and understand the problems so we can continue to improve our technology, staffing, and policies to address them. These teams of experts help us uncover patterns of harmful behavior so we can disrupt it. They have helped us to find and disrupt gangs and traffickers exploiting our platforms.

It is important to keep in mind that bad actors intent on breaking our rules will continue to update their tactics—and even the terminology they use—to avoid detection. We therefore do not provide specific details on our policy enforcement to avoid bad actors using that information as a playbook to game the system, and we work to stay on top of emerging trends in order to find and remove illicit content more quickly. And as a general matter, we continue to review and revisit our policy in line with feedback we receive from external stakeholders and our internal research.

For all policies and protocols, Meta provides comprehensive and rigorous training for content reviewers. This training begins at orientation, where content reviewers learn about the job and the resources available to them. In addition, ad reviewers must attend policy and procedure training, which consists of multiple weeks of instructor-led training, as well as hands-on practice, shadowing and training on all policies and protocols related to the work they do. Finally, reviewers are regularly trained after starting the position, and are tested with specific examples on how to uphold our policies and take correct action on a violation. This includes additional training when policies are clarified, or as they evolve.

Question 2. **In Feb. 2022, I and Sens. Tillis, Cotton, Cruz, and Cramer sent a letter to Meta regarding the contradiction between Meta's Terms of Service and their allowance of information on Facebook about human smuggling. Meta responded to our letter by stating: "Content that offers to provide or facilitate human smuggling is prohibited and removed when detected, for the reasons stated above."**

How much content "that offers to provide or facilitate human smuggling" has Meta removed from Facebook? In answering this question, please provide numbers for the past 10 years and break the numbers down on a monthly basis. Of the content described in your answer to the previous question, how long was the content available on Facebook prior to its removal? In answering this question please break down the numbers provided in response to the previous question according to the time on Facebook prior to removal. Additionally, please provide the analytics of each post, including the amount of shares and impressions each post had accumulated prior to removal.

Content that offers to provide or facilitate human smuggling is prohibited and removed when detected. Between July 2019 and December 2021, Meta removed over 240,000 pieces of content from Facebook for violating this policy. Approximately 96.9% of this content was identified before it was reported by any user.

Question 3. **As you may be aware, 8 U.S.C. 1325 makes it a criminal offense to enter the U.S. without authorization. Your terms of service state that "You may not use our products to do or share anything … That is unlawful." Do you believe that content that would facilitate an individual's violation of 8 U.S.C. 1325 would be a violation of your terms of service? If not, why not?**

Please see the response to your Question 1.

Question 4. **Your March 15 letter response to me noted: "With its policy, Meta intends to prohibit content relating to the business of human smuggling but not interfere with people's ability to exercise their right to seek asylum, which is recognized by international and U.S. law." As you may be aware, the Immigration and Nationality Act state that "any alien who is physically present in the United States or who arrives in the United States … may apply for asylum." (emphasis added) Do you believe individuals located outside of the United states have a right to seek asylum in the United States if they have not yet stepped foot on U.S. soil? If so, what is the basis for this right? Please share any analysis and provide a citation for any specific legal or administrative authority used in answering this question.**

Meta is not in a position to offer a legal opinion on the rights of specific individuals. We continually review and refine our policies' language in line with feedback from external experts and our internal research.

Question 5. **Has Meta assessed how WhatsApp is used as a tool to facilitate unlawful activity with respect to individuals crossing the U.S. border illegally?**

As a private encrypted messaging service, WhatsApp uses a combination of techniques to enforce its policies and prevent serious abuse. These include machine learning systems to tackle automated or bulk messaging accounts and unwanted contact, deliberate design to make it easy for people to report or block other users, and assisting law enforcement by complying with valid legal requests. In addition, WhatsApp makes it simple for users to report abusive or unwanted users, messages, media, groups, or communities. WhatsApp has a dedicated Trust and Safety

team with presence all over the world and in multiple languages, including Spanish, to support our global user base.

Question 6. **Has Meta assessed how Instagram and Facebook have been used to facilitate the unlawful activity of individuals crossing the U.S. border illegally? Please answer yes or no. If yes, please share this study with the Committee.**

We prohibit content that facilitates or coordinates the exploitation of humans on Facebook and Instagram and remove it when detected. Our global team of over 40,000 people working on safety and security includes native speakers covering over 50 languages working 24/7, educational resources, and partnerships with local experts and third-party fact checkers to help keep people safe. We also continually review and refine our policies' language in line with feedback from external experts and our internal research.

It is important to keep in mind that bad actors intent on breaking our rules will continue to update their tactics—and even the terminology they use—to avoid detection. We therefore do not provide specific details on our policy enforcement to avoid bad actors using that information as a playbook to game the system, and we work to stay on top of emerging trends in order to find and remove illicit content more quickly.

We also have policies in place to protect against repeat offenders and recidivist behavior. We impose increasingly severe account restrictions for people who repeatedly violate our policies, culminating in the account being permanently disabled. And if we identify credible threats of real-world harm, we may contact law enforcement and immediately disable the account. We handle disclosures to law enforcement on a case-by-case basis to ensure consistency with our data privacy practices and legal requirements.

We have an ongoing review of our larger enforcement protocols to see how we can continue to improve them.

Question 7. **In Feb. 2022, I and Sens. Tillis, Cotton, Cruz, and Cramer sent a letter to Meta regarding the contradiction between Meta's Terms of Service and their allowance of information on Facebook about human smuggling. Meta responded that it has developed its current policies around content related to border crossing with input from "civil society leaders and human rights leaders." Please provide a list of each civil society leader, human rights leader, academic, or other individual with whom Meta has consulted in developing these policies.**

We continually review and refine our policies' language in line with feedback from external experts and our internal research. Our definition of Human Exploitation adapts the definitions of human trafficking and human smuggling from the Protocol to Prevent, Suppress, and Punish Trafficking in Persons ("Palermo Protocol") and Protocol Against Smuggling of Migrants.

We do not share the names of all of the groups and individuals we consult with for a number of reasons—among them safety and security concerns, which are especially acute in places where the government may exercise censorship or control, and the fact that groups may not want to be

named. That said, we typically engage with civil society organizations, activist groups, and thought leaders in a variety of areas, including human rights. We also engage with academics who have relevant expertise.

We work with law enforcement around the world to help keep our community safe, both on and offline. This sometimes means providing information that will help them respond to emergencies, including those that involve the immediate risk of harm, such as those related to human exploitation. We also work with organizations around the world to provide resources and support for victims and survivors of human trafficking, and we direct people to expert organizations like Polaris and the National Human Trafficking Hotline. Our Help Center Page also provides our community with information on how to report human trafficking-related content on the platforms. In collaboration with our Safety partners, we share contact details of over 20 local, regional and global anti-trafficking organizations, including Polaris and the National Human Trafficking Hotline, to provide resources and assist victims and survivors of human trafficking. We work continually to update and expand the Help Center Page.

Question 8. **Meta usually requires an automated review of a Facebook ad and sponsored post prior to having a human review the ad or post as necessary. What keywords, images, hashes, or other content and indicators do Meta's algorithms recognize as "content that offers to provide or facilitate human smuggling" with respect to Facebook ads or sponsored posts? In answering this question, please provide all keywords, images, hashes, or other content Meta's algorithm recognizes as "content that offers to provide or facilitate human smuggling" with respect to Facebook ads or sponsored posts.**

What keywords, images, hashes, or other content and indicators do Meta's algorithms recognize as "content seeking information on how to cross borders" with respect to Facebook ads or sponsored posts? In answering this question, please provide all keywords, images, hashes, or other content Meta's algorithm recognizes as "content seeking information on how to cross borders" with respect to Facebook ads or sponsored posts.

How do Meta's human content reviewers make this distinction with respect to Facebook ads or sponsored posts? What keywords, images, hashes, or other content and indicators do Meta's human content reviewers recognize as "content that offers to provide or facilitate human smuggling"? In answering this question, please provide all keywords, images, hashes, or other content Meta's algorithm recognizes as "content that offers to provide or facilitate human smuggling" with respect to Facebook ads or sponsored posts.

What keywords, images, hashes, or other content and indicators do Meta's human content reviewers recognize as "content seeking information on how to cross borders" with respect to Facebook ads or sponsored posts? In answering this question, please provide all keywords, images, hashes, or other content Meta's algorithm recognizes as "content seeking information on how to cross borders" with respect to Facebook ads or sponsored posts.

Our Community Standards guide what is prohibited on Facebook and apply to all types of content, including ads. We also require advertisers to follow our Advertising Policies in addition to our Community Standards. Both of these policies are publicly available for review.

When it comes to enforcement, our ad review system is designed to review ads before they go live. This system relies primarily on automated technology to apply our Advertising Policies to the millions of ads that run across our apps. While our review is largely automated, we rely on our teams to build and train these systems, and in some cases, to manually review ads.

While ad review is typically completed within 24 hours, it may take longer, and ads can be reviewed again, including after they're live. Based on the results of the review, an ad is either rejected or allowed to run. If an ad is rejected, an advertiser can create a new ad—either with new ad creative or by revising the rejected ad—or request another review if they believe their ad was incorrectly rejected. If an ad was rejected for not complying with our Advertising Policies, depending on the violation, an advertiser may be able to correct the issue by updating the policy violating component like the image, video, text, targeting, or landing page.

Unlike the initial ad review, we rely more heavily on teams of human reviewers to process re-review requests from advertisers. Ads remain subject to review and re-review at all times.

Feedback from our community of advertisers, global experts, and people who use our platform everyday are an integral part of this effort and work. We will continue to work alongside them as we regularly evaluate our policies, improve our enforcement, and stay accountable to our progress over time.

Beyond reviewing individual ads, we may also review and investigate advertiser behavior, like the number of previous ad rejections and the severity of the type of violation, including attempts to get around our advertising review process. We want advertisers to understand our policies and how our enforcement system works. Account Quality provides a centralized place to monitor issues related to compliance with our Advertising Policies, Community Standards, and other Facebook policies and terms. Advertisers who violate our policies may have actions taken against them to restrict their advertising tools. These restrictions can include how much advertisers can spend, the advertising features they can use, or a loss of access to all advertising.

We know that both machines and people make mistakes, which is why the ad review system and enforcement aren't perfect. We are constantly working to improve our systems.

Question 9. **We have seen reports suggesting that individuals contemplating illegally crossing the U.S. border are now able to access reviews of border smuggling services on your platform.[1] Do you believe this content is within Meta's Terms of Service? Do you support this content being on your platform?**

How many Facebook groups in which an individual has shared a review of border smuggling services currently operate on Facebook? Has Facebook removed any groups

[1] https://www.npr.org/2021/10/14/1046140249/how-social-media-has-changed-migration-to-the-united-states

that allow for reviews of border smuggling services? Do its algorithms track this type of group or content within groups? Has Facebook referred any individuals who have participated in such a group to law enforcement? Does Facebook currently allow for law enforcement to view or observe content and activity in such groups?

We do not maintain statistics on these particular data points. However, Meta employs a robust series of tools on Facebook and Instagram to detect and remove content that violates its policies concerning human exploitation.

We prohibit content that facilitates or coordinates the exploitation of humans on Facebook and Instagram and remove it when detected. Meta defines human trafficking in our Community Standards as the "exploitation of humans in order to force them to engage in commercial sex, labor, or other activities against their will. It relies on deception, force and coercion, and degrades humans by depriving them of their freedom while economically or materially benefiting others." Meta adapted the definitions of human trafficking and human smuggling from the Protocol to Prevent, Suppress, and Punish Trafficking in Persons ("Palermo Protocol") and Protocol Against Smuggling of Migrants.

With respect to criminal organizations, as a threshold matter, we prohibit organizations or individuals that proclaim a violent mission or that are engaged in violence to have a presence on Facebook, including in Groups. We are also using AI to demote content that likely violates our policies.

If we identify serious violations that constitute a credible threat of real-world harm, we may contact law enforcement. We handle disclosures to law enforcement on a case-by-case basis to ensure consistency with our data privacy practices and legal requirements. Moreover, in responding to an emergency, a law enforcement official may submit a request through the Law Enforcement Online Request System, where any law enforcement agent or emergency responder authorized to gather evidence for an official investigation or investigate an emergency involving the danger of serious physical injury or death may request records from Meta. We disclose account records in accordance with our terms of service and applicable law. Our online records request system allows us to scale support to the over 18,000 law enforcement agencies and nearly 700,000 law enforcement officers in the U.S. alone.

Our Community Standards also apply to all Groups—public and private—and many of our proactive detection tools work across both. We may remove an entire Group if it repeatedly breaks our rules or if it was set up with the intent to violate our Community Standards. For people who repeatedly violate our Community Standards, we may take action to prevent them from creating new Groups. Our recidivism policy seeks to stop the administrators of a previously removed Group from creating another Group similar to the one removed. Similarly, an administrator or moderator who has had Groups taken down for policy violations cannot create any new Groups for a period of time. Posts from members who have violated any Community Standards in a Group must be approved by an administrator or moderator for 30 days following the violation. This stops their post from being seen by others until an administrator or moderator approves it. If administrators or moderators repeatedly approve posts that violate our Community Standards, we may remove the Group. We know there is more to do to keep Groups safe on

Facebook, and we'll keep improving our technology and policies to ensure Groups remain places where people can connect and find support.

Question 10. **We understand that ads and sponsored posts on Facebook are allowed to be geo-located and targeted based on a number of criteria. Please provide an exhaustive list of all criteria an individual can use to target an ad or sponsored post on Facebook.**

We strongly believe that the best advertising experiences are personalized. They enable people to discover products and services from small businesses that may not have the ability to market them on broadcast television or other forms of media. They also enable nonprofits, social causes, and organizations to reach the people most likely to support and benefit from them, such as connecting people to fundraisers for charitable causes they care about.

At the same time, we want to better match people's evolving expectations of how advertisers may reach them on our platform, and address feedback from civil rights experts, policymakers, and other stakeholders on the importance of preventing advertisers from abusing the targeting options we make available.

Meta's ads generally allow marketers to select audiences for their ads based on a variety of factors including age, gender, location, interests, and behaviors. These audience selection (or targeting) tools are not available for ads related to housing, employment, and credit.

Additionally, we give people ways to tell us that they would rather not see ads based on their interests or on their activities on other websites and apps, such as through controls within our ad settings. We also know that young people may not be well equipped to make these decisions on their own, which is why we take a more precautionary approach. In 2021, we began limiting advertisers' ability to target ads to people under 18 (or older in certain countries), allowing them only to target based on their age, gender, and location. This means that previously available targeting options for users under 18, like those based on interests or on their activity on other apps and websites, are no longer available to advertisers. These changes are global and apply to Instagram, Facebook, and Messenger. When young people become adults, we notify them about targeting options that advertisers can now use to reach them and the tools we provide to them to control their ad experience.

Meta is committed to being transparent about its ad delivery process; we publish information about it, including in our Business Help Center. For more information, please visit: https://www.facebook.com/business/news/good-questions-real-answers-how-does-facebook-use-machine-learning-to-deliver-ads.

Question 11. **Please provide us a copies of all Facebook ads and sponsored post Meta has approved and all ads and sponsored posts that have been denied since Jan. 20, 2021 regarding informational posts about how to enter a country illegally or how to be smuggled. In providing this information, please provide a list of all criteria that was used to target each ad or sponsored post.**

Please see the responses to your Questions 8, 9, and 10.

Question 12. **Does Facebook assess the funding source of any paid ads or sponsored posts regarding "content seeking information on how to cross borders"? For each paid ad or sponsored post regarding such content, did Facebook assess the funding source to ensure it is not profiting off of organized crime, drug cartels, or human smuggling organizations? If not, why not?**

We prohibit organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on Facebook, including through the use of paid ads or sponsored posts. This includes criminal organizations, including, but not limited to, those designated by the United States government as Specially Designated Narcotics Trafficking Kingpins ("SDNTKs"). We assess entities based on their behavior both online and offline—most significantly, their ties to violence. Under this policy, we designate individuals, organizations, and networks of people, and we prohibit these designated parties to have a presence on our platform, nor do we allow content that praises, substantively supports, or represents events that Facebook designates as terrorist attacks, hate events, mass murders or attempted mass murders, serial murders, hate crimes, or violating violent events.

Question 13. **For accounts whose paid ads or sponsored posts regarding content related to crossing the United States border was rejected, was that information turned over to law enforcement in the account holder's country of origin or in the United States? What is Meta and Facebook's process for dealing with the account information of ads which Meta or Facebook denies for violating its terms of service with regards to crossing the border. For each denied ad or sponsored post described in Question 11, please mark which ads had their account holders' information turned over to law enforcement in the United States or in the account holders' countries of origin. Please also indicate the specific (United States Federal, United States State, United States local, United States Tribal, and its foreign equivalents) law enforcement entity to whom such information was turned over.**

Law enforcement plays a critical role in keeping people safe, and we have a long history of working successfully with law enforcement to address a wide variety of threats. If we identify serious violations that constitute a credible threat of real-world harm, we may contact law enforcement. We handle disclosures to law enforcement on a case-by-case basis to ensure consistency with our data privacy practices and legal requirements. Moreover, in responding to an emergency, a law enforcement official may submit a request through the Law Enforcement Online Request System, where any law enforcement agent or emergency responder authorized to gather evidence for an official investigation or investigate an emergency involving the danger of serious physical injury or death may request records from Meta. We disclose account records in accordance with our terms of service and applicable law. Our online records request system allows us to scale support to the over 18,000 law enforcement agencies and nearly 700,000 law enforcement officers in the U.S. alone.

Question 14. **If prohibited content is posted on a Meta platform, what tools are available to Meta's human reviewers in regards to acting on the account which posted the content? In answering this question, please describe each tool that could be used by a human reviewer.**

Is a single human reviewer able to decide whether content only warrants removal? Is a single human reviewer able to decide whether content warrants suspension? Is a single human reviewer able to decide whether a piece of content warrants a ban from a Meta platform? In answering these questions, please describe in detail how decisions are made by a human reviewer or multiple human reviewers regarding removal, suspension, or a ban. In answering these questions, please also describe whether a single individual or multiple individuals are able to make each decision.

Decisions about whether to remove content are based on whether the content violates our policies. Our Community Standards are global, and all reviewers use the same guidelines when making decisions. We seek to write policies that clearly distinguish between violating and non-violating content and ensure that the content review process is as objective as possible.

Artificial intelligence ("AI") has improved to the point that it can detect violations across a wide variety of areas without relying on users to report content to Facebook, often with greater accuracy than reports from users. This helps us detect harmful content and prevent it from being seen by hundreds or thousands of people. AI has also helped scale the work of our content reviewers. Our AI systems automate decisions for certain areas where content is highly likely to be violating. This helps scale content decisions without sacrificing accuracy so that our reviewers can focus on decisions where more expertise is needed to understand the context and nuances of a particular situation. Instead of simply looking at reported content in chronological order, our AI prioritizes the most critical content to be reviewed, whether it was reported to us or detected by our proactive systems. Our teams focus on cases where it's essential to have people review, and we leverage technology to help us scale our efforts in areas where it can be most effective.

More than 15,000 reviewers across the globe review potential violations on Facebook and Instagram in more than 70 languages. They receive in-depth training and often specialize in certain policy areas and regions. Content is assessed against our policies, and reviewers use a wide range of resources in carrying out this work. For example, review teams undergo extensive training to ensure they have a strong grasp on our policies, the rationale behind our policies, and how to apply our policies accurately. Reviewers spend at least 80 hours in training with a live instructor. From there, they have hands-on practice using a facsimile of the review system, so they can apply what they've learned in a simulated environment. After this hands-on learning, reviewers get a report highlighting the areas where they apply our policies consistently and accurately and areas where they need more practice. To ensure they're up-to-speed on the latest information, reviewers receive regular coaching, refresher sessions, and additional trainings when policies are clarified or as they evolve. Our reviewers' work is also regularly audited to ensure quality enforcement.

Decisions about the consequences for a particular violation (for example, whether the user's account should be removed) are governed by our policies. For more information about our strike system and the penalties associated with violations, please refer to our Transparency Center: https://transparency.fb.com/enforcement/taking-action/restricting-accounts/.

We are committed to providing platforms where people can express themselves, and this is a guiding principle in our approach to content moderation. If someone believes that we have gotten

a content moderation decision wrong, that person can generally appeal our decision. Once a user asks us to take another look, we aim to review the content again, usually within 24 hours. If we find we've made a mistake, we'll let the user know, and their post, photo or video will be restored. If a user requests a review and still does not agree with our decision, the user may be able to appeal to the Oversight Board. Not all decisions are eligible for appeal to the Oversight Board and the Board only selects a certain number of eligible appeals, so they may not choose the user's appeal to review. Transparency in our appeals process is important, so our Community Standards Enforcement Report now includes the number of content appeals we receive, as well as how much content we restore upon appeal. Gathering and publishing those statistics keeps us accountable to the broader community and helps us to continue improving our content moderation. For more information, see https://transparency.facebook.com/community-standards-enforcement.

Question 15. **How long does it take Meta platforms to detect and take down child sexual abuse material, both previously detected material and new material?**

Our work on child safety has spanned more than a decade. We take a comprehensive approach to child safety that includes zero-tolerance policies prohibiting child exploitation; cutting-edge technology to prevent, detect, remove, and report policy violations; and victim resources and support. More than 40,000 people work on security and safety at Meta. We also collaborate with industry child safety experts and civil society around the world to fight the online exploitation of children because our commitment to safeguarding children extends beyond our apps to the broader internet.

We prohibit content that endangers children, such as content that contains nudity or physical abuse or content that sexually exploits children on Facebook and Instagram. When we find this type of violating content, we remove it, regardless of the context or the person's motivation for sharing it. We may also disable the account of the person who shared it, unless it appears the intent was not malicious (for example, to spread awareness of child exploitation).

In addition to using photo and video matching technologies to identify known child sexual abuse material ("CSAM"), we use artificial intelligence ("AI") and machine learning to proactively detect child nudity and previously unknown child exploitative content. We're using this and other technology to more quickly identify this content, hash it, and report it to the National Center for Missing and Exploited Children ("NCMEC"), in accordance with US law. We also use AI to find accounts that engage in potentially inappropriate interactions with children on Facebook and Instagram so that we can take action on the account and prevent additional harm.

Additionally, when a post on Facebook is reported, our specially trained teams determine whether to take action on the post. This team works 24 hours a day, 7 days a week around the globe. We are constantly innovating to develop ways we can do more to encourage reporting, make it more accessible to more people, and surface it at key moments that might signal abuse. After consultations with child safety experts and organizations, we've made it easier to report content for violating our child exploitation policies. To do this, we added the option to choose

"involves a child" under the "Nudity & Sexual Activity" category of reporting in more places on Facebook and Instagram. These reports are prioritized for review. We also started using Google's Content Safety API to help us better prioritize content that may contain child exploitation for our content reviewers to assess. We recognize that every upload of CSAM content victimizes a child, and any instance in which CSAM is present on our platform is abhorrent. We work aggressively to find and remove this content.

Question 16. **How many employees do you have who are dedicated to ensuring child sexual abuse material is appropriately removed and reported?**

For more than a decade, we have invested in teams and technology to combat child exploitation online. We have more than 40,000 people working around the world on safety and security across our platforms, and in 2021 alone we spent about $5 billion on safety and security. We have specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations to review potentially violating content and report findings to the NCMEC.

Child protection requires a global and comprehensive response from industry, law enforcement, government, civil society, and families, which is why we are committed to working with child-safety stakeholders to build and support the child-safety ecosystem. We also collaborate across the industry through organizations like the Technology Coalition, an industry association dedicated solely to eradicating the sexual exploitation of children online. In 2020, Meta joined Google, Microsoft, and 15 other member companies of the Technology Coalition to launch Project Protect, a plan to combat online child sexual abuse. This project includes a renewed commitment and investment from the Technology Coalition, expanding its scope and impact to protect kids online and guide its work for years to come. Project Protect focuses on five key areas: tech innovation, collective action, independent research, information and knowledge sharing, and transparency and accountability.

Additionally, we work closely with safety advisors, which include leading online safety nonprofits, as well as over 400 safety experts and NGOs from around the world, including specialists in combating child-sexual exploitation and aiding its victims. Our efforts include developing industry best practices, building and sharing technology to fight online child exploitation, and supporting victim services, among other things.

Question 17. **Would you describe Meta's efforts to take down and report child sexual abuse material as active or passive? In other words, is material passively scanned for known matches, or is there an active element to how you detect and take down this material?**

We prohibit content or activity that sexually exploits or endangers children, and we actively take down and report CSAM. When we become aware of apparent child exploitation, we report it to NCMEC, in compliance with applicable law.

We work with external experts, including the Facebook Safety Advisory Board, to discuss and improve our policies and enforcement around online safety issues, especially with regard to children. We also collaborate with industry child safety experts and civil society around the world to fight the online exploitation of children because our commitment to safeguarding children extends beyond our apps to the broader internet.

We also provide tools for users to report violating or potentially violating content. According to a recent study from Thorn, children who experience something negative online are more inclined to use in-app safety tools than to seek help offline. So we made it easier for minors to report by making our reporting tools easier to find, reducing the number of steps to report and pointing them toward reporting at key moments, such as after they block someone. In addition to user reports, we work closely with law enforcement to help them respond to emergencies. Our dedicated, in-house team provides the information that is most often used in criminal investigations, including basic subscriber information, IP addresses, email addresses, phone numbers, and traffic data.

While online platforms offer young people access to opportunities we never thought possible, we know those who seek to harm children seek out places where they can access children. Criminals will try to find ways to exploit online platforms. This is why we not only have strict policies against CSAM, but also policies against the sexualization of minors and other activities that can lead to child exploitation. It's also why we are constantly improving and developing new technologies to stay ahead of these criminals, and it's why we have made some of our technologies to fight abuse available at no cost to other companies.

In the past, our approach focused predominantly on proactive detection of CSAM on our platforms. This was the only tool we had in our toolbox—find the content and report it. As more companies have adopted this approach and we have improved our own efforts, the number of reports across industry has increased every year, but detection and reporting alone have not been successful at solving the problems of CSAM and grooming. Focusing solely on detecting harm after it's happened is not going to stop it from happening in the first place.

When it comes to prevention, we work to prevent harm from happening in the first place by disrupting an offender's ability to find other offenders or potential victims. We use a combination of strong default protections for young people, alongside technologies like machine learning to identify and address potentially malicious activity. Over the last few years, we've worked on several new features, tools, and technologies designed to help us do just that. For examples of the industry-leading tools and technologies we use to fight child sexual exploitation, please visit our Safety Center here: https://www.facebook.com/safety/onlinechildprotection/tools.

Question 18. **There has been an alarming uptick in sextortion (meaning a child is threatened or blackmailed with the potential release of sexual material in order to receive additional sexual content or money) and grooming on social media platforms. Between**

2019 and 2021, the number of reports involving sextortion more than doubled. What has your platform done to address this alarming trend?

We have zero tolerance policies regarding sexual exploitation. We prohibit content or activity that sexually exploits or endangers children. As stated previously, when we become aware of apparent child exploitation, we report it to NCMEC, in compliance with applicable law. We also work with external experts to discuss and improve our policies and enforcement around online safety issues, especially with regard to children.

We provide tools for users to report violating or potentially violating content, and we work closely with law enforcement to help them respond to emergencies.

Whether created in a relationship or in the context of exploitation, the non-consensual sharing of an intimate image is extremely traumatic for minors. This is particularly the case when someone tries to use those images to force a minor to share additional images, engage in sexual contact, or pay money to avoid those images being shared publicly. Both the Internet Watch Foundation and NCMEC have initiated important report-remove programs. We are proud to support NCMEC in their effort to build a global platform for minors who are worried their self-generated images will be shared on public online platforms without their consent. Powered by Meta funding, NCMEC hopes to see the platform launch in the near future. We will work closely with NCMEC, experts, academics, and victim advocates globally to develop the platform and ensure it responds to the needs of children in these horrific situations. In the initial version of the platform, minors will privately generate a hash of their images or videos directly on their own devices, without having to upload their content to the platform. NCMEC will securely house the hashes, at which point participating tech companies like ours can take those hashes and use them to remove and thwart the sharing of that content—helping to return power and control back to the victim.

Additionally, we have resources available for anyone seeking support and information related to sextortion. In 2019, together with Thorn and their partners, we launched an online "Stop Sextortion" hub on our Safety Center. With these resources, we hope to help destigmatize the issue so victims seek the help and support they need. We also want to educate others to help prevent further victims. "Stop Sextortion" includes tips for young people about what to do if they or one of their friends is a victim of this crime. It also includes tips for parents, caregivers, and educators about what they can do to help young people avoid situations that could lead to sextortion. It is available in more than 50 different languages. We also have a guide in the "Education Hub" of the Meta Family Center about talking to teens about not sharing intimate images. This conversation guide provides parents with tips on how they can openly talk to their teens about handling situations that deal with intimate images online.

Question 19. **On June 26, a board certified OBGYN posted a video on her Instagram account explaining that treating ectopic pregnancies or miscarriages is not the same as abortion. The purpose of abortion is to end an unborn child's life. The purpose of treating**

an ectopic pregnancy or treating a miscarriage is to save the life of a mom. Surprisingly, this post was flagged as "partly false information."

- **What information in the video does Meta consider to be false?**
- **Did Meta take time to flag posts from pro-abortion organizations that falsely claimed that the Supreme Court's decision in *Dobbs* would prohibit care for ectopic pregnancy and miscarriage? Or are only prolife organizations restricted from sharing information?**

Facebook and Instagram are platforms for users to share information and ideas from a variety of viewpoints and across the political spectrum. We strive to make Facebook and Instagram places where people can access reliable information and discuss important issues, like abortion, and we craft and enforce policies that support these goals.

A version of the referenced video was rated "Partly False" by an independent third-party fact-checker. In general, we do not think a private company like Meta should be deciding what's true or false, which is why we work with over 90 fact-checking organizations who independently review and rate potential misinformation on our platform. Fact-checkers review content (including, as relevant here, video content and any accompanying text), check facts, and rate the accuracy of content. These organizations are all certified through the non-partisan International Fact-Checking Network ("IFCN") and they independently decide what content to review and what rating to apply to that content.

Questions from Senator Ossoff

Question 1. **Does Meta or any Meta subsidiary have access to, possess, collect, or store any health care, health care related, or medical data related to its user or to any other persons in the United States?**

a. If so, what data?
b. How does Meta maintain and store that data?
c. For what purpose does Meta maintain and store that data?

Third-party developers can choose to use Meta's self-service Business Tools to provide certain data about the actions people take on their websites, apps, and in their stores for a variety of purposes that include ads measurement, targeting, and personalization. But Meta does not want developers to use Meta's Business Tools to send us sensitive data of any kind, and we contractually prohibit developers from using our Business Tools to send us sensitive data, specifically including personal health data.[2]

We take a number of steps to educate developers about these contractual obligations. For example, we provide publicly available guidance to developers about how to integrate the Business Tools in a compliant manner in numerous locations, including on Meta for Developers, Meta for Business, and Meta Help Center pages.[3] And, as a further preventative measure, if a developer does not heed its contractual obligations and our education, Meta's integrity systems are designed to filter out data that Meta categorizes as *potentially* sensitive health data from websites and apps categorized as health-related *before* it can be stored and used in our advertising systems. When we detect and filter out potentially sensitive health data, we send notifications to the developer that identify the events that included the potentially violating data we filtered out and instruct them to take any necessary action to fix their integrations, as required by our terms.

---

[2] *See* Facebook Business Tools Terms, https://www.facebook.com/legal/technology_terms; *see also* Meta Commercial Terms, https://www.facebook.com/legal/commercial_terms.

[3] *See, e.g.*, Meta Business Help Center, "About Restricted Meta Business Tools Data," https://www.facebook.com/business/help/1057016521436966; Meta Business Help Center, "Best Practices for Privacy and Data Use for Meta Business Tools," https://www.facebook.com/business/help/363303621411154; Meta Business Help Center, "About Sensitive Health Information," https://www.facebook.com/business/help/361948878201809; Meta Business Help Center, "About Personally Identifiable Information (Contact Information)," https://www.facebook.com/business/help/876665706470981; Meta Business Help Center, "About Sensitive Financial Information," https://www.facebook.com/business/help/2770378636585929; Meta Business Help Center, "Troubleshoot Facebook Business Tools Data Policy Violations," https://www.facebook.com/business/help/414614186228298; Meta Business Help Center, "How to Troubleshoot Meta Pixel Error and Warning Messages in Events Manager," https://www.facebook.com/business/help/490934944618455. Further, during the Business Tools account ID creation process, Meta instructs third-party developers not to send Meta sensitive user data, linking to both the Business Tools Terms and to Meta Business Help Center content about restricted data ("About Restricted Meta Business Tools Data").

Questions from Senator Padilla

Question 1. **Meta's users span the globe, and it makes the same commitments to Western, English-speaking communities as it does to non-Western, non-English language speaking communities. Last year, it was disclosed that at Facebook, 87% of all spending combating misinformation was spent on English language content. Only 9% of Facebook's users are English language speakers, and here in the United States and around the globe the native language of millions of people is a language other than English.**

a. What are the top five non-English languages for content that users within the United States encounter on each of your services (including Instagram and Facebook) and what percentage of users in the United States encounter content within each of those languages on a weekly basis?

The majority of content on Facebook and Instagram is language-agnostic, meaning that local language expertise is not required for content review. The 87% statistic is out of context. The majority of resources we have in place to prevent misinformation on our platform focus on content originating outside the United States. One of the primary ways we combat misinformation is through our third-party fact-checking program, which includes more than 90 global partners who review content in over 60 languages. The majority of those partners review content originating from outside the United States.

We also remove false and harmful content that violates our Community Standards, including more than 27 million pieces of false COVID-19 and vaccine content in more than a dozen languages. The majority of these content removals occurred on content originating from outside the U.S.

b. In your testimony before the committee you said that 40,000 people work on trust and safety issues. For each of these questions, please break down by employment status (full time in-house employees or contract-level).

i. Do those 40,000 people exclusively work on trust and safety issues, or are trust and safety issues only part of their responsibilities?

We have over 40,000 people working on safety and security, and we invested about $5 billion last year alone. This includes over 15,000 reviewers across the globe who review potential violations of our policies on Facebook and Instagram.

ii. How many people work exclusively on trust and safety issues?

Please see the response to your previous question.

iii. Of the people working exclusively on trust and safety issues, how many of them focus on non-English language content in the United States? Of the people working exclusively on trust and safety issues,

how many of them focus on non-English language content for users outside of the United States?

We have a global team of over 40,000 people working on safety and security. This includes more than 15,000 content reviewers who review content in more than 70 languages. Our reviewers receive in-depth training and often specialize in certain policy areas and regions.

When it comes to misinformation, we partner with over 90 fact-checking organizations around the world who rate content in more than 60 languages. In the US, we partner with 11 fact-checking organizations, seven of which cover content in Spanish: AFP, AP, PolitiFact, Reuters, FactCheck.org, TelevisaUnivision, and Lead Stories. We have 14 fact-checking partners in Latin America, covering Spanish and Portuguese; this includes 2 fact-checking partners we recently added in Brazil (now totaling 6 in Brazil). We've also partnered with Reuters, AFP, Telemundo, and TelevisaUnivision to launch fact-checking tip lines in Spanish on WhatsApp to help connect people with reliable information.

iv. How many of them focus on each of those top five non-English languages?

Please see the response to your previous question.

v. How many of them focus primarily on United States-based users, and how many of them focus primarily on non-United States based users?

Please see the response to your previous question.

c. Across each of Meta's services (including Instagram and Facebook) do your employees review all content for compliance with your community standards in their original language or are some or all subject to automated translation before being reviewed? And what percentage of total content reviewed is subject to automated translation prior to the company choosing to action, or not action that content?

We ensure that content reviewers are supported by teams with regional and linguistic expertise, including local or regional context that may be relevant to the content. In the last few years, we have tripled the number of people working on safety and security and now have over 40,000, of whom over 15,000 are content reviewers who review content in more than 70 languages.

In the first instance, content reviewers generally review content in the language in which it was posted. Similarly, the machine learning models that we use to proactively detect potentially violating content are trained on local languages, and not translations.

d. What is Meta doing to ensure that it has cultural competency for all the communities it serves, across each of its services?

Please see the response to your previous question.

e. Are the community standards that govern each of your services available in each of the languages your users speak?

Our Community Standards are available in approximately 68 languages, and we continue looking for ways to increase this number.

Question 2. **In your testimony you shared that Meta found and removed 95% of hate speech content before it was ever reported.**

a. Where did that 95% of hate speech content originate?

Overall, we focus on prevalence and calculate this metric by selecting a sample of content seen on Facebook and then labeling how much of it shouldn't be there. To measure prevalence we focus on how much content is seen, not how much sheer content violates our rules. In other words, we do not treat all content equally: a post seen 1 million times is 1 million times more likely to be sampled. In Q2 2022, hate speech prevalence was about 0.02%, or two views for every 10,000 views on Facebook and Instagram. Measuring views, rather than the amount of hate speech on our platforms better reflects the impact on the community.

Prevalence tells us how often content that violates our standards is seen relative to the total amount of times any content is seen on Facebook. It is how we objectively evaluate our progress and reflects our efforts to reduce hate speech distribution. Some pieces of content get many more views than others, so removing that content likely has a greater effect on how much hate speech is actually seen by users. Meta's enforcement teams typically prioritize review of content that is likely to accumulate a high number of views, and therefore the removals Meta makes may have an outsized positive effect on reducing user encounters of hate speech. We talk about and report on prevalence in our Community Standards Enforcement Report every quarter and describe it in our Transparency Center. We do not report prevalence by location in our Community Standards Enforcement Report, but per a recommendation from the Oversight Board, we are working on measuring enforcement data by location.

Of the content we remove, we flag over 95% of the violating Facebook content we action for hate speech before it is reported to us. We believe ours is the most comprehensive effort to combat hate speech of any major consumer tech company. As of Q2 2022, for every 10,000 views of content on Facebook and Instagram, about two views were on content that would violate our hate speech policies.

b. Of the remaining 5% of hate speech content, what was the median number of users who were recommended that content in their newsfeed before it was removed? What was the median number of United States based users who were recommended that content in their newsfeed before it was removed?

We do not track this information.

c. Of the remaining 5% of hate speech content, what was the median length of time a post determined to include hate speech content remained accessible before being removed? For posts that originated in the United States and were determined to include hate speech content, what was the median length of time it remained accessible prior to being removed?

Please see the response to your previous question. Moreover, given the nature of the internet, the amount of times content is seen is not evenly distributed. A small amount of content could go viral and get a lot of distribution in a very short span of time, whereas other content could be on the internet for a long time and not be seen by anyone.

For this reason, we consider prevalence to be a critical metric because it helps us measure how violations may impact people on Facebook. We care most about how often content that violates our standards is seen relative to the total amount of times any content is seen on Facebook.

We believe independent audits and assessments are crucial to hold us accountable and help us do better. That is why we commissioned Ernst & Young ("EY") to conduct an independent assessment focused on the metrics we report in our Community Standards Enforcement Report. EY found that the calculation of the metrics in our 2021 fourth quarter Community Standards Enforcement Report were fairly stated, and our internal controls are suitably designed and operating effectively. We look forward to expanding on our commitment to transparency and third-party accountability across our integrity systems and processes.

d. How does Meta assess its false positive rate—content you actioned for perceived violation of community standards, but that didn't actually violate the rules—and what is that rate on Facebook and Instagram? Please break this down by the top five languages on your services.

As a general matter, when we are made aware of incorrect content removal trends, we review them with our Community Operations team and content reviewers to prevent similar mistakes in the future.

If someone believes that we have gotten a content moderation decision wrong, that person can generally appeal our decision. In most cases, we also provide users with the reason for our decision and the specific policy that was violated.

In order to request re-review of a content decision we made, users are often given the option to "Request Review" or to provide feedback by stating they "Disagree with Decision." We try to make the opportunity to request this review or give this feedback clear, either via a notification or interstitial, but we are always working to improve.

Transparency in our appeals process is important, so we now include in our Community Standards Enforcement Report how much content people appealed and how much content was restored upon appeal. Gathering and publishing those statistics keeps us accountable to the broader community and enables us to continue improving our content moderation. For example, in Q2 2022, users appealed removal of 2.7 million pieces of content actioned for hate speech.

After re-review, 238,000 pieces of content, or less than 10% of content appealed, were restored. In addition, we restored 11,800 pieces of content removed for hate speech that we proactively re-reviewed.

We have also established an Oversight Board so people in the community can appeal our content decisions to a body that has independent judgment. The Oversight Board uses its independent judgment to decide some of our most significant and difficult cases, and the decisions it makes are binding.

e. Internally, what metric do you use to evaluate the success of moderation of hate speech content, and enforcement of your community standards generally?

Our metric is the same internally as it is externally—prevalence. Our goal is to minimize how often people see content that violates our policies. We gauge how we're doing primarily by measuring prevalence. Prevalence tells us what violating content people see because we missed it. It's how we most objectively evaluate our progress, as it provides the most complete picture. We talk about prevalence in our Community Standards Enforcement Report every quarter and describe it in our Transparency Center.

In Q2 2022, hate speech prevalence was about 0.02%, or two views for every 10,000 views on Facebook and Instagram. Measuring views, rather than the amount of hate speech on our platforms better reflects the impact on the community.

For additional information please see response to subpart c.

Question 3. **Election integrity is more important than ever. Platforms have a crucial role to play in ensuring that voters are able to access good, accurate election information, and in maintaining the basic integrity of our elections system. Meta has made announcements about what it is doing around the midterm elections, though much of your strategy for the midterms are similar to steps taken two years ago.**

a. Have you studied how effective your efforts were during the last election to ensure that your users knew where, how, and when to vote, and if so, what were the results of those studies?

Meta has dozens of teams working on elections, and we spent approximately $5 billion globally on safety and security last year alone. This work includes our advanced security operations, an industry-leading global fact-checking network, transparency around political ads, and new measures to help keep poll workers safe. With each major election around the world—including national elections this year in France and the Philippines—we incorporate the lessons we learn to help stay ahead of emerging threats.

While each election will bring its own unique set of challenges, we're working diligently to apply the lessons we've learned from previous years to forthcoming elections. Through our Facebook Open Research Transparency ("FORT") initiative, we provided researchers with data around the

US 2020 elections to enable them to examine the impact of Facebook and Instagram on elections and democracy.

We worked hard to do our part in protecting the integrity of the 2020 election, and we're proud of the work we've done and continue to do to support our democracy.

Ahead of the 2022 midterm elections, we're taking steps to combat election interference and misinformation, including covert influence operations conducted on behalf of foreign actors, while also working to help people vote. We're enforcing our policies against voter interference content and ads that call into question the legitimacy of the upcoming election. As we did in 2020, Meta is implementing a restriction period for new ads about social issues, elections, or politics in the US for the week leading up to General Election Day. We are putting this restriction period in place again because we found that the restriction period achieves the right balance of giving campaigns a voice while providing additional time for scrutiny of political ads in the Ad Library.

Our elections work is ongoing, and we continue to invest effort and resources to help protect elections, not just during campaign season, but at all times.

b. What do you assess to be the biggest threats to election integrity on your platforms in advance of the 2022 elections? What do you assess to be the biggest threats to election integrity on your platforms in advance of the 2024 elections? What steps are you taking to mitigate these identified threats? To the extent the nature of the threats differ by platform, please provide platform-specific answers to this question.

While each election will bring its own unique set of challenges, we are working diligently to apply the lessons we've learned from previous years to forthcoming elections. We are committed to continuing our work to help protect the right to vote.

Our approach to the 2022 US midterm elections is largely consistent with the policies and safeguards we had in place during the 2020 US presidential election. We work with a range of partners to ensure we're as prepared as possible, including state and local election officials, the federal Cybersecurity and Infrastructure Security Agency, and industry peers. We have hundreds of people across more than 40 teams working on the midterms, and we spent approximately $5 billion globally on safety and security last year alone. We're focused on preventing voter interference, connecting people with reliable information, and providing industry-leading transparency for ads about social issues, elections, and politics.

Through our Election Operations Center we monitor and respond to emerging election risks in real time. Our teams have been on high alert to identify emerging threats and respond as quickly as we can, including reducing the spread of misinformation, removing covert influence operations, and addressing foreign state-controlled media. We work hard to tackle all three of these challenges.

c. How are you staying ahead of the evolving threats to election integrity? To the extent the nature of the threats differ by platform, please provide platform-specific answers to this question.

Please see the response to your previous question.

d. What is Meta doing to keep election officials safe from doxxing and other online harassment campaigns?

We do not tolerate harassment on our platforms, and our Community Standards expressly prohibit doxxing, including the release of Personally Identifiable Information.

Furthermore, we know that certain people such as election officials can be targeted by bad actors on social media platforms, including Facebook and Instagram. Accounts that face additional threats during an election cycle may need additional protection, and that's why we created a program called Facebook Protect. Facebook Protect helps election officials, candidates, their campaigns, and elected officials adopt stronger account security protections, like two-factor authentication, and monitors for potential hacking threats.

In the lead-up to the 2020 election, we expanded our definition of high-risk people to include election officials, in order to help prevent any attempts to pressure or harm them, especially while they're fulfilling their critical obligations to oversee the vote counting. We're investing in proactive threat detection and expanding our policies to help address coordinated harassment and threats of violence against election officials and poll workers.

Additionally, law enforcement plays a critical role in keeping people safe, and we have a long history of working successfully with them to address a wide variety of threats. We dedicate significant resources to addressing law enforcement concerns, and we carefully review, validate, and respond to the legal requests we receive from them as soon as possible. This includes prioritizing requests related to emergency situations. And when we see a credible threat on our platform, we don't hesitate to reach out to law enforcement proactively.

Question 4. **In August, it was reported that when a user opens up a link to a website in Facebook or Instagram's respective iOS apps, the in-app browser that is used tracks user's behavior on those external websites. In a statement, Meta said that "[This code] allows us to aggregate user data for targeted advertising," but this tracking occurs both on external websites that partner with Meta and on those that do not. This doesn't appear to be standard practice and can be quite intrusive.**

a. How come nearly every other major iOS app is able to open up links without injecting tracking code but Facebook and Instagram cannot?

b. Is Meta considering opening up links in Instagram and Facebook's iOS apps in the user's default browser as WhatsApp does?

In-app web browsers are common across the industry. TikTok, Snapchat, LinkedIn, and Google Search are all social apps on iOS with in-app browsers. At Meta, we use in-app browsers to enable safe, convenient, and reliable experiences, such as making sure auto-fill populates properly or preventing people from being redirected to malicious sites. Adding any of these kinds of features requires additional code. People can also use the menu in our in-app browser to open links in their native browser.

The claims made about this issue by the researcher and in the press misrepresent how Meta's in-app browser and Pixel work. We add code for different purposes. One piece of code—the one that is the main focus of the researcher's article—was intentionally developed to honor people's App Tracking Transparency ("ATT") choices on our platforms. Our in-app browser is built to allow advertisers to work within Apple's ATT requirements. In order to do this, Meta has to rely on JavaScript code. The code in question allows us to respect people's privacy choices by helping aggregate events (such as making a purchase online) from Pixels already on websites, before those events are used for advertising or measurement purposes. Importantly, if an advertiser has not embedded Pixel, then the code remains dormant and it will not have any effect on the webpage, nor will it record data. And to be clear, the only party that can choose to embed a Meta Pixel on a website is the operator of the webpage (in practice, this would be the advertiser, not Meta).

Separately, we use another piece of code to suggest auto-fill options, like when you put in your information before making a purchase. Users who choose to have their auto-fill feature enabled can have pre-saved data fields auto-filled in the in-app browser—such as their postal address or phone number. In order for auto-fill to work, the code needs to understand when a user taps on a website element that can be auto-filled. It is misleading to say this code tracks user behavior within our in-app browser.

Questions from Senator Scott

Question. **How many accounts have you found advertising counterfeit pills on your platform?**

- **How many of those accounts were reported to you by a user of your platform?**

We prohibit the sale of illicit drugs on Facebook and Instagram, and have developed technology to find and remove this content proactively. Instagram's Community Guidelines and Facebook's Community Standards make it very clear that buying, selling, or trading non-medical or pharmaceutical drugs is prohibited. We're working hard to keep this content off Facebook and Instagram while surfacing communities of support that help those struggling with addiction. Any time we become aware of content on Facebook or Instagram that is facilitating activity like illicit drug sales, we remove it.

We enforce our policies through a combination of artificial intelligence, which we have developed to remove this content proactively, as well as reports from our community and human review. In the second quarter of 2022, over 98% of this content was removed from Facebook and nearly 97% of this content was removed from Instagram before users reported it. In some cases, content requires human review to understand the context in which it was posted (for example, to ensure it wasn't posted in the context of education or awareness raising). Today, we have 40,000 people working on safety and security and have invested more than $16 billion in teams and technology in this area since 2016. We have approximately 15,000 reviewers globally who work every day to review content in line with our policies and keep Facebook and Instagram safe places for all our users. These reviewers go through training to ensure they understand our policies and can enforce those policies accurately and consistently at scale. Our reviewers' work is audited to ensure quality enforcement.

In the second quarter of 2022 alone, we removed about 3.9 million pieces of content related to drug sales on Facebook and 1.9 million pieces of content related to drug sales on Instagram, and due to our improving detection technology, the prevalence of such content is about 0.05% of content viewed. In other words, out of every 10,000 views of content on Facebook and Instagram, we estimate no more than 5 of those views contained content that violated this policy. We publish this information and more in our quarterly Community Standards Enforcement Report. (https://transparency.fb.com/data/community-standards-enforcement/regulated-goods/facebook) We also routinely block hashtags when we find violations, such as #mdma, #buyfentanyl, and #buyxanax, among many others, and we continuously review additional hashtags to understand if there are further violations of our policies. We'll continue to improve in this area in our ongoing efforts to keep Facebook and Instagram safe, particularly for our youngest community members.

In addition to taking down violating content when we become aware of it, we use a strike system to count violations, educate users on policy, and hold users accountable for the content they post. For most violations, if a user continues to post content that goes against our Community Standards on Facebook (or our Community Guidelines on Instagram), despite warnings and

restrictions, we will disable the user's account. We also remove accounts that we assess are dedicated to the sale of restricted goods, such as illicit drugs.

Question. **How many of those accounts did you find on your own? Are you committed to working with appropriate local, state, and federal law enforcement to create a Trusted Reporter program to help bring down these accounts?**

Please see the response to your previous question. We are committed to working with law enforcement, and we deeply respect and support the work law enforcement agencies do to keep us safe. Meta has a long history of working successfully with many federal, state, and local government agencies to address a wide variety of threats. When law enforcement alerts us about illegal, drug-related activity on Instagram, we work to mitigate that threat and have developed tools designed to help law enforcement obtain quick feedback on their inquiries and gather evidence in connection with official investigations. User data privacy is one of our core values, and we handle disclosures to law enforcement on a case-by-case basis to ensure consistency with our data privacy practices and legal requirements.

As explained in our Privacy Policy, formerly known as the Data Policy, consistent with federal law, Meta may share user information with law enforcement when it has a good-faith belief it is necessary to detect, prevent, and address fraud, unauthorized use of our tools, violations of our terms or policies, or other harmful or illegal activity; to protect ourselves (including our rights, property, or tools), people who use our apps or others, including as part of investigations or regulatory inquiries; or to prevent death or imminent bodily harm.

Meta offers a robust online request system by which law enforcement can submit legal requests for user data as part of evidence gathering for investigative efforts. The Law Enforcement Online Request System is available at http://www.facebook.com/records and enables law enforcement officials to submit requests for information and track the progress of their requests. Meta has also created a guide, entitled the Law Enforcement Guidelines, to detail the process and legal requirements for obtaining user records. As detailed in the Law Enforcement Guidelines, Meta also allows law enforcement officers to submit emergency requests for information, which are addressed on an expedited basis, and requests for account preservations. These preservation requests can be continuously submitted so that data continues to be preserved while law enforcement officials continue their investigations.

Meta also conducts training for law enforcement agencies across the country to help law enforcement submit and track their requests. This includes training on how to leverage the online request system, as well as training on the legal requirements necessary for obtaining account information. Meta also provides law enforcement with a short guide to teach law enforcement how to submit and track their requests for data more efficiently.

Question. **Do you proactively refer accounts mentioning counterfeit pills or fentanyl to law enforcement?**

- **Do you notify the account holder that they are being investigated by law enforcement, potentially leading them to evade and continue selling their product and killing people?**

- **Which law enforcement entities are you referring accounts suspected of drug sales and/or trafficking to?**

When our teams identify a credible threat, we may reach out to law enforcement and disclose information necessary to help prevent real-world harm, in accordance with the law and our Terms of Service. We contact federal, state, or local law enforcement depending on the specific circumstances of a threat. We also have robust processes in place to handle government requests we receive, and we disclose account records in accordance with our terms of service and applicable law. We have law enforcement response teams available around the clock to respond to emergency requests.

With respect to notification, we notify users about requests for their information before disclosing it, unless exceptional circumstances exist, in emergencies, or where we are prohibited by law from doing so. We will also provide delayed notice upon expiration of a specific non-disclosure period in a court order and when we have a good faith belief that exceptional circumstances no longer exist and we are not otherwise prohibited by law from doing so.

For more information on our work with law enforcement, please see the response to your second question.

Question. **Over the last year, on average, how many subpoenas have you received per month related to accounts being used to sell or distribute drugs?**

Meta publishes data on the number of requests we receive from governments, including subpoenas. From July to December 2021, we received almost 11,000 subpoenas overall. For more information, please visit: https://transparency.fb.com/data/government-data-requests/.

Question. **What is the average turnaround time for responding to a subpoena issued by a law enforcement agency (Not including an automated or "received" style response)?**

We strive to respond promptly to law enforcement requests and within the requirements of the law. Moreover, in responding to an emergency, a law enforcement official may submit a request through the Law Enforcement Online Request System, where any law enforcement agent or emergency responder authorized to gather evidence for an official investigation or investigate an emergency involving the danger of serious physical injury or death may request records from Meta. We review requests we receive for legal sufficiency, and we may reject or require greater specificity on requests that appear overly broad or vague.

Question. **Do your moderators search for accounts or posts that use the DEA's Emoji Drug Code, or various permutations of the Emoji Drug Code?**

- **If no, why not?**

As a general matter, content containing emojis is not exempt from moderation. We know that people often use tactics like this in an attempt to evade our detection systems and bypass our policies. We regularly work with experts to identify terms, phrases, slang, and emojis that could

be used in this manner. That said, we do not share detailed descriptions of our work to avoid providing bad actors with the means to evade our detection system. For more information about our work to keep this content off of our platforms, please see the response to your first and second questions.

Question. **What has your company done to help spread awareness of the DEA's One Pill Can Kill campaign?**

We care deeply about opioid addiction and the fentanyl pill crisis in our communities, and we are committed to doing our part to address the crisis. We are aware of the DEA's One Pill Can Kill campaign and are very supportive of its goals. Where permitted by law and applicable ethics rules, we have provided support for such campaigns. We actively work to increase awareness about counterfeit prescription drugs and the specific risks to youth. For example, we have partnered closely with Song for Charlie, a leading non-profit working to raise awareness of the fentapill (i.e., fake pills made of fentanyl) crisis. We are supporting their public awareness campaigns that alert young people and parents about the dangers of fake pills and fentanyl. Song for Charlie recently reported that their campaign had reached 1.2 million people in the past 30 days. We are currently exploring other ways to bring this important message to life.

We have found other ways to support the work of the DEA to keep prescription drugs out of the hands of youth. In an effort to bolster the DEA's Prescription Drug Take Back Day that occurs twice a year, we have worked with our long-time partner the Center for Safe Internet Pharmacies ("CSIP") to create and promote an automated Facebook messenger experience on CSIP's Facebook page that provides people with quick and direct access to their closest disposal location, personalized FAQs, and treatment resources.

These two aforementioned efforts are an example of programs we run to address the opioid and addiction crisis. Beyond these programs, we allow people to talk about their recovery from substance misuse, as well as post content in the context of education and awareness raising (so long as this discussion does not also speak positively about, encourage the use of, coordinate, or provide instructions to make or use non-medical drugs). In the US, we have a long history of programs to try to address the addiction crisis by raising awareness of the stigma of addiction, while also connecting individuals with resources and help. For example, when people search for drugs or help on Facebook and Instagram, we direct them to the Substance Abuse and Mental Health Services Administration National Helpline ("SAMHSA") to help connect people with support and recovery services. Meta consults with experts and partners with federal, state, and local authorities, as well as non-profits, on innovative ways they can use social media as a tool to respond to the opioid epidemic. We have seen that Meta apps and tools can complement work on prevention, education, de-stigmatization, addiction support, and awareness, and we continue to support community groups and NGOs that have used our platform for good. From working with the Partnership to End Addiction to use Facebook Messenger as an effective tool to reach people during crisis, to the national Stop Opioid Silence PSA campaign that has reached over 140 million people in an effort to destigmatize addiction and connect people with recovery resources, we are committed to doing our part to implement solutions, recognizing that no company or government organization can address the dangers of fentanyl alone.

Questions from Senator Sinema

Cartel-Sponsored Content

Question. **In my capacity as Chair of the Border Management Subcommittee, I asked TikTok, Meta, Twitter, and YouTube to each commit to sharing cartel recruitment content with the Department of Homeland Security (DHS) as quickly as possible. You responded: "we would commit to that, provided privacy and legal concerns were addressed." I appreciate Meta's commitment to work with DHS to help protect Arizonan youth from being targeted by cartels.**

- **Please elaborate. Are there any circumstances where known and identified cartel recruitment content would not be shared with DHS due to privacy or legal concerns? If so, please describe in detail under what, if any, circumstances cartel-sponsored content on Facebook or Instagram that seeks to recruit American minors would not be shared with DHS.**

Law enforcement plays a critical role in keeping people safe, and we have a long history of working successfully with law enforcement to address a wide variety of threats. If we identify serious violations that constitute a credible threat of real-world harm, we may contact law enforcement. We handle disclosures to law enforcement on a case-by-case basis to ensure consistency with our data privacy practices and legal requirements. Moreover, in responding to an emergency, a law enforcement official may submit a request through the Law Enforcement Online Request System, where any law enforcement agent or emergency responder authorized to gather evidence for an official investigation or investigate an emergency involving the danger of serious physical injury or death may request records from Meta. We disclose account records in accordance with our terms of service and applicable law. Our online records request system allows us to scale support to the over 18,000 law enforcement agencies and nearly 700,000 law enforcement officers in the U.S. alone.

For many years, we have had investigative and subject-matter expert teams working across the company to look for potential threats to public safety. These teams include former federal, state, and local law enforcement agents and officers with decades of experience in safety and security fields. We also work with partners and experts in government, law enforcement, other companies (e.g., GIFCT), and civil society to help identify credible, real-world threats. We've invested more than $16 billion in safety and security since 2016, and we continue to prioritize it.

- **Will you commit to engage with the Border Patrol and the sheriffs' departments in counties along the Arizona-Mexico border regarding these concerns and provide each of them with a designated point-of-contact at Meta?**

Meta has a long history of working successfully with law enforcement and other government agencies around the world to address a wide variety of threats to our platform. We offer a robust online legal request system by which law enforcement can submit legal requests for user data as

part of evidence gathering for investigative efforts. The Law Enforcement Online Request System is available at www.facebook.com/records and enables law enforcement officials to submit legal process and track the progress of their requests. Meta has also created a guide, entitled the Law Enforcement Guidelines, to detail the process and legal requirements for obtaining user records. As detailed in the Law Enforcement Guidelines, Meta also allows law enforcement officers to submit emergency requests for information, which are addressed on an expedited basis, and requests for account preservations, which last 90 days pending receipt of formal legal process.

Moreover, Meta conducts training for law enforcement agencies across the country to help law enforcement submit and track their requests. This includes training on how to leverage the online request system, as well as training on the legal requirements necessary for obtaining account information. Meta also provides law enforcement with a short guide to teach law enforcement how to submit and track their requests for data more efficiently.

We are committed to continuing to engage with law enforcement at all levels to ensure that we are doing our part to support them in their efforts to keep people safe.

Deepfakes

Question. **Deepfakes – AI-generated videos that make people appear to do or say things they didn't actually do or say – present serious homeland security risks. In fact, in other countries such as Gabon and Malaysia, we've already seen fake videos of leaders trigger political turmoil. In Ukraine, there was a deepfake of President Zelensky supposedly telling Ukrainians to lay down their arms against Russia. While platforms have increasingly adopted policies to ban many deepfakes, these policies only can be enforced if you know what is or is not real.**

- **What is Meta doing to improve deepfake detection? And in cases where deepfakes are too advanced to be detected, how do you go about determining a video's authenticity?**

Under our policy regarding misleading manipulated videos, we remove content that meets the following criteria:

1. It has been edited or synthesized—beyond adjustments for clarity or quality—in ways that aren't apparent to an average person and would likely mislead someone into thinking that a subject of the video said words that they did not actually say; and
2. It is the product of artificial intelligence or machine learning that merges, replaces, or superimposes content onto a video, making it appear to be authentic.

This policy does not extend to content that is parody or satire, or video that has been edited solely to omit or change the order of words. Consistent with our existing policies, audio, photos, or videos—whether a deepfake or not—will be removed from Facebook if they violate any of

our other Community Standards, including those governing nudity, graphic violence, voter suppression, and hate speech. And videos that don't meet these standards for removal are still eligible for review by one of our independent third-party fact-checkers, which include over 90 partners worldwide fact-checking in over 60 languages. If content is rated false or partly false by a fact-checker, we significantly reduce its distribution in News Feed and reject it if it's being run as an ad. And critically, people who see it, try to share it, or have already shared it will see warnings alerting them that it's been rated false by a fact-checker.

Our enforcement approach has several components, from investigating AI-generated content and deceptive behaviors, like fake accounts; to partnering with academia, government, and industry; to exposing people behind these efforts. We also continue to invest in partnerships, including with journalists, academics, and independent fact-checkers, to help us reduce the distribution of false news and misinformation, as well as to better inform people about the content they encounter online.

XR Product Development

***Question.* XR technology holds great potential, but it also raises important privacy questions – both in the public and private spheres.**

- **What steps is FRL taking to protect the privacy interests of non-participants of Project Aria in public areas, particularly with respect to being recorded without knowledge?**
- **What steps is FRL taking to protect the privacy interests of Project participants, particularly with respect to the 3D mapping of homes and other private places?**

We believe in the power of technology to collapse physical distance and help people connect with those who matter most. We want everyone to feel like they're in control of their VR experience and to feel safe, and that's why we're consulting with experts (privacy, safety, and civil liberties) and gathering feedback to develop policies, tools, and resources.

Privacy is an integral part of everything we do—we embed privacy across the Company and invest in addressing it at scale. That's why we are starting the conversation around building a responsible metaverse early—bringing outside experts and partners into the debate from the start.

In developing Project Aria, our research project designed to help us build the first generation of wearable AR devices, we've taken steps to protect people's privacy. Our privacy and data protection controls were developed with the input of privacy experts, including Eli Dourado, Senior Research Fellow at the Center for Growth and Opportunity, and Daniel Castro, Director of the Center for Data Innovation at the Information Technology and Innovation Foundation. All Project Aria devices display a prominent white light that indicates when data is being collected. Under our policies, recording is not permitted in sensitive areas like restrooms, prayer rooms, locker rooms, or in sensitive meetings and other private situations—and it is only allowed in the homes of wearers with consent from all members of the household. All devices have a physical

mute button that will stop collecting data when pressed. Project Aria participants are not permitted to record in venues such as stores or restaurants without written consent from such venues. Furthermore, once uploaded, captured data is kept under quarantine (not available to researchers) for three days. During that time, for any data set gathered in a public place, the system automatically blurs faces and license plates.

Responses to Questions for the Record
U.S. Senate Homeland Security and Governmental Affairs Committee
"Social Media's Impact on Homeland Security"
September 14, 2022

Neal Mohan
Chief Product Officer, YouTube

Written Questions Submitted by Chairman Peters to Neal Mohan

1. Please provide the following information regarding YouTube's current employees as of September 1, 2022: (i) the total number of full-time engineers at your company, (ii) the total number of full-time engineers working full time on ensuring trust and safety or integrity of your platform, and (iii) the total number of full-time engineers working full time on product development.

Google invests billions of dollars a year on research and development to come up with next generation technologies, including breakthroughs in quantum computing and innovations in artificial intelligence. We are constantly working to improve our products to meet the needs of our consumers and have made numerous improvements to our products to provide all Google consumers with world-class privacy and security protections, consistent with the privacy commitments we make to our users and our legal obligations. YouTube is a global platform with over 2 billion users in more than 100 countries speaking 80 different languages — and our engineering workforce is a critical component across all our efforts.

We have reviewed your question carefully, and the information you requested is not the type of information that Google customarily releases to the public. The law, as reflected in the Trade Secrets Act (18 U.S.C. § 1905) and Freedom of Information Act (5 U.S.C. § 552(b)(4)) recognizes the critical and sensitive nature of confidential business, proprietary, and trade secret information, and – as such – protects against the disclosure of such information. Public disclosure of this type would likely cause substantial competitive harm to the company. For instance, in the case at hand, disclosure through publication would arm Google's competitors with information that could be used to undermine the company in business negotiations and future transactions – as well as potentially compromising Google's credibility and relationships with users, partners, and advertisers.

As the information requested is confidential, business, and trade secret information and not publicly disclosed by Google in the normal course of business, the Company will continue coordinating appropriately with Committee counsel regarding the availability and confidential

treatment of such commercial confidential information under applicable Senate and Committee rules.

2. Please provide all metrics evaluated in your A/B testing for each of the last 2 years, including a description of each metric and the percentage of A/B tests in which the metric was evaluated.

We test our products extensively before they are launched. Deep thought about how to meet the needs of our users and how to uphold our responsibility to users goes into all the products, tools, and features we bring to creators and viewers on YouTube. We gather feedback and consider a range of options to determine the best path to execute at scale. Critically, before making it to the final product launch, we take a special approach to experiments that makes each of these advances possible.

We launch experiments throughout the year, lasting for as long as it takes to gather statistically significant data to meaningfully inform our product decisions. We start by developing a set of clear hypotheses we want to test. Then all the teams that work together to launch new features collaborate on designing the experiments or research studies that can give insight into these hypotheses. While A/B tests can be an effective way to test products to better understand how they may work at scale, they are not the only – or even the best – way to understand how our products impact users. In addition to quantitative testing, we also rely on qualitative assessments like user experience research sessions.

One critical area where we rely on both quantitative and qualitative information to inform our decision making is around fulfilling our commitments to user trust and safety. A key part of our approach involves anticipating problems before they emerge and combating them before they pose a larger threat. We have teams in place that continually monitor the news, social media, and user reports from around the world; detect new trends; and work with relevant teams to investigate and address them. An important metric we use to measure the effectiveness of our work is violative view rate (VVR), explained in greater detail in our response to Question 3. Additionally, we have a robust process where feature teams consult with our Trust & Safety teams prior to launching new products. Our Trust & Safety teams conduct a risk assessment of those products, make recommendations to reduce risk, and work to put appropriate safety measures in place before launching.

3. Does YouTube use metrics to measure whether certain demographics see a disproportionate amount of harmful content on your platform, such as, for example, the number of users who have repeated exposures to harmful content, or the amount of harmful content seen by the user in the 99th percentile of exposure to harmful content? If yes, please share a list of these metrics and their current values and values from last year.

YouTube strives to remove content that violates our Community Guidelines before users are exposed to this content. To measure our progress on removing violative videos, we have developed a metric called Violative View Rate (VVR). This metric estimates the percentage of views on violative videos.

VVR data gives critical context around how we protect our community. Although metrics like the turnaround time to remove a violative video are important, those statistics do not fully capture the actual impact of violative content on the viewer. We believe the VVR is the best way for us to understand how harmful content impacts viewers, and to identify where we need to make improvements. We are committed to continuing to be transparent about this metric and to reduce it over time.

We calculate VVR by taking a sample of videos on YouTube and sending it to our content reviewers who tell us which videos violate our policies and which do not. By sampling, we gain a more comprehensive view of the violative content we might not be catching with our systems. However, the VVR will fluctuate— both up and down. For example, immediately after we update a policy, you might see this number temporarily go up as our systems ramp up to catch content that is newly classified as violative. Our methodology for this report has been validated by MIT Sloan professor of statistics Dr. Arnie Barnett as "thoroughly sensible and statistically sound" (Dr. Barnett's analysis is available at https://www.analysisgroup.com/globalassets/insights/publishing/2021-youtube-vvr-methodology-a-statistical-assessment-arnold-barnett.pdf).

In response to your questions regarding metrics:

- In the second quarter of 2022, we removed over 4.4 million videos for violating the Community Guidelines. Over 4.1 million of the videos removed were first detected by our automated flagging system. Of the content automatically detected and removed, more than 72% received 10 or fewer views before they were removed. Over this time period, VVR was in the range of .09% to .11% (i.e., out of every 10,000 views on YouTube, only 9-11 come from violative content).
- In the first quarter of 2022, we removed over 3.8 million videos for violating the Community Guidelines. Over 3.5 million of the videos removed were first detected by our automated flagging system. Of the content automatically detected and removed, approximately 67% received 10 or fewer views before they were removed. Over this time period, VVR was in the range of .09% to .11%
- In the fourth quarter of 2021, we removed over 3.7 million videos for violating the Community Guidelines. Over 3.4 million of the videos removed were first detected by our automated flagging system. Of the content automatically detected and removed,

approximately 70% received 10 or fewer views before they were removed. Over this time period, VVR was in the range of .12% to .14%

- In the third quarter of 2021, we removed over 6.2 million videos for violating the Community Guidelines. Over 5.9 million of the videos removed were first detected by our automated flagging system. Of the content automatically detected and removed, approximately 72% received 10 or fewer views before they were removed. Over this time period, VVR was in the range of .09% to .11%.

4. What is the total number of views that violative content has received across your platform over the last year, month, and day? Please break down by category of violative content and platform.

As explained in response to Question 3, YouTube developed the VVR metric to measure our progress in removing violative videos. The VVR for the second quarter of 2022 was in the range of .09% to .11%. For the first quarter of 2022, the VVR was also in the range of .09% to .11%. For the fourth quarter of 2021, the VVR was in the range of .12% to .14% and for the third quarter of 2021, the VVR was in the range of .09% to .11%.

The policy violations for video removal during the second quarter of 2022 is as follows:

- Child safety – 1,383,028 videos (30.8%);
- Violent or graphic – 900,014 videos (20.0%);
- Nudity or sexual – 666,315 videos (14.8%);
- Harmful or dangerous – 533,896 videos (11.9%);
- Harassment and cyberbullying – 499,719 videos (11.1%);
- Spam, misleading and scams – 150,833 videos (3.4%);
- Hateful or abusive – 145,688 videos (3.2%);
- Misinformation – 122,660 videos (2.7%);
- Promotion of violence and violent extremism – 72,990 videos (1.6%); and
- Other – 21,790 videos (0.5%).

The policy violations for video removal during the first quarter of 2022 is as follows:

- Child safety – 968,178 videos (24.9%);
- Violent or graphic – 824,899 videos (21.2%);
- Nudity or sexual – 654,415 videos (16.9%);
- Harmful or dangerous – 481,327 videos (12.4%);
- Harassment and cyberbullying – 422,320 videos (10.9%);
- Spam, misleading and scams – 367,787 videos (9.5%);
- Hateful or abusive – 95,947 videos (2.5%);
- Promotion of violence and violent extremism – 60,721 videos (1.6%); and
- Other – 7,090 videos (0.2%).

The policy violations for video removal during the fourth quarter of 2021 is as follows:

- Child safety – 1,182,403 videos (31.5%);
- Violent or graphic – 748,943 videos (19.9%);
- Nudity or sexual – 692,147 videos (18.4%);
- Spam, misleading and scams – 339,763 videos (9.1%);
- Harassment and cyberbullying – 322,627 videos (8.6%);
- Harmful or dangerous – 304,187 videos (8.1%);
- Hateful or abusive – 88,678 videos (2.4%);
- Promotion of violence and violent extremism – 71,789 videos (1.9%); and
- Other – 3,678 videos (0.1%).

The policy violations for video removal during the third quarter of 2021 is as follows:

- Child safety – 1,986,073 videos (31.9%);
- Violent or graphic – 1,453,805 videos (23.3%);
- Nudity or sexual – 1,146,956 videos (18.4%);
- Spam, misleading and scams – 607,621 videos (9.8%);
- Harassment and cyberbullying – 283,955 videos (4.6%);
- Harmful or dangerous – 281,568 videos (4.5%);
- Promotion of violence and violent extremism – 250,962 videos (4.0%);
- Hateful or abusive – 114,854 videos (1.8%); and
- Other – 104,088 videos (1.7%).

5. Are trust and safety metrics used to determine compensation of all product development employees in their individual goals? If no, why not? If yes, please provide which metrics are used and how they affect compensation.

Our first priority is to protect our users, and as described in more detail below, this priority informs how we measure employee performance both at the individual and collective level. From the way we staff our teams to our advertising policies and the choices we make on what content to monetize, trust and safety considerations underpin how we operate our business.

Content moderation efforts at Google and YouTube are primarily managed by Trust & Safety teams across the company. These teams are made up of engineers, content reviewers, and others who work across Google to address content that violates any of our policies. They also work with our legal and public policy teams, oversee the people we hire to help us scale our content moderation efforts, and provide the language expertise and the 24-hour coverage required of a global platform. Google employs review teams across many offices globally and across the U.S. to ensure we have a diverse set of reviewers who are reviewing publisher sites, apps, and content.

We do not have to choose between safety and growing our business. In fact, our business depends on our providing a safe environment for users and advertisers. How YouTube lives up to its publicly declared goals is a core part of the product, and much of what we do on the product team is to deliver on Trust and Safety objectives and live up to YouTube's responsibilities.

As we consider the YouTube product experience, we must take into account how our content policies impact our viewers and our creators. The Community Guidelines act as the "rules of the road" for what is and is not allowed, and we apply them consistently across YouTube. This means the policies and how they work are a core part of the overall product experience for our entire ecosystem — including our viewers, creators, and advertisers. Our responsibility relies on using our systems to connect viewers to high-quality information and to minimize the chances they'll see problematic content. It also involves assessing how creators–the vast majority of whom are looking to do the right thing – are rewarded. These levers work together to form a comprehensive approach to responsibility.

Our advertiser-friendly content guidelines and the YouTube Partner Program detail YouTube's revenue sharing program, and if a creator fails to comply with those guidelines, we remove them from the program and prevent them from making money on YouTube. When we do that, YouTube foregoes revenue as well.

Our business model therefore relies on raising up and rewarding quality content while reducing or removing potentially harmful content, and all Google employees are responsible for prioritizing the safety of our platform. In 2020, the most recent year for which we have figures, more than 20,000 people across the globe helped enforce Google's policies and moderate content, and we spent nearly $1.2 billion on content moderation across the company. Considerations regarding the trust and safety of our users is an integral part of the YouTube product. A large part of our work on the product team is devoted to delivering on these important objectives and living up to YouTube's responsibilities.

If an employee builds a product that does not take into account trust and safety, we simply would not launch that product for our users. While failing to consider user safety would factor into an employee's performance review, we have numerous safeguards in place to ensure that all products accrue to the benefit of our users from the time they are launched.

In terms of specific metrics, YouTube's focus on responsibility is clearly reflected in our Objectives and Key Results, referred to as "OKRs." OKRs are a method for setting and communicating goals throughout an organization, and ensuring the organization is working towards milestones in support of those goals. The Objectives are the goals that the organization is working to accomplish, and the Key Results are specific, measurable results that adjust over time to help the organization stay on track to accomplish those objectives.

For years, YouTube's OKRs both at the individual and company level have included the objective to provide "the most trusted experience possible" for our users. To reach that objective, the key results track our four "Rs" of responsibility: Remove content that violates our policies as quickly as possible, Raise up authoritative voices, Reduce the spread of content that brushes up against our policy (i.e., borderline content), and Reward trusted creators. We also focus on improving the quality of experience for users and creators, delivering a reliable service.

6. Are growth, engagement, or revenue metrics used to determine compensation of all product development employees in their individual goals? If yes, please provide which metrics are used and how specifically they affect compensation.

In addition to providing a safe and trusted experience for our users, we expect Google employees to deliver the most helpful products for all users, to build and improve products and features at scale, internally and externally, and to grow a scaleable, financially sustainable businesses. These expectations, too, are reflected in our individual and company Objectives and Key Results.

As noted above, in response to Question 5, all of our objectives work together, and we simply cannot build our business or deliver the best product possible without keeping our platform safe and maintaining the trust of our users. We have both a responsibility and a business incentive to continue to encourage our employees to build the safest, most helpful product possible for our users.

7. How much has YouTube spent in the last year on trust and safety?

In 2020, the most recent year for which we have figures, more than 20,000 people across the globe help enforce Google's policies and moderate content, and we spent nearly $1.2 billion on content moderation across the company. Trust and Safety is an integral part of the YouTube product. A large part of our work on the product team is devoted to delivering on Trust and Safety objectives and living up to YouTube's responsibilities.

8. How much has YouTube spent in the last year on product development?

We publicly report our financial results for YouTube as part of the Google Services segment of our business. Google Services includes products and services such as ads, Android, Chrome, hardware, Google Maps, Google Play, Search, and YouTube. Our quarterly and yearly consolidated operating income and expense numbers for all our business units — including YouTube as part of Google Services — are detailed in our public filings, available at https://abc.xyz/investor/.

These filings include information on various aspects of our business, including the amount we spend each year and each quarter on research and development expenses. As noted in our Form 10-K filing (available at https://abc.xyz/investor/static/pdf/20220202_alphabet_10K.pdf?cache=fc81690) in 2021, we spent over $31.5 billion dollars on research and development.

9. Please provide the number of employees who specifically research extreme content on your platform and whether that research includes the extent to which your recommendation algorithms promote extreme content.

At YouTube, we take a comprehensive approach to combating online extremism, abuse, and other harmful content. We continue to significantly increase our investments in the systems and processes that allow us to remove violative content, raise up authoritative content, reward trusted creators and artists, and reduce the spread of borderline content. With respect to extremist content, we have a network of robust policies in place that work together to combat it. Our system is designed to elevate authoritative voices and surface relevant information panels for additional context. On topics such as news, politics, medical, and scientific information, we take the additional step of raising up authoritative sources. Through a combination of human reviewers and machine learning, we reduce borderline content —or content that comes close to violating, but does not quite violate, our rules.

A growing number of independent researchers have been looking into how tech platforms impact the consumption of borderline content, and, while ongoing study continues, certain recently published papers conclude YouTube recommendations do not steer viewers towards extreme content. Instead, consumption of news and political content on YouTube more generally reflects personal preferences that can be seen across their online habits.

- Academics from Harvard and the University of Pennsylvania have concluded that YouTube's algorithm does not cause echo chambers. Instead, they have concluded that radical content on YouTube appears to reflect broader patterns of news (this study is available at https://arxiv.org/pdf/2011.12843.pdf).
- Researchers at UC Berkeley have found that YouTube's recommendation algorithm actively discourages viewers from visiting radicalizing or extremist content by favoring mainstream media and cable news content over independent YouTube channels (this study is available at https://firstmonday.org/ojs/index.php/fm/article/view/10419). The study found: "...data suggest that YouTube's recommendation algorithm actively discourages viewers from visiting radicalizing or extremist content. [The] study thus suggests that YouTube's recommendation algorithm fails to promote inflammatory or radicalized content, as previously claimed by several outlets." A 2022 ADL report is also consistent with these findings (this study is available at

https://www.adl.org/resources/reports/exposure-to-alternative-extremist-content-on-youtube).

- The Pew Research Center has shown that YouTube's recommendation system points users to popular content, no matter which criterion was used to select the starting video (this study is available at https://www.pewresearch.org/internet/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/pi_2018-11-07_youtube_0-10-2/).
- Another set of researchers at Berkeley confirmed that YouTube's policies have led to a reduction in conspiratorial recommendations (this study is available at https://arxiv.org/abs/2003.03318).
- Another research article published earlier this year on the consumption of radical content on YouTube found "no evidence that engagement with far-right content is caused by YouTube recommendations systematically" (this study is available at https://www.pnas.org/content/118/32/e2101967118).
- An April 2022 study by A. Chen, B. Nyhan, et. all., found that "non-subscribers are rarely recommended videos from alternative and extremist channels" (this study is available at https://arxiv.org/abs/2204.10921).

To encourage further research, we recently expanded the YouTube Research Program, which gives academic researchers access to our raw data in a user privacy sensitive way, allowing them to derive metrics or develop insights of their own based on that data (more information about this program is available at https://research.youtube/).

10. What restrictions, if any, does YouTube place on the type of research employees can perform or the type of data they can access?

Google has numerous policies and controls in place to ensure that user data is managed appropriately. Google's data access controls dictate that access is limited to only the data necessary to perform a role, and that access is reviewed by managers and the privacy team before being granted.

We tightly restrict employee access through a number of industry-leading safeguards, including: limiting access to user data to necessary individuals; requiring a justification to access such data, multi-stage review before access is granted to sensitive data; and monitoring for access anomalies and violations. We have a strong build and test infrastructure and access to the production environment is limited.

Access grants also expire and must be renewed and re-reviewed regularly. Specific data may only be accessed for a properly justified purpose for a specific period of time and is subject to review by our privacy team.

11. What is the takedown rate of (i) hate speech, and (ii) all violative content over the last year, month, and day?

As discussed in response to Question 3, at YouTube, we publish a quarterly report on YouTube Community Guidelines Enforcement, which includes data on removals, appeals and reinstatements, and detection by human and machine flagging. As discussed above, in 2021, we started releasing the Violative View Rate (VVR). That metric, which we report on a quarterly basis, indicates the percentage of views on YouTube that comes from content that violates our policies. In order to calculate VVR, we take a sample of the views on YouTube and send the sampled videos for review. Once we receive the decisions from reviewers about which videos in the sample are violative, we aggregate these decisions in order to arrive at our estimate. In the first half of 2022, the VVR was .09%-.11%. In other words, for every 10,000 views on YouTube, only 9-11 went to content that proved to be violative. Our methodology for this report has been validated by MIT Sloan professor of statistics Dr. Arnie Barnett as "thoroughly sensible and statistically sound."

In the second quarter of 2022, we removed over 4.4 million videos for violating the Community Guidelines. The breakdown of the reasons for video removal during this period is as follows:

- Child safety – 1,383,028 videos (30.8%);
- Violent or graphic – 900,014 videos (20.0%);
- Nudity or sexual – 666,315 videos (14.8%);
- Harmful or dangerous – 533,896 videos (11.9%);
- Harassment and cyberbullying – 499,719 videos (11.1%);
- Spam, misleading and scams – 150,833 videos (3.4%);
- Hateful or abusive – 145,688 videos (3.2%);
- Misinformation – 122,660 videos (2.7%);
- Promotion of violence and violent extremism – 72,990 videos (1.6%); and
- Other – 21,790 videos (0.5%).

More than 93% of videos removed were first detected by our automated flagging system. Of the content automatically detected and removed, more than 72% received 10 or fewer views before they were removed. Over this time period, VVR was in the range of .09% to .11% (i.e., out of every 10K views on YouTube, only 9-11 come from violative content). We are committed to continuing to be transparent about this metric and to reducing it over time.

The breakdown of the reasons for video removal during the first quarter of 2022 is as follows:

- Child safety – 968,178 videos (24.9%);
- Violent or graphic – 824,899 videos (21.2%);
- Nudity or sexual – 654,415 videos (16.9%);
- Harmful or dangerous – 481,327 videos (12.4%);
- Harassment and cyberbullying – 422,320 videos (10.9%);

- Spam, misleading and scams – 367,787 videos (9.5%);
- Hateful or abusive – 95,947 videos (2.5%);
- Promotion of violence and violent extremism – 60,721 videos (1.6%); and
- Other – 7,090 videos (0.2%).

The breakdown of the reasons for video removal during the fourth quarter of 2021 is as follows:

- Child safety – 1,182,403 videos (31.5%);
- Violent or graphic – 748,943 videos (19.9%);
- Nudity or sexual – 692,147 videos (18.4%);
- Spam, misleading and scams – 339,763 videos (9.1%);
- Harassment and cyberbullying – 322,627 videos (8.6%);
- Harmful or dangerous – 304,187 videos (8.1%);
- Hateful or abusive – 88,678 videos (2.4%);
- Promotion of violence and violent extremism – 71,789 videos (1.9%); and
- Other – 3,678 videos (0.1%).

The breakdown of the reasons for video removal during the third quarter of 2021 is as follows:

- Child safety – 1,986,073 videos (31.9%);
- Violent or graphic – 1,453,805 videos (23.3%);
- Nudity or sexual – 1,146,956 videos (18.4%);
- Spam, misleading and scams – 607,621 videos (9.8%);
- Harassment and cyberbullying – 283,955 videos (4.6%);
- Harmful or dangerous – 281,568 videos (4.5%);
- Promotion of violence and violent extremism – 250,962 videos (4.0%);
- Hateful or abusive – 114,854 videos (1.8%); and
- Other – 104,088 videos (1.7%).

12. How many hours of QAnon content were viewed on YouTube (i) before you began to down-rank QAnon in 2019, and (jj) since you began to down-rank QAnon in 2019?

The rise of misinformation in recent years led us to further expand the ways we use our recommendation system to address problematic misinformation and borderline content. This includes conspiracy theory videos ("the moon landing was faked") or other content that spreads misinformation ("orange juice can cure cancer").

We are able to do this by using classifiers to identify whether a video is "authoritative" or "borderline". To determine authoritativeness, we look at a few key questions: Does the content deliver on its promise or achieve its goal? What kind of expertise is needed to achieve the video goal? What is the reputation of the speaker in the video and the channel it is on? What is the main topic of the video (e.g., News, Sports, History, Science, etc)? Is the content primarily meant to be satire? These answers and more determine how authoritative a video is. The

higher the score, the more the video is promoted when it comes to news and information content.

To determine borderline content, we assess factors that include — but aren't limited to — whether the content is: inaccurate, misleading or deceptive; insensitive or intolerant; or harmful or with the potential to cause harm. The results are combined to give a score for how likely it is that the video contains harmful misinformation or is borderline. Any video classified as borderline is reduced in recommendations. These evaluations then train our system to model their decisions, and we now scale their assessments to all videos across YouTube.

To create a safe environment for our users at YouTube, we approach our content recommendations with our four "Rs" of responsibility: we remove content that violates our policies as quickly as possible, we raise up authoritative voices, we reduce the spread of content that brushes up against our policy (i.e., borderline content), and we reward trusted creators.

- **We remove content that violates our policies**. In Q2 2022 alone, we removed more than 4.4 million videos for violating our Community Guidelines, more than 93% of which were first flagged by our automated systems. More than 72% of those first flagged by our systems received 10 or fewer views. Most recently, we began disclosing our violative view rate (VVR), or the number of views on content that violates our Community Guidelines as a portion of total viewership. More information on our VVR is available at https://transparencyreport.google.com/youtube-policy/views. In Q2 2022 our VVR was 0.09-0.11%, meaning that out of every 10,000 views on YouTube, 9-11 came from violative content.
- **We raise up authoritative voices when people are looking for news and information**. For example, we raise content from authoritative channels for newsworthy events in YouTube search results and "Up Next" recommendations panels. In addition to current events, authoritativeness is also important when it comes to topics prone to misinformation, such as vaccines. In these cases, we aim to surface videos from experts, like public health institutions, in search results. We also provide informational panels to provide contextual information on specific content in both the search results page and the video watch page. In October 2020, we updated YouTube's hate and harassment policy to prohibit content targeting individuals or protected groups based on conspiracy theories used to justify real world violence (e.g., QAnon and Pizzagate).
- **We reduce recommendations of borderline content.** As referenced above, in January 2019, we announced improvements to our recommendation system to greatly reduce recommendations of borderline content and content that could misinform users in harmful ways. This resulted in a 70 percent drop in views coming from our search and discovery systems. Further, we saw the number of views that come from

non-subscribed recommendations to prominent QAnon-related channels dropped by over 80 percent from January 2019 to October 2020.

- **We reward trusted creators by allowing those channels to make money on our site.** We have strict policies on the kind of videos we allow ads to appear on that are strictly enforced, and creators must meet specific eligibility requirements to join our monetization program.

We have been removing QAnon-related content from the YouTube platform for years, given the content's potential for incitement to violence and pursuant to our policies prohibiting hate, harassment, violent extremism, and graphic violence. After updating our hate speech policy in June 2019, we removed and terminated tens of thousands of QAnon-related videos and channels for violating our policies, including videos and channels belonging to prominent QAnon figures.

In October 2020, we updated our harassment policies to explicitly prohibit content that targets an individual or group with conspiracy theories — including QAnon — that have been used to justify real-world violence. Due to the evolving nature and shifting tactics of groups promoting violative content, we continuously review and adapt our policies to stay ahead of bad actors. We work continuously to improve our defenses against these threats, including by investing in technological mechanisms that can protect our platform at scale.

13. How many hours of QAnon content were viewed on YouTube between 2017 and 2019 as a result of algorithmic recommendation?

At YouTube, we take a comprehensive approach to combating harmful content, including QAnon-related content. Over the last few years, we have significantly increased our investments in the systems and processes that allow us to remove violative content, raise up authoritative content, reward trusted creators and artists, and reduce the spread of borderline content. Our teams work hard every day to ensure that we are allowing for a broad range of political speech while adhering to our responsibility of making sure that our platform is not abused to incite real-world harm or spread harmful misinformation.

Managing misinformation and harmful conspiracy theories is challenging because the content is always evolving, but we take this issue very seriously. Due to the shifting tactics of groups promoting conspiracy theories, we continuously invest in the policies, resources, and products needed to protect our users from harmful content. We have clear and public policies that we apply consistently.

In January 2019, we announced improvements to our recommendation system to greatly reduce recommendations of borderline content and content that could misinform users in harmful ways. This resulted in a 70 percent drop in views coming from our search and

discovery systems. Further, we saw the number of views that come from non-subscribed recommendations to prominent QAnon-related channels dropped by over 80 percent from January 2019 to October 2020.

In October 2020, we expanded our hate and harassment policies to prohibit content that targets an individual or group with conspiracy theories that have been used to justify real-world violence (e.g., QAnon and Pizzagate). We also provide informational panels to provide contextual information on specific content in both the search results page and the video watch page. For example, since 2018, we have seen 25 million impressions on our QAnon info panel. We continue to adapt our policies to stay current and remain committed to taking the steps needed to live up to this responsibility.

Written Question Submitted by Sen. Carper to Neal Mohan

Social media has fundamentally transformed the way we stay in touch with loved ones, create connections, and the way businesses large and small reach customers around the world. Unfortunately, it is also used to recruit, influence, and mobilize individuals to commit violent attacks.

The rate at which information is shared using social media has significantly impacted the speed at which individuals may be radicalized and inspired to violence, which can narrow the window of opportunity for law enforcement to stop them before they commit violent attacks.

a. **What information does your platform proactively share with law enforcement to prevent or flag the spread of violent or hateful content on social media before an act of violence is committed? What more needs to be done?**

With regard to government requests for user information, a variety of laws allow government agencies around the world to request the disclosure of user information for civil, administrative, criminal, and national security purposes. Each request is carefully reviewed to make sure it satisfies applicable laws. For more information, see our policies for how Google handles government requests for user information (available at https://policies.google.com/terms/information-requests).

We also publicly share information on government requests to remove content to shed light on the scale and scope of government requests for content removals. Data on government removal requests can be located on the Google Transparency Report (available at https://transparencyreport.google.com/government-removals/overview?hl=en).

There are also a number of ways we proactively work with law enforcement agencies to assess threats and to counter attempts to deceive, harm, or take advantage of people using our platforms. Our interactions with law enforcement include the following:

- First, we maintain regular communication channels with law enforcement, as communication with law enforcement, industry partners, and the federal government is a key component of our efforts to keep users safe. We rely on information from industry peers and law enforcement to ensure the integrity of our platform and act swiftly in response to crises or when we detect abuse that may threaten public safety or the integrity of democratic processes.
- Second, we have dedicated teams of analysts and security experts who work around the clock to identify and investigate possible instances of coordinated influence operations on our platform. Google's Threat Analysis Group works to identify bad actors and terminate their channels and accounts, works with other technology

companies to share intelligence and best practices, and shares threat information with law enforcement. For more information regarding our Threat Analysis Group, please see https://blog.google/threat-analysis-group/.

- Third, we have processes in place to proactively refer to law enforcement imminent threats and certain illegal activity occurring on our platform. These imminent threats are raised to us in a variety of ways. In addition to our ongoing threat analysis just noted, we also have processes in place for our various internal product teams, as well as external sources, to escalate potential threats and criminal activity if they see it on our platforms. Google has a team - the CyberCrime Investigation Group (CCIG) - in place to assess escalated threats, and that team refers matters to law enforcement whenever appropriate.

Additionally, we developed the YouTube Trusted Flagger program to help provide more sophisticated reporting processes for government agencies and non-governmental organizations (NGOs) that are particularly effective at notifying YouTube of content that violates our Community Guidelines. The program provides these partners with dedicated reporting processes and a channel for ongoing discussion and feedback about YouTube's approach to various content areas. The program is part of a network of around 300 government partners and NGOs that bring valuable expertise to our enforcement systems. Participants in the Trusted Flagger program receive training in enforcing YouTube's Community Guidelines, and because their flags have a higher action rate than the average user, we prioritize them for review. Once flagged, our YouTube content reviewers make the call on what content is or is not removed by applying our YouTube policies. We apply our policies transparently and consistently, without taking into account either the political leanings of the author, or their place in society.

Written Questions Submitted by Sen. Sinema to Neal Mohan

Cartel-Sponsored Content

In my capacity as Chair of the Border Management Subcommittee, I asked TikTok, Meta, Twitter, and YouTube to each commit to sharing cartel recruitment content with the Department of Homeland Security (DHS) as quickly as possible. You responded: "we would cooperate, as long as there's a due legal process with the DHS and other law enforcement as well." I appreciate YouTube's commitment to work with DHS to help protect Arizonan youth from being targeted by cartels.

- **Please elaborate on what you mean by a "due legal process" and describe under what, if any, circumstances known and identified cartel-sponsored content on YouTube that seeks to recruit American minors would not be shared with DHS.**
- **Will you commit to engage with the Border Patrol and the sheriffs' departments in counties along the Arizona-Mexico border regarding these concerns and provide each of them with a designated point-of-contact at YouTube?**

As a preliminary matter, our Community Guidelines prohibit terrorist groups or cartels from using YouTube. Specifically, we prohibit content created by, praising, or recruiting for these groups. We have a dedicated team that responds to law enforcement around the clock, every day of the year, including on issues related to violent organizations.

With regard to government requests for user information, a variety of laws allow government agencies around the world to request the disclosure of user information for civil, administrative, criminal, and national security purposes. Each request is carefully reviewed to make sure it satisfies applicable laws. For more information, see our policies for how Google handles government requests for user information (available at https://policies.google.com/terms/information-requests).

We also publicly share information on government requests to remove content to shed light on the scale and scope of government requests for content removals. Data on government removal requests can be located on the Google Transparency Report (available at https://transparencyreport.google.com/government-removals/overview?hl=en).

There are also a number of ways we proactively work with law enforcement agencies to assess threats and to counter attempts to deceive, harm, or take advantage of people using our platforms. Our interactions with law enforcement include the following:

- First, we maintain regular communication channels with law enforcement, as communication with law enforcement, industry partners, and the federal government is a key component of our efforts to keep users safe. We rely on information from

industry peers and law enforcement to ensure the integrity of our platform and act swiftly in response to crises or when we detect abuse that may threaten public safety or the integrity of democratic processes.

- Second, we have dedicated teams of analysts and security experts who work around the clock to identify and investigate possible instances of coordinated influence operations on our platform. Google's Threat Analysis Group works to identify bad actors and terminate their channels and accounts, works with other technology companies to share intelligence and best practices, and shares threat information with law enforcement. For more information regarding our Threat Analysis Group, please see https://blog.google/threat-analysis-group/.

- Third, we have processes in place to proactively refer to law enforcement imminent threats and certain illegal activity occurring on our platform. These imminent threats are raised to us in a variety of ways. In addition to our ongoing threat analysis just noted, we also have processes in place for our various internal product teams, as well as external sources, to escalate potential threats and criminal activity if they see it on our platforms. Google has a team - the CyberCrime Investigation Group (CCIG) - in place to assess escalated threats, and that team refers matters to law enforcement whenever appropriate.

Additionally, we developed the YouTube Trusted Flagger program to help provide more sophisticated reporting processes for government agencies and non-governmental organizations (NGOs) that are particularly effective at notifying YouTube of content that violates our Community Guidelines. The program provides these partners with dedicated reporting processes and a channel for ongoing discussion and feedback about YouTube's approach to various content areas. The program is part of a network of around 300 government partners and NGOs that bring valuable expertise to our enforcement systems. Participants in the Trusted Flagger program receive training in enforcing YouTube's Community Guidelines, and because their flags have a higher action rate than the average user, we prioritize them for review. Once flagged, our YouTube content reviewers make the call on what content is or is not removed by applying our YouTube policies. We apply our policies transparently and consistently, without taking into account either the political leanings of the author, or their place in society.

Deepfakes

Deepfakes – AI-generated videos that make people appear to do or say things they didn't actually do or say – present serious homeland security risks. In fact, in other countries such as Gabon and Malaysia, we've already seen fake videos of leaders trigger political turmoil. In Ukraine, there was a deepfake of President Zelensky supposedly telling Ukrainians to lay down their arms against Russia. While platforms have

increasingly adopted policies to ban many deepfakes, these policies only can be enforced if you know what is or is not real.

What is YouTube doing to improve deepfake detection? And in cases where deepfakes are too advanced to be detected, how do you go about determining a video's authenticity?

YouTube is a global platform with over 2 billion users in more than 100 countries speaking 80 different languages. Responsibility to our users across the globe is our top priority, and to advance it we remove content that violates our publicly available Community Guidelines, including those against manipulated media such as deepfakes (available at https://www.youtube.com/howyoutubeworks/policies/community-guidelines/). We continuously work to improve on the processes and technology that help us enforce our policies, including against threats related to manipulated content on our platform.

Our Community Guidelines include a number of policies to protect users from violent or dangerous content, including misleading or deceptive content that may pose a serious risk of egregious harm. Pursuant to YouTube's deceptive practices policy, YouTube prohibits the use of manipulated media to take advantage of the YouTube community, whether this media is AI-enabled or otherwise (available at https://support.google.com/youtube/answer/2801973?hl=en#zippy=%2Cmisleading-metadata-or-thumbnails%2Cvideo-spam). An actor using deepfake technology to potentially cause such harm would see their content subject to our policies like anyone else. Additionally, we terminate channels that attempt to impersonate another person or channel, misrepresent their country of origin, or conceal their association with a government actor (available at https://support.google.com/youtube/answer/2801947?hl=en#).

YouTube has teams that - in conjunction with external experts, trusted flaggers, and NGOs, among other groups - continuously survey YouTube and other online platforms to track new symbols, terminology and memes. We determine whether these symbols, terms, and memes have a presence on our platform, assess them against our existing policies, and take action accordingly. For example, our teams continue to closely monitor the ongoing war in Ukraine, including analyzing guidance from outside experts and government bodies, and we have removed more than 76,000 videos and 9,000 channels related to the ongoing war in Ukraine for violating our Community Guidelines and Terms of Service.

With respect to paid content, our misrepresentation ads policy across Google Ads and AdSense specifically prohibits "deceptively doctoring media related to politics, social issues, or matters of public concern."

We work diligently to address manipulated media on YouTube across dozens of languages. As is the case with our other policies, we enforce our deceptive practices policy globally, and our automated systems flag violative videos in multiple languages.

Written Questions Submitted by Sen. Padilla to Neal Mohan

1. YouTube users span the globe, and it makes the same commitments to Western English- speaking communities as it does to non-Western, non-English language speaking communities. There has been strong congressional interest in ensuring technology companies are investing in protecting every user, not just English language natives, or those from Western countries.

a. **What are the top five non-English languages for content that users within the United States encounter on YouTube and what percentage of users in the United States encounter content within each of those languages on a weekly basis?**

b. **For each of these questions, please break down by employment status (full time in-house employees or contract-level).**

   i. **How many employees focus exclusively on trust and safety issues at YouTube?**

   ii. **Of the people working exclusively on trust and safety issues, how many focus on the content moderation of non-English language content, how many of them focus on each of those top five non-English languages, and how many of them focus on non-United States users?**

   iii. **Of the people working exclusively on trust and safety issues, how many of them focus on non-English language content in the United States? Of the people working exclusively on trust and safety issues, how many of them focus on non-English language content for users outside of the United States?**

   iv. **How many of them focus on each of those top five non-English languages?**

   v. **How many of them focus primarily on United States-based users, and how many of them focus primarily on non-United States based users?**

c. **When making community standards enforcement decisions is all content reviewed in its original language or are some or all subject to automated translation before being reviewed? What percentage of total content reviewed is subject to automated translation?**

d. **Are YouTube's community standards available in each of the languages of your users?**

e. **What is YouTube doing to ensure that it has cultural competency for all the communities it serves?**

YouTube is a global platform with over 2 billion users in more than 100 countries speaking 80 different languages — and responsibility is our top priority. To achieve our responsibility goals,

YouTube removes content that violates our publicly available Community Guidelines (available at https://www.youtube.com/howyoutubeworks/policies/community-guidelines/).

Before we do the work of removing content that violates our policies, we have to make sure the line between what we remove and what we allow is drawn in the right place — with a goal of preserving free expression while also protecting and promoting a vibrant community. To that end, we have a dedicated policy development team that systematically reviews all of our policies to ensure that they are current, keep our community safe, and do not stifle YouTube's openness. These policies are global, and when developing them, we consider regional differences by working with our internal policy teams, as well as outside experts. This is because YouTube's Community Guidelines are enforced consistently across the globe, regardless of the language or location of where the content is uploaded, and we want to make sure proposed policies can be applied fairly around the world. Across Google, more than 20,000 people in 2020 worked in a variety of roles to help enforce our policies and moderate content. These individuals are located across the globe, and possess a diverse set of backgrounds, including an array of linguistic capabilities and varied regional contexts.

We endeavor to be transparent in our work to enforce our Community Guidelines and release a quarterly report that provides a detailed overview of our work. For example, in the second quarter of 2022, a majority of content removed from YouTube was from users outside of the United States. More specifically, of the top 30 countries with the highest number of content removed in Q2, 88.3% was removed outside of the United States. It is also important to note that English is not the primary language in 9 of the top 10 countries where videos were removed.

In addition to our work to remove violative content, YouTube also strives to provide context and raise authoritative information before and during viewer engagement. This is especially important in the case of topics that are prone to misinformation, such as COVID-19. A range of information panels on video watch pages and in search results are a key mechanism to provide users additional context from authoritative sources on such topics. In the U.S., these info panels are available in several different languages, including Spanish, French, and Chinese. Additionally, we work to prominently surface authoritative content in search results in a wide diversity of languages. For example, if a user searched for "COVID-19 vacuna," they would receive authoritative content about the COVID-19 vaccine in Spanish. Since March 2020, we have shown information panels in Spanish in the United States across a range of topics, including elections and COVID-19, over 700M times. YouTube continues to elevate authoritative content in Spanish for users in the U.S., from news sources such as Noticias Telemundo (6.32 million subscriptions), Univision Noticias (7.36 million subscriptions), and CNN en Español (3.47 million subscriptions), among other sources.

2. In your testimony you shared that videos identified as borderline content are demoted in recommendations and the company saw a 70% drop in watch time on this content.

a. **What percentage of total watch time on YouTube comes from content identified as borderline content?**

b. **Is total watch time a metric that the company uses to assess its performance? Within that, does YouTube track what percentage of that watch time is devoted to borderline content?**

c. **How does YouTube assess its false positive rate—content you actioned for perceived violation of community standards, but that didn't actually violate the rules—and what is that rate? Please break this down by the top five languages on your platform.**

d. **Internally, what metric do you use to evaluate the success of enforcement of your community standards?**

In addition to removing content that violates our policies, we also work to ensure that we do not proactively expose users to content that is potentially harmful. We reduce recommendations of the "borderline content", or content which comes close to but does not violate our policies, that represents less than 1% of the content watched on YouTube. In January 2019, we announced that we would begin reducing recommendations of borderline content or videos that could misinform users in harmful ways, and we continue to extend these efforts to more countries outside of the United States and into non-English-language markets.

Determining what is harmful misinformation or borderline content is challenging, especially given the wide variety of videos uploaded to YouTube. To do so, we rely on external evaluators from around the world to provide input on the quality of a set of videos. These evaluators are trained on the same rater guidelines as Google Search to guide their work (these guidelines are available at https://support.google.com/youtube/answer/9230586). Each evaluated video receives up to nine different ratings, with some content requiring ratings from certified experts in the field. For example, medical doctors provide guidance on the validity of videos about specific medical treatments to limit the spread of medical misinformation. Based on consensus input from these raters, we use well-tested machine learning systems to build models that help review hundreds of thousands of hours of videos every day to identify and limit the spread of borderline content. The accuracy of these systems continues to improve over time.

With respect to content that we remove for having violated our policies, we track the percentage of views on YouTube that comes from this content. We refer to this metric as the Violative View Rate, or VVR. In the first half of 2022, the VVR was .09%-.11%. In other words, for every 10,000 views on YouTube, only 9-11 went to content that proved to be violative. In order to hold ourselves accountable to our users and the public at large, we publish a quarterly

Community Guidelines Enforcement Report (available at https://transparencyreport.google.com/youtube-policy/views?hl=en) that provides data on VVR and our enforcement across channels, videos, comments, user ags, and appeals and reinstatements. Our methodology for this report has been validated by MIT Sloan professor of statistics Dr. Arnie Barnett as "thoroughly sensible and statistically sound" (Dr. Barnett's analysis is available at: https://www.analysisgroup.com/globalassets/insights/publishing/2021-youtube-vvr-methodology-a-statistical-assessment-arnold-barnett.pdf). We are committed to reducing this metric over time and to transparency around related efforts.

The openness of our platforms has helped creativity and access to information thrive. It is our responsibility to protect that, and to prevent our platforms from being used to incite hatred, harassment, discrimination, and violence. We are committed to taking the steps needed to live up to this responsibility today, tomorrow, and in the years to come.

3. Algorithms play an important role in suggesting content to users on YouTube—they not only suggest another video, but the recommendation algorithms themselves guide the types of content that creators make to reach the most people. Both honest and dishonest creators know this. In your testimony, you mention that you demonetize content that is in violation of your community guidelines, including content that results in harassment & cyberbullying. Is the company considering efforts to dis-incentivize the creation of any inflammatory content—whether or not it directly violates your community guidelines?

Our mission at YouTube is to "give everyone a voice and show them the world." As such, our search and recommendation systems are not designed to filter or demote videos or channels based on specific political perspectives. We take extraordinary lengths to build our products and enforce our policies in a consistent manner. Our business model is dependent on being a useful and trustworthy source of information for everyone, so we have a natural, long-term business incentive to prevent anyone from interfering with the integrity of our products.

When developing and updating our policies, we solicit perspectives from a range of voices, including creators, subject-area experts, free speech proponents, and policy organizations from across the political spectrum. Once a policy has been developed, we invest significant time making sure new policies are consistently enforced by our global team of reviewers who follow objective guidelines Our work is also guided by our four pillars of responsibility: removing violative content; raising up authoritative content; reducing the spread of borderline content; and rewarding trusted creators — what we refer to as the 4Rs of responsibility. YouTube's Community Guidelines outline what content is permissible and what content is violative on our platform. We enforce these Community Guidelines using a combination of human reviewers and machine learning, and apply them to everyone equally — regardless of the subject or the creator's background, political viewpoint, position, or affiliation.

We believe strongly in the freedom of expression and access to information, and we know that the overwhelming majority of creators follow our guidelines and understand they are part of a large, influential, and interconnected community. However, we also know that we have a responsibility to protect our users, which is why we have policies prohibiting hate speech, terrorist content, and other content that violates our policies, as well as stricter standards for who can monetize their content. Each of the products and services we offer has a different purpose, and we tailor our approach carefully to the content that should be available on each product and service. For example, in October 2020, we updated YouTube's hate and harassment policy to prohibit content targeting individuals or protected groups based on conspiracy theories used to justify real world violence (e.g., QAnon and Pizzagate).

Extreme content is both contrary to our responsibility objectives and bad for business. It undermines user trust and satisfaction and it pushes advertisers away from the platform. Our system is both designed and incentivized to elevate authoritative voices and surface relevant information over inflammatory content. Creators who violate those rules may have their content removed or their accounts terminated. When we detect a video that violates our Community Guidelines, we remove the video and apply a strike to the channel. The strike restricts a creator's ability to post or create content on the platform for one week. If the creator's behavior warrants another strike within 90 days from the first, a new two-week prohibition from posting or creating content is implemented. A third strike within 90 days results in permanent removal of a channel from YouTube. Creators can appeal those strikes if they believe we are mistaken. We also terminate entire channels if they are dedicated to posting content prohibited by our Community Guidelines or contain a single egregious violation, like child sexual abuse material.

We reward our trusted creators through monetization programs. Creators who earn money on YouTube must follow YouTube's channel monetization policies (available at https://support.google.com/youtube/answer/1311392), which include YouTube's Community Guidelines, Terms of Service, Copyright, and Google AdSense program policies. Violation of our YouTube channel monetization policies may result in monetization being suspended or permanently disabled. Additionally, creators who follow our policies can apply to join the YouTube Partner Program, a program that sets a higher bar for which channels can make money on our site (for more information on our Partner Program, please see https://support.google.com/youtube/answer/72851).

4. Election integrity is more important than ever. Platforms have a crucial role to play in ensuring that voters are able to access good, accurate election information, and in maintaining the basic integrity of our elections system. YouTube has made announcements about what it is doing around the midterm elections, and much of your strategy for the midterms are similar to steps taken two years ago.

a. **Have you studied how effective your efforts were during the last election to ensure that your users knew where, how, and when to vote, and if so, what were the results of those studies?**

b. **What do you assess to be the biggest threats to election integrity on YouTube in advance of the 2022 elections? What do you assess to be the biggest threats to election integrity on YouTube in advance of the 2024 elections? What steps are you taking to mitigate these identified threats?**

c. **How are you staying ahead of evolving threats to election integrity?**

d. **What is YouTube doing to keep election officials safe from doxxing and other online harassment campaigns?**

Throughout the 2020 U.S. election season, given the extraordinary circumstances of the COVID-19 pandemic, YouTube served as a helpful destination for people to learn more about where and how to vote and to gain information about specific candidates or issues. Our main goal going into the election season was to make sure we connected people with authoritative information, while limiting the reach of misinformation and removing harmful content.

Following the 2020 U.S. election, YouTube completed a retrospective report (available at https://kstatic.googleusercontent.com/files/a5a96bfa10fa6b28cfbf9750e0730da82ce1ca638398b57a00f0c276bc42e78151a297577f382e402087a514dac728a59a6b6655f1c4c5823ccea945d16ed528), which provides a thorough overview of our work to support election integrity in the 2020 U.S. Presidential and Congressional elections. Among other findings, the report notes that, between September and November 2020, election results information panels were collectively shown over 8 billion times and candidate information panels surfaced for viewers over 6 million times.

With respect to informing users about voting information, we showed an information panel at the top of results when viewers searched YouTube for specific queries related to how to vote. This panel linked to Google's "How to vote" feature, which provided authoritative information about how to vote in a viewer's state, including details such as ID requirements, registration and voting deadlines, and guidance for different means of voting, such as in-person voting or voting by mail. More information about our platform's engagement with users about elections is available at https://elections.google/#engaging-voters.

To help viewers find relevant information, our search and recommendation algorithms are designed to raise authoritative content, and reduce the spread of borderline content and harmful misinformation. Between September and November, 2020, we saw that on average 88% of the videos in top 10 search results related to the 2020 U.S. elections came from authoritative news sources (the rest include content such as news-based late-night shows,

creator videos, and commentary). In addition, over 70% of non-subscribed recommendations on U.S. election-related topics came from authoritative news sources, and the top recommended videos and channels for election-related content were primarily authoritative news. In fact, when it comes to election-related content, the top 10 authoritative news channels were recommended over 14 times more than the top 10 non-authoritative channels. The most viewed channels and videos were from news channels like NBC and CBS.

We continue to learn from the 2020 election period and to apply the lessons learned to our work going forward. Among other updates, we have made it easier for people using Google Search to spot misinformation online and make more informed decisions about the sites they want to visit. We created a menu icon next to search results that users can tap to learn more about the result or feature and the source of the information. We also announced a new feature on Search that will provide the user notice when a topic is rapidly evolving and a range of sources has not yet weighed in (more information about this feature is available at https://blog.google/products/search/new-notice-search-rapidly-evolving-results/). In addition, we recently created a resource page to help users evaluate the information they find on Google and its sources (available at https://support.google.com/websearch/answer/12003459?hl=en). And on YouTube, we continuously use new data to train our system to be faster and more accurate when it comes to identifying viral misinformation narratives.

With so many users around the world coming to YouTube to learn about political leaders, participate in civic engagement, and develop informed opinions about current events, we have a responsibility to support an informed citizenry and foster healthy political discourse. To deliver on this responsibility, we remove policy-violative content, raise authoritative news sources, and provide a range of resources for civics partners such as government officials, candidates, civics organizations, and political creators to ensure a broad range of voices are heard.

At YouTube, we are constantly working to make sure we can be a reliable source for timely news and information. Over many years, we have built policies, systems and teams that raise authoritative content and limit the spread of harmful misinformation. Our dedicated team of elections experts has been focused on preparing for the 2022 U.S. midterm elections. Whether it's learning about when and where to vote, or finding information about political candidates, we take seriously our commitment to connecting viewers with trusted resources. More information about our work at YouTube supporting elections is available at https://blog.youtube/news-and-events/the-2022-us-midterm-elections-on-youtube/.

As we have done for previous elections, we rolled out features on Google Search to connect voters with accurate information about voter registration and how to vote. When people search for "how to vote" in either English or Spanish, they find election information panels

sourced from Democracy Works, a nonpartisan and nonprofit data provider that works directly with state election administrators to aggregate information about how they can vote in their state, along with key dates and deadlines and guidance for options like voting early, in person or by mail (more information on Democracy Works is available at https://www.democracy.works/). The information panels will also link to state government official websites, which provide additional information. As with prior U.S. elections, we are working with The Associated Press to present authoritative election results on Google.

To train and protect the security of people working on elections, Google partners with organizations like Defending Digital Campaigns, which has provided free security keys to over 300 national committees, state parties, and political campaigns in all 50 states to date (more information on these election security efforts is available at https://blog.google/technology/safety-security/furthering-our-support-election-security/). To ensure the strongest account and site-level protections, we offer our free Advanced Protection Program and free Project Shield service to safeguard against digital attacks (more information on these programs is available at https://landing.google.com/advancedprotection/ and https://projectshield.withgoogle.com/landing). Through our Campaign Security Project, we've teamed up with organizations across the political spectrum to train over 4,000 campaign and election officials on security best practices, including products and tools they can use to stay safe online. More information about these efforts is available at https://blog.google/technology/safety-security/safer-internet-day-2022/ and https://blog.google/outreach-initiatives/civics/our-ongoing-work-to-support-the-2022-us-midterm-elections/.

As Election Day approaches, we recognize that we need to stay vigilant and nimble in our approach, and we will continue to closely monitor new developments and make needed adjustments along the way. Elections are one of the cornerstones of our society, and we are committed to keeping viewers informed and protected this November, and for elections to come (more information concerning YouTube's efforts concerning the 2022 elections is available at https://blog.youtube/news-and-events/the-2022-us-midterm-elections-on-youtube/).

Written Questions Submitted by Sen. Lankford to Neal Mohan

1. An October 2021 NPR report discussed how migrants and cartels have used YouTube to provide information on where to cross the U.S. southern border and routes to take to get to the U.S. southern border.[1] During our hearing, I asked you about a specific video, which had over 2.7 million views and had been posted over 2 years ago. That video described where U.S. Border Patrol (USBP) would be stationed in a specific sector and how to cross the border without being detected by USBP. Why has that video been allowed to receive millions of views on YouTube and remain on your platform? How many other videos are currently on YouTube with similar content?

This particular video includes documentary content that discusses U.S. border policy, border security measures, and the evolution of illegal immigration over time, U.S.-Mexico economic relationships, factors motivating illegal immigration, and other topics commonly found in television or film. Videos that might otherwise violate our policies may be allowed to stay on YouTube if the content offers a compelling reason with visible context for viewers. We often refer to this exception as "EDSA", which stands for Educational, Scientific, Documentary, or Artistic context. The categories describe the reasons why we might allow videos that would otherwise violate our Community Guidelines.

To help determine whether a video might qualify for an EDSA exception, we look at multiple factors, including the video title, descriptions and the context provided in the video's audio or imagery. Context is important, and we make these nuanced decisions on a case by case basis by looking at multiple factors, including the video title, descriptions and other context. EDSA exceptions are a critical way we make sure that important speech stays on YouTube, while protecting the wider YouTube ecosystem from harmful content. For more information on how we treat EDSA content, please see https://blog.youtube/inside-youtube/look-how-we-treat-educational-documentary-scientific-and-artistic-content-youtube/.

2. What are YouTube's content moderation policies around information related to crossing the U.S. border? Please describe these policies in detail.

Our Community Guidelines set forth what content is not allowed on YouTube (available at https://support.google.com/youtube/answer/9288567?hl=en). With respect to violent and dangerous content, we have a network of robust policies in place that work together to combat it.

Content intended to praise, promote, or aid violent criminal organizations is not allowed on YouTube. This includes content aimed at recruiting new members to violent criminal or terrorist organizations designated by the U.S. government (more information on this policy is available at https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436). Our

policies also prohibit content that encourages dangerous or illegal activities that risk serious physical harm or death, as well as content that incites others to commit violent acts against individuals or a defined group of people (more information on these policies is available at https://support.google.com/youtube/answer/2801964 and https://support.google.com/youtube/answer/2802008.

In addition to removing content that violates our policies, we also work to ensure that we do not proactively expose users to content that is potentially harmful. We reduce recommendations of the "borderline content", or content which comes close to but does not violate our policies, that represents less than 1% of the content watched on YouTube. We also raise up authoritative voices by providing users with more information about the content they are seeing to allow them to make educated choices. For more information on information panels, please see https://support.google.com/youtube/answer/9004474, and on fact-checks, please see https://support.google.com/youtube/answer/9229632.

We recognize that sometimes videos that might otherwise violate our policies may be allowed to stay on YouTube if the content offers a compelling reason with visible context for viewers. As mentioned in the response to your previous question, we refer to this exception as "EDSA", which stands for Educational, Scientific, Documentary, or Artistic context. The categories describe the reasons why we might allow videos that would otherwise violate our Community Guidelines. For example, we may allow news coverage of border crossings or graphic footage taken by a citizen, recognizing there may be a documentary public interest in that content. Context is important, and we make these nuanced decisions on a case by case basis by looking at multiple factors, including the video title, descriptions and other context.

There are also certain types of content where we don't allow an EDSA exception under any circumstances because of the sensitivity and egregiously harmful nature of the content, or when it violates the law. For example, content that endangers children or any content with footage of deadly violence filmed by the perpetrator is not allowed on YouTube, regardless of the context.

We have a dedicated team that responds to law enforcement around the clock, every day of the year, including on issues related to violent organizations. We also work proactively with law enforcement agencies in a variety of ways to assess threats and to counter attempts to deceive, harm, or take advantage of people using our platforms. Our interactions with law enforcement include the following:

- First, we maintain regular communication channels with law enforcement, as communication with law enforcement, industry partners, and the federal government is a key component of our efforts to keep users safe. We rely on information from industry peers and law enforcement to ensure the integrity of our platform and act

swiftly in response to crises or when we detect abuse that may threaten public safety or the integrity of democratic processes.

- Second, we have dedicated teams of analysts and security experts who work around the clock to identify and investigate possible instances of coordinated influence operations on our platform. Google's Threat Analysis Group works to identify bad actors and terminate their channels and accounts, works with other technology companies to share intelligence and best practices, and shares threat information with law enforcement. For more information regarding our Threat Analysis Group, please see https://blog.google/threat-analysis-group/.

- Third, we have processes in place to proactively refer to law enforcement imminent threats and certain illegal activity occurring on our platform. These imminent threats are raised to us in a variety of ways. In addition to our ongoing threat analysis just noted, we also have processes in place for our various internal product teams, as well as external sources, to escalate potential threats and criminal activity if they see it on our platforms. Google has a team - the CyberCrime Investigation Group (CCIG) - in place to assess escalated threats, and that team refers matters to law enforcement whenever appropriate.

Additionally, we developed the YouTube Trusted Flagger program to help provide more sophisticated reporting processes for government agencies and non-governmental organizations (NGOs) that are particularly effective at notifying YouTube of content that violates our Community Guidelines. The program provides these partners with dedicated reporting processes and a channel for ongoing discussion and feedback about YouTube's approach to various content areas. The program is part of a network of around 300 government partners and NGOs that bring valuable expertise to our enforcement systems. Participants in the Trusted Flagger program receive training in enforcing YouTube's Community Guidelines, and because their flags have a higher action rate than the average user, we prioritize them for review. Once flagged, our YouTube content reviewers make the call on what content is or isn't removed by applying our YouTube policies. We apply our policies transparently and consistently, without taking into account either the political leanings of the author, or their place in society.

3. Were these policies developed in consultation with any entity outside of YouTube? If so, with whom?

When developing and updating our policies, we solicit perspectives from a range of voices, including creators, subject-area experts, free speech proponents, and policy organizations from across the political spectrum. Once a policy has been developed, we invest significant time making sure new policies are consistently enforced by our global team of reviewers who follow objective guidelines Our work is also guided by our four pillars of responsibility:

removing violative content; raising up authoritative content; reducing the spread of borderline content; and rewarding trusted creators — what we refer to as the 4Rs of responsibility. YouTube's Community Guidelines outline what content is permissible and what content is violative on our platform. We enforce these Community Guidelines using a combination of human reviewers and machine learning, and apply them to everyone equally — regardless of the subject or the creator's background, political viewpoint, position, or affiliation.

4. Has YouTube ever studied the prevalence of information on its platform related to crossing the U.S. border? If so, please share any such study with the Committee.

We constantly work to improve our systems to reduce recommendations on borderline content and videos that could misinform users in a potentially harmful way, particularly in areas that rely on veracity — such as news, medicine, historical events, or science.

Determining what is harmful misinformation or borderline content is challenging, especially given the wide variety of videos uploaded to YouTube. To do it, we rely on external evaluators from around the world to provide input on the quality of a set of videos. These evaluators use the same rater guidelines as Google Search to guide their work (these guidelines are available at https://support.google.com/youtube/answer/9230586). Each evaluated video receives up to nine different ratings, with some content requiring ratings from certified experts in the field. Based on consensus input from these raters, we use well-tested machine learning systems to build models that help review hundreds of thousands of hours of videos every day to identify and limit the spread of borderline content. The accuracy of these systems continues to improve over time.

To encourage further research, we recently expanded the YouTube Research Program, which gives academic researchers access to our data in a user privacy sensitive way, allowing them to derive metrics or develop insights of their own based on that data (more information on this program is available at https://research.youtube/).

5. How do YouTube's algorithms recognize content which violates YouTube's terms of service related to human smuggling, drug smuggling, cartel activity, or crossing the U.S. southern or northern borders? What keywords, images, hashes, or other content and indicators regarding such subject matter do YouTube's algorithms recognize as violating YouTube's terms of service? In answering this question, please provide all keywords, images, hashes, or other content regarding such subject matter which YouTube would consider as violating its terms of service.

More than 500 hours of video content on average is uploaded every minute to YouTube. Our Community Guidelines prohibit content that encourages dangerous or illegal activities that risk serious physical harm or death.

We rely on a combination of people and technology to flag inappropriate content and enforce these guidelines. Flags can come from our automated flagging systems, from members of the Trusted Flagger program (NGOs, government agencies, and individuals) or from users in the broader YouTube community. YouTube strives to prevent content that breaks our rules from being widely viewed—or viewed at all—before it's removed. Automated flagging enables us to act more quickly and accurately to enforce our policies.

We sometimes use hashes (or "digital fingerprints") to catch copies of known violative content before they are ever made available to view. For some content, like child sexual abuse images (CSAI) and terrorist recruitment videos, we contribute to shared industry databases of hashes to increase the volume of content our machines can catch at upload.

In addition to providing choice, we believe that transparency around our processes is important. As an indication of this commitment, we recently expanded the YouTube Research Program. The program gives academic researchers access to our raw data in a user privacy sensitive way, allowing them to derive metrics or develop insights of their own based on that data. This increased access includes expanded data quotas, ability to derive analytics, and in-house YouTube support. As we expand this access to our API, we will work closely with researchers to gain feedback and improve collaborations.

Finally, we provide users with the ability to influence the ads they see online. As explained in the Google Privacy Policy, if you use a Google Account, Google may use information you provide to show you relevant ads. Users can always learn more information concerning personalized advertising and controlling the ads they see on Google's support pages (available at https://support.google.com/google-ads/answer/2549116 and https://support.google.com/accounts/answer/2662856). In addition, users can turn off ad personalization in their Google Account settings (available at https://adssettings.google.com/).

6. How do YouTube's human content reviewers recognize content which violates YouTube's terms of service related to human smuggling, drug smuggling, cartel activity, or crossing the U.S. southern or northern borders? What keywords, images, hashes, or other content and indicators regarding such subject matter do YouTube's human content reviewers recognize as violating YouTube's terms of service? In answering this question, please provide all keywords, images, hashes, or other content regarding such subject matter which YouTube would consider as violating its terms of service.

Content intended to praise, promote, or aid violent criminal organizations is not allowed on YouTube. This policy (available at https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436) also applies to channels and content produced by violent criminal or terrorist organizations designated by the U.S. government, including content:

- praising or memorializing prominent terrorist or criminal figures to encourage others to carry out acts of violence;
- praising or justifying violent acts carried out by violent criminal or terrorist organizations;
- aimed at recruiting new members to violent criminal or terrorist organizations designated by the U.S. government;
- depicting hostages or posted with the intent to solicit, threaten, or intimidate on behalf of a violent criminal or terrorist organization; and depicting the insignia, logos, or symbols of violent criminal or terrorist organizations to praise or promote them.

As noted in the response to question 5, with hundreds of hours of new content uploaded to YouTube every minute, we use a combination of people and machine learning to detect problematic content at scale. Machine learning is well-suited to detect patterns, which helps us to find content similar to other content we have already removed, even before it can be viewed. For example, over time we may train machines to identify symbols of iconography an extremist group uses to promote its "brand" or attempt to recruit new members. Our teams monitor trends in risky forms of content, and use this information to ensure that we are prepared to address them appropriately before they can become a larger issue.

The YouTube community also plays an important role in flagging content they think is inappropriate. We rely on our users, as well as experts in our Trusted Flagger program, to help us spot potentially problematic content by reporting it directly to us. Once such content is identified, human content reviewers evaluate whether it violates our policies. If it does, we remove the content and use it to train our machines for better coverage in the future. Our content reviewers also assess whether flagged content has a clear educational, documentary, scientific, or artistic (EDSA) purpose and should, for that reason, be protected.

7. Please provide a list of each piece of content on YouTube that YouTube's moderators have removed that regard human smuggling, drug smuggling, cartel activity, or crossing the U.S. southern or northern borders. In providing this list, please include how much each piece of content was viewed, liked, and shared. Please also include the time each piece of content was allowed to remain on YouTube's platform prior to being removed.

Given the volume of content on our platform, we do not maintain lists of individual pieces of content by subject area. YouTube does, however, publish quarterly data in its Community Guidelines enforcement report. This report provides public data about the number of videos we remove from our platform under each of our content policies, as well as information about channel removals, views before removals, appeals and reinstatements, and human and machine flagging. To measure our progress on removing violative videos, we have developed a metric called the Violative View Rate (VVR), which is an estimate of the proportion of video views that violate our Community Guidelines, excluding spam, in a given quarter. In Q2 of 2022,

YouTube's VVR was 0.09-0.11%, meaning that out of every 10,000 views on YouTube, 9-11 are views of content that violates any of our Community Guidelines. Our data science teams have spent years refining this metric, which we consider to be an important indicator in measuring the effectiveness of our efforts to fight and reduce abuse on YouTube. Our objective is to continually reduce the VVR rate over time. The VVR metric was validated by a MIT Sloan professor of statistics as "thoroughly sensible and statistically sound" (Dr. Barnett's analysis is available at https://www.analysisgroup.com/globalassets/insights/publishing/2021-youtube-vvr-methodology-a-statistical-assessment-arnold-barnett.pdf). We will continue to expand the information we share through our transparency report, cross-industry initiatives, blog posts, public disclosures, and other mechanisms like tools for researchers. Our goal is to achieve transparency and accountability by providing meaningful information while protecting our platform.

8. How long does it take YouTube to detect and take down child sexual abuse material, both previously detected material and new material?

9. How many employees do you have who are dedicated to ensuring child sexual abuse material is appropriately removed and reported?

10. Would you describe YouTube efforts to take down and report child sexual abuse material as active or passive? In other words, is material passively scanned for known matches, or is there an active element to how you detect and take down this material?

Because the answers to these questions are related, we have grouped together our response to Questions Nos. 8 through 10.

Detecting and taking down child sexual material as quickly as possible is a priority for YouTube, and we have always had clear policies against videos, playlists, thumbnails and comments on YouTube that sexualise or exploit children. We have policies that prohibit harmful and dangerous content involving minors, and our Community Guidelines prohibit sexual content or content with other inappropriate themes involving minors. Child Sexual Abuse Material (CSAM) represents a fraction of one percent of the content we remove. Detailed information about our policies and approach on enforcement concerning child safety is available in our transparency report (available at https://transparencyreport.google.com/youtube-policy/featured-policies/child-safety).

We use machine learning systems to proactively detect violations of these policies and have human reviewers around the world who quickly remove violations detected by our systems or flagged by users and our trusted flaggers. While some content featuring minors may not violate our policies, we recognise that the minors could be at risk of online or offline

exploitation. This is why we take an extra cautious approach when enforcing these policies. Our machine learning systems help to proactively identify videos that may put minors at risk and apply our protections at scale, such as restricting live features, disabling comments, and limiting video recommendations. And if a user comes across content that they think is depicting a child in danger or an abusive situation, they can flag the video or file a report. If we believe a child is in danger based on content that has been reported to us, we will investigate and take appropriate action as quickly as possible. Should law enforcement seek additional information from us, we will provide that pursuant to valid legal process or in accordance with applicable laws.

We have heavily invested in engineering resources to detect CSAM in ways that are precise and effective, and have long used this technology to prevent the distribution of known child sexual abuse imagery (CSAI) videos on YouTube. Our proprietary CSAI Match technology, which we license to a number of other technology companies, allows us to detect known CSAI videos. In cases where a video contains CSAI or a user solicits CSAI through comments or other communications, our team reports it to the National Center for Missing and Exploited Children (NCMEC), who then liaise with global law enforcement agencies. Once we have identified a video as illegal and reported it to NCMEC, the content is hashed—i.e., given a unique digital fingerprint—and used to detect matching content. This hashing and scanning technology is highly precise at detecting known CSAI and enables us to detect illegal content more quickly. We maintain a database of known CSAI hashes and any content that is matched against this list is removed and reported to NCMEC.

In addition to our long-standing efforts to combat CSAI video, we have made large investments to detect and remove content which may not meet the legal definition of CSAI, but where minors are still being sexualized or exploited. We continue to invest more resources to ensure children and families have a safe experience on YouTube.

11. There has been an alarming uptick in sextortion (meaning a child is threatened or blackmailed with the potential release of sexual material in order to receive additional sexual content or money) and grooming on social media platforms. Between 2019 and 2021, the number of reports involving sextortion more than doubled. What has your platform done to address this alarming trend?

We have clear policies that prohibit content that exploits or endangers minors on YouTube and have committed significant time and resources toward removing violative content as quickly as possible. We regularly review these policies and engage with external experts to ensure they are current.. In Q2 of 2022 alone, we removed over 1.3 million videos for violations of our child safety policies.

In addition to our long-standing efforts to combat CSAI video, we have made large investments to detect and remove content which may not meet the legal definition of CSAI, but where minors are still being sexualized or exploited. We use machine learning systems to proactively detect violations of these policies and have human reviewers around the world who quickly remove violations detected by our systems or flagged by users and our trusted flaggers. Our machine learning systems help to proactively identify videos that may put minors at risk and apply our protections at scale, such as restricting live features, disabling comments, and limiting video recommendations. We continue to invest more resources to ensure children and families have a safe experience on YouTube (more information is available at https://protectingchildren.google/#fighting-abuse-on-our-own-platform-and-services).

12. When YouTube notifies a user that their content has been removed does it include a timestamp to point to any specific instances of community guideline violations? If not, why?

13. If requested, are users able to receive the exact quote or action that resulted in a violation of community guidelines or terms of use? If not, why?

14. Please detail the appeals process for violations of community guidelines and terms of use. Please include the average response time to resolve an appeal request.

15. Once a video is removed, does YouTube grant access of the file to the owner of the video or does the company retain the file?

Because the answers to these questions are related, we have grouped together our response to Questions Nos. 12 through 15.

When we let a user know about a strike, we let them know what kind of strike they received. If the strike is for a Community Guidelines violation we also tell the user which specific policy their content violated. Additionally, Community Guidelines strikes will be visible in a user's Channel settings.

We also provide our users with information about videos that have been removed from YouTube. For example, if a user sees the "Video removed: Inappropriate content" message next to one of their uploaded videos, it means the video in question has been found to violate our Community Guidelines. If a user sees the "Video removed: Terms of Use violation" message next to one of their videos, the video may have been rejected due to a Terms of Use or copyright violation.

We recognize that we sometimes make mistakes in enforcing our Community Guidelines, including in our removals of potentially violative content. Creators have 30 days to submit an appeal after a video's removal. In Q2 of 2022, we saw 223,286 such appeals.

If a creator chooses to submit an appeal, it goes to human review, and the original decision is either upheld or reversed. The appeal request is reviewed by a senior reviewer who did not make the original decision to remove the video. The creator receives a follow up email with the result.

Following an appeal, if we find that content did not violate our Community Guidelines, we will reinstate it and remove the strike from the channel. If we find the content did not violate our Community Guidelines, but is nevertheless not appropriate for all audiences, we will apply an age restriction. If we find that the content was in violation of our Community Guidelines, the strike will stay and the video will remain off the site. There is no additional penalty for appeals that are rejected.

Privacy is for everyone. We have a suite of tools designed to give users security, transparency, and meaningful control over their data while making things as simple as possible. Over 1 billion users have reviewed their Google privacy settings, while more than 20 million people visit their Google Account page–where they can review their privacy settings and see and delete data stored with their account– every day. In 2021, we announced a new auto-deletion default for the Location History, Web & App Activity settings. We recently also expanded that to YouTube Search history.

Written Questions Submitted by Sen. Scott to Neal Mohan

1. **How many accounts have you found advertising counterfeit pills on your platform?**

2. **How many of those accounts were reported to you by a user of your platform?**

3. **How many of those accounts did you find on your own?**

4. **Are you committed to working with appropriate local, state, and federal law enforcement to create a Trusted Reporter program to help bring down these accounts?**

Because the answers to these questions are related, we have grouped together our response to Questions Nos. 1 through 4.

YouTube prohibits the sale or promotion for sale of counterfeit goods, or goods containing a trademark or logo that is identical to, or substantially indistinguishable from, the trademark of another (more information is available at https://support.google.com/youtube/answer/6154227?hl=en). These goods mimic the brand features of the product in an attempt to pass themselves off as a genuine product of the brand owner. Channels that promote or sell counterfeit goods may be terminated.

YouTube's Community Guidelines also clearly prohibit the sale of counterfeit pills. Our policy on the sale of illegal or regulated goods and services bans content on YouTube that aims to directly sell, link to, or facilitate access to a wide range of regulated goods and services, including controlled narcotics, other drugs, and pharmaceuticals without a prescription (more information is available at https://support.google.com/youtube/answer/9229611?hl=en). The policy strictly prohibits any content that makes the sale of these items or facilitates the use of these services possible by posting links, email, phone number or other means. This policy applies to videos, video descriptions, comments, live streams, and any other YouTube product or feature.

Our policies and procedures are designed to be dynamic and responsive to emerging trends in substance abuse disorders and the online sale of drugs. We are proud that these policies and processes have benefitted from feedback and close partnerships with law enforcement and regulatory agencies in the United States, including the Department of Health & Human Services (HHS), the Substance Abuse and Mental Health Services Administration (SAMHSA), the Office of the Surgeon General (OSG), and Drug Enforcement Administration (DEA), and the Food and Drug Administration (FDA). These partnerships have resulted in innovative and creative techniques to combat the sale of illicit drugs, and to make the Internet safer and more useful for users seeking treatment and recovery.

Finally, Google contracts with LegitScript, an independent company, to significantly increase the breadth and scope of non-pharmaceutical products prohibited by Google's policies from being surfaced on its products, including in ads, shopping, and YouTube. LegitScript keeps track of thousands of internet pharmacies and dangerous health products, and tracks merchants of dangerous health products such as "legal highs" and tainted supplements. LegitScript also runs sweeps which it then reports to Google for enforcement. These notices are received from LegitScript twice a week and have resulted in the removal of tens of thousands of ads from Google's systems.

We have also worked with LegitScript to monitor and prohibit thousands of new products, ingredients, and psychoactive substances, regardless of legal status, based on legitimate concerns about the safety of these products and substances from health professionals, regulators, and law enforcement entities around the world. LegitScript also runs regular sweeps on YouTube for videos of purported offers to sell illicit opioids which are reported for removal.

5. Do you proactively refer accounts mentioning counterfeit pills or fentanyl to law enforcement?

a. **Do you notify the account holder that they are being investigated by law enforcement, potentially leading them to evade and continue selling their product and killing people?**

b. **Which law enforcement entities are you referring accounts suspected of drug sales and/or trafficking to?**

6. Over the last year, on average, how many subpoenas have you received per month related to accounts being used to sell or distribute drugs?

7. What is the average turnaround time for responding to a subpoena issued by a law enforcement agency (Not including an automated or "received" style response)?

Because the answers to these questions are related, we have grouped together our response to Questions Nos. 5, 6, and 7.

Google has consistently worked with the FDA, the DEA, and other regulatory and law enforcement agencies involved in enforcing laws and regulations concerning the sale of drugs online. For years, Google has made numerous referrals to law enforcement, including the FDA's Office of Criminal Investigations and the DEA's Special Operations Division. We provide proactive referrals to the FDA and DEA in different ways, including phone calls, meetings, and emails.

In regards to Google's commitment to combat the opioid and fentanyl epidemic, beginning in 2018, we partnered with the DEA on National Prescription Drug Take Back Day, promoting the DEA's initiative across Google properties twice a year (more information about this partnership is available at https://www.washingtonpost.com/news/the-switch/wp/2018/04/25/google-is-promoting-the-deas-take-back-day-on-its-homepage-to-help-combat-the-opioid-crisis/). Through this partnership, we developed a Google Maps API locator tool to help people dispose of their prescription drugs at temporary event locations (more information is available at https://blog.google/outreach-initiatives/public-policy/opioid-crisis-maps-search/). With the help of this tool, the DEA and its local partners have collected 7,000 tons of unused medication since the program's inception.

Google works cooperatively with law enforcement while respecting the privacy of our users. When we receive a request from a government agency, we carefully review it to ensure it satisfies applicable laws and send an email to the user account before disclosing the appropriate information. If the account is managed by an organization, we provide notice to the account administrator. We do not give notice to users when such notice is legally prohibited under the terms of the request, though we will provide notice after any legal prohibitions are lifted, such as when a statutory or court-ordered gag period has expired. In cases when an account has been disabled or hijacked, we might not give notice. Similarly, we might not give notice in the case of emergencies, such as threats to a child's safety or threats to someone's life. In these cases, we will provide notice if we learn that the emergency has passed.

Each year, Google receives hundreds of thousands of requests for information from law enforcement and government agencies. We were the first major company to publish a public Transparency Report (available at https://transparencyreport.google.com/user-data/overview?hl=en) that presents a comprehensive data set encompassing all demands we receive from government agencies for user information. The Global Transparency Report includes categories of government requests from all countries that we have processed or completed during the time periods specified by the report (including subpoenas), except for requests issued under U.S. national security laws, which are detailed separately.

8. Do your moderators search for accounts or posts that use the DEA's Emoji Drug Code, or various permutations of the Emoji Drug Code? If no, why not?

As noted in response to Question 2, we use a combination of people and machine learning to detect problematic content at scale. Flags can come from our automated flagging systems, from members of the Trusted Flagger program (NGOs, government agencies, and individuals) or from users in the broader YouTube community.

9. What has your company done to help spread awareness of the DEA's One Pill Can Kill campaign?

Google is combining our products and technology with government and non-profit efforts to empower families and individuals to prevent, cope, and recover from drug addiction. We are committed to continuing to improve our products and collaborate with key public and private sector stakeholders to help curb the sale of illicit drugs online and combat drug addiction more generally.

In an effort to help spread awareness about the dangers of fentanyl overdose, specifically, we updated our Recover Together site to include information about the risks associated with fentanyl and "fentapills" and how to recognize and respond to overdoses (more information is available at https://recovertogether.withgoogle.com/overdose-awareness/). We launched this content in conjunction with Fentanyl Awareness Day on May 10, 2022, and promoted it on the Google homepage to call attention to this important issue.

Search surfaces knowledge panels in response to search queries relating to opioid addiction. The panels provide information on the prevalence and dangers of opioid abuse and dependence, as well as on the symptoms of substance abuse disorder and treatment options. When users search for queries like "help for opioid addiction", we also surface a text box in Google and YouTube search results containing information about the Substance Abuse and Mental Health Services Administration (SAMHSA), such as the agency's National Hotline and a link to FindTreatment.Gov. Across Search and YouTube, we have provided financial support upwards of one million dollars for the Partnership to End Addiction's public service announcements, including the organization's "Start the Connection" campaign, aimed at families dealing with the opioid epidemic (more information is available at https://www.youtube.com/watch?v=29fxb07Pc6w&t=5s).

TikTok

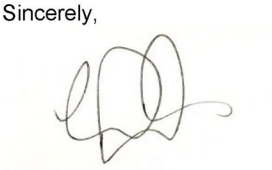
October 28, 2022

Chairman Gary Peters
U.S. Senate Committee on Homeland
Security and Government Affairs
340 Dirksen Senate Office Building
Washington, D.C. 20510

Dear Chairman Peters,

Thank you for your questions for the record from the September 14, 2022 hearing titled "Social Media's Impact on Homeland Security". Per your request, attached are TikTok's answers for the record.

Sincerely,



Vanessa Pappas
Chief Operating Officer, TikTok

TikTok Inc.
5800 Bristol Pkwy, Suite 100
Culver City, CA 90230

TikTok

Post-Hearing Questions for the Record
Submitted to Vanessa Pappas
From Senator Gary C. Peters

"Social Media's Impact on Homeland Security"
September 14, 2022

1. Please provide the following information regarding TikTok's current employees as of September 1, 2022: (i) the total number of full-time engineers at your company, (ii) the total number of full-time engineers working full time on ensuring trust and safety or integrity of your platform, and (iii) the total number of full-time engineers working full time on product development.

More than 40,000 people globally work exclusively on trust and safety issues, which includes in-house and contract moderators as well as teams focused on safety policy, product, and operations.

Approximately 4,900 full-time engineers work for TikTok. Approximately 400 engineers focus on trust and safety. Approximately 2,000 engineers focus on product development.

2. Please provide all metrics evaluated in your A/B testing for each of the last 2 years, including a description of each metric and the percentage of A/B tests in which the metric was evaluated.

TikTok takes a multi-faceted approach to product development and launch. One component is A/B tests, which are often used by TikTok in an effort to ensure that product features are performing satisfactorily. There is no single set of metrics used to evaluate potential product features; the team designing an A/B test can select commonly-used metrics that the system offers, as well as add their own metrics, depending on what the product or feature is and what the team is testing.

In addition to A/B testing, TikTok may use phased roll-outs to smaller groups of users in order to evaluate how a product launch is performing. TikTok may also delay launches of products if a product is not performing as expected or if there is a need for enhanced moderation capabilities.

Some of the metrics used in A/B tests may include:

- (i) content safety metrics, such as the number of reported videos and the number of videos found to be in violation of our Community Guidelines;
- (ii) user interaction metrics, such as likes, comments, shares, and follows;
- (iii) feature-specific metrics, such as LIVE, search, or effect;

TikTok

- (iv) performance metrics, such as the time it takes to respond to a request and the time it takes to show the first For You Feed; and

- (v) user metrics, such as user numbers, user retention, and view viewed.

3. Does TikTok use metrics to measure whether certain demographics see a disproportionate amount of harmful content on your platform, such as, for example, the number of users who have repeated exposures to harmful content, or the amount of harmful content seen by the user in the 99th percentile of exposure to harmful content? If yes, please share a list of these metrics and their current values and values from last year.

TikTok does not maintain such information.

4. What is the total number of views that violative content has received across your platform over the last year, month, and day? Please break down by category of violative content and platform.

Internally, we look at Community Guidelines Violation Rates (CGVR) which is an estimate of the proportion of views on videos that violate our Community Guidelines. To derive this estimate, moderators review a sample of videos and tell us the percentage that were violative.

Please see attached file.

5. Are trust and safety metrics used to determine compensation of all product development employees in their individual goals? If no, why not? If yes, please provide which metrics are used and how they affect compensation.

During the performance cycle, an employee is evaluated based on a mix of his/her own self-evaluation against his/her key outputs, 360-degree reviews against his/her key outputs, and how he/she demonstrates our company values and leadership principles. Building a trusted and safe platform is core to the overall goals of the company, so this too factors into performance evaluations.

6. Are growth, engagement, or revenue metrics used to determine compensation of all product development employees in their individual goals? If yes, please provide which metrics are used and how specifically they affect compensation.

TikTok

During the performance cycle, an employee is evaluated based on a mix of his/her own self-evaluation against his/her key outputs, 360-degree reviews against his/her key outputs, and how he/she demonstrates our company values and leadership principles.

7. How much has TikTok spent in the last year on trust and safety?

In 2021, TikTok spent approximately $1 billion on trust and safety.

8. How much has TikTok spent in the last year on product development?

In 2021, TikTok spent approximately $460 million on research and development.

9. Please provide the number of employees who specifically research extreme content on your platform and whether that research includes the extent to which your recommendation algorithms promote extreme content.

TikTok has approximately 50 employees globally who work directly on Violent Extremism and Hateful Behavior-related content, policies, and research. That does not include moderators and other cross-functional staff who work to enforce our policies on the platform. As part of effective enforcement, TikTok uses feedback loops, whereby Violent Extremism and Hateful Behavior team members are informed of emerging trends of concern by content moderators, and then use that information to improve both enforcement and development of new policies. TikTok also makes investments in product development to enhance our auto-detection capabilities for violative content. Additionally, TikTok seeks feedback on our policies from our Content Advisory Council and works with third-party threat intel vendors, focused on tracking possible extremist content, both on platform and off.

10. What restrictions, if any, does TikTok place on the type of research employees can perform or the type of data they can access?

TikTok has established internal protocols and processes for the review and approval of TikTok user research, which, among other things, imposes privacy safeguards to minimize data collection, limit data access, and retain data for only as long as necessary. Relevant teams are trained to adhere to the protocols and processes.

11. What is the takedown rate of (i) hate speech, and (ii) all violative content over the last year, month, and day?

TikTok's Community Guidelines establish a set of norms and common code of conduct aimed at creating a safe and welcoming platform. (https://www.tiktok.com/community-

guidelines?lang=en). Our Hateful Behavior policies prohibit content that contains hate speech or involves hateful behavior, and such content is removed from our platform when we discover it. The tables below reflect the video removal rate for both (i) hate speech violations and (ii) all content violative of the Community Guidelines for the specified time period.

10/1/21-09/30/22		
Total Uploads	Hate Speech Violations	All CG Violations
Takedown Rates:	**0.005%**	**0.908%**

09/1/22-09/30/22		
Total Uploads	Hate Speech Violations	All CG Violations
Takedown Rates:	**0.004%**	**0.772%**

09/30/2022		
Total Uploads	Hate Speech Violations	All CG Violations
Takedown Rates:	**0.004%**	**0.700%**

TikTok

Post-Hearing Questions for the Record
Submitted to Vanessa Pappas
From Senator Thomas R. Carper

"Social Media's Impact on Homeland Security"
September 14, 2022

1. Social media has fundamentally transformed the way we stay in touch with loved ones, create connections, and the way businesses large and small reach customers around the world. Unfortunately, it is also used to recruit, influence, and mobilize individuals to commit violent attacks.

   The rate at which information is shared using social media has significantly impacted the speed at which individuals may be radicalized and inspired to violence, which can narrow the window of opportunity for law enforcement to stop them before they commit violent attacks.

   a. What information does your platform proactively share with law enforcement to prevent or flag the spread of violent or hateful content on social media before an act of violence is committed? What more needs to be done?

TikTok is committed to cooperating with law enforcement while respecting the privacy and other rights of its users. To achieve this, TikTok has externally facing Law Enforcement Guidelines (https://www.tiktok.com/legal/law-enforcement) as well as internal policies and procedures governing how TikTok handles and responds to these law enforcement requests.

TikTok has a dedicated Law Enforcement Response Team ("LERT") that handles all reports from law enforcement agencies and officers. These teams are staffed by experienced professionals, including individuals with prior law enforcement experience.

When we have a good faith belief that an emergency can be prevented through disclosure of data, we disclose user data necessary to provide law enforcement with 1) content of the threat or emergency issue (e.g., video, comment, direct message, etc.) and 2) limited information to locate and/or identify the potential perpetrator or victim as applicable to the situation (e.g. IPs, registration information).

TikTok knows that our work is never done and that there is no finish line when it comes to safety on our platform. We continuously review our policies and our features and consult with experts like our US Content Advisory Council, academics, and NGOs to consider how we can enable creative expression while protecting against potential harms.

TikTok

Post-Hearing Questions for the Record
Submitted to Ms. Vanessa Pappas
From Senator Kyrsten Sinema

"Social Media's Impact on Homeland Security"
September 2022

Cartel-Sponsored Content

In my capacity as Chair of the Border Management Subcommittee, I asked TikTok, Meta, Twitter, and YouTube to each commit to sharing cartel recruitment content with the Department of Homeland Security (DHS) as quickly as possible. You responded: "yes . . . following our legal process and privacy policy." I appreciate TikTok's commitment to work with DHS to help protect Arizonan youth from being targeted by cartels.

- Please elaborate. Are there any circumstances where known and identified cartel recruitment content would not be shared with DHS due to your legal process or privacy policy? If so, please describe in detail under what, if any, circumstances cartel-sponsored content on TikTok that seeks to recruit American minors would not be shared with DHS.
- It was further stated at the hearing that "when reports [of cartel-sponsored content] have been sent [to TikTok], that content is immediately taken down." Over the last two years, how many such reports have you received and how many responsive posts have you removed?
- Will you commit to engage with the Border Patrol and the sheriff's department in each county along the Arizona-Mexico border regarding these concerns and provide each of them with a designated point-of-contact at TikTok?

Our Law Enforcement Response Team (LERT) routinely discloses relevant user data in response to valid legal requests from law enforcement agencies, including DHS. TikTok may share content or account information directly with law enforcement in the absence of a request when it believes in good faith that there is an emergency involving imminent harm or risk of death or serious physical injury to a person.

TikTok uses a combination of machine-based moderation tools and human moderators to detect and remove content that violates our Community Guidelines, and users and third parties, including law enforcement, may also report content that they believe violates the Guidelines. From October 1, 2021-Sept 30, 2022, TikTok removed through our different detection methods 2,467 pieces of content in the US relating to violations of our human exploitation and human smuggling policies.

TikTok has a robust law enforcement outreach team that is dedicated to meaningful engagement with law enforcement officers across the federal, state, and local levels. Since May 2021, this team has delivered its outreach training presentation to approximately 6,000 officers across the United States and Canada. The team also regularly hosts booths and makes presentations at nationwide conferences, such as the recent annual Dallas Crimes Against Children Conference, which is attended by thousands of law enforcement officials. As part of the efforts of this team,

TikTok

we would be pleased to engage with United States Border Patrol as well as local Sheriff's Departments in jurisdictions along the Arizona-Mexico border.

Biometrics

More than 100 million Americans – nearly one-third of our nation's population – use TikTok. In the course of using your service, they have the right to understand how their biometric information is being collected, processed, and stored. Given TikTok's links to ByteDance, a China-based company, and publicly-disclosed efforts of the Chinese government to collect personal information from Americans, the United States itself also has a critical national security interest in better understanding this issue. These questions are an opportunity for TikTok to provide the American people and government with much-needed clarity on TikTok's practices relating to biometric information.

In 2020, class action litigation was initiated against TikTok that alleged, among other claims, TikTok was violating Illinois' Biometric Information Privacy Act by collecting users' biometric information without authorization. In the course of this litigation, attorneys representing TikTok made the following statement in a federal court filing dated June 5, 2020 (emphasis added):

> *"The filters . . . allow users to add visual effects to videos (e.g., by altering a video's background or by changing a user's eye color).* ***Such filters are possible because the App tracks the location of users' facial features based on facial landmarking****, not any biometric identifying scan of users' faces. In other words, it marks the location of a user's eyes, nose, or mouth in a video frame so that the filter can be properly positioned. TikTok does not collect this landmarking and filter data from users' devices. It is stored locally on a user's device."*[1]

The crux of TikTok's argument in this filing was that "*such landmarking data is not unique to any user, nor does the filter specifically identify any individual user, and thus does not constitute a "****biometric indicator****" or "****biometric information****" as defined by [Illinois' Biometric Information Privacy Act]*" (emphasis added).[2] The Plaintiffs in the case disagreed with TikTok's assertion.

The case was settled without a final ruling from the court on this question. However, on June 2, 2021 (approximately one year after the date of the above statement), TikTok updated its U.S. Privacy Policy to read as follows (emphasis added):

> "*We may collect* ***biometric identifiers*** *and* ***biometric information*** *as defined under US laws,* ***such as faceprints and voiceprints****, from your User Content. Where required by law, we will seek any required permissions from you prior to any such collection."*[3]

---

[1] "Defendant TikTok's Statement in Support of Final Approval of Class Settlement and Opposition to Motion to Intervene", United States District Court for the Northern District of Illinois, *T.K. et. al. v. ByteDance Technology Co. Ltd. et. al.*, June 5, 2020, available at: https://www.courtlistener.com/docket/16540316/34/tk-v-bytedance-technology-co-ltd/.

[2] Id.

[3] Privacy Policy, TikTok, available at: https://www.tiktok.com/legal/privacy-policy.

TikTok

The statements made by TikTok in court are challenging to reconcile with the statements made by TikTok in its U.S. Privacy Policy. In court, TikTok unequivocally states that it does not collect information that "constitute[s] a biometric indicator or biometric information," but in its updated Privacy Policy, TikTok explicitly states "we may collect biometric identifiers and biometric information."

Meanwhile, at the recent hearing before the Senate Homeland Security and Governmental Affairs Committee, you provided the following testimony in response to my questioning regarding the possibility that TikTok could collect biometric information on Americans and share such information with PRC authorities:

> "*I think biometrics is . . . a topic that is hard to define and everybody has their own definition of what biometrics means. So I will be clear in how TikTok sees this. We do not use any sort of facial voice or audio . . . or . . . body recognition that would identify an individual. So there's no way that we would be able to identify. The way that we use facial recognition, for example, would be is if we're putting an effect on the creator's video. So you were uploading a video and you wanted to put sunglasses or dog ears on your video, that's when we do facial recognition. All of that information is stored only in your device. And as soon as it's applied to - like that filter is supplied and posted, that data is deleted. So we don't have that data.*"[4]

While your answer partially addresses how TikTok purports to *use* biometric data, we know for a fact that TikTok and other similar social media services possess users' sensitive biometric information. For instance, in a recent American Civil Liberties Union blog post entitled "New Trends May Help TikTok Collect Your Personal, Unchangeable Biometric Identifiers," the ACLU highlighted the trending "Euphoria Effect" on TikTok.[5] In essence, users put the lens of their smartphone immediately in front of their eyes and then TikTok's applies a high-resolution filter to highlight the details of their irises with bright colors and effects. According to *Biometric Update*, "more than 700,000 videos of iris closeups have been posted to TikTok."

Irises are immutable biometric characteristics, so users who have posted closeups of their irises on TikTok or other social media platforms may find themselves as victims of identity theft as iris recognition becomes a more common form of identification, or even more disturbingly, tracked by those with extensive surveillance capabilities in public places. Young users of social media applications are surely not aware of these risks when they perform the "Euphoria Challenge" or apply other similar effects. Although the iris closeups uploaded by users are available to anymore who visits TikTok's app or website, because users submit their real names, dates of birth, and phone numbers/email addresses when they sign up for TikTok, your company has information that uniquely links the publicly-available iris image to the frequently-private actual identity of the user.

---

[4] Social Media's Impact on Homeland Security, Hearing before the Senate Committee on Homeland Security and Governmental Affairs, 117th Congress, Testimony of Ms. Vanessa Pappas in Response to Questioning of Senator Kyrsten Sinema.

[5] "New Trends May Help TikTok Collect Your Personal, Unchangeable Biometric Identifiers", American Civil Liberties Union, Apr. 14, 2022, available at: https://www.aclu.org/news/privacy-technology/new-trends-may-help-tiktok-collect-your-personal-unchangeable-biometric-identifiers.

TikTok

In light of the above, please provide answers to the following:

- As you stated in your Senate testimony, "everybody has their own definition of what biometrics means." However, despite the acknowledged ambiguity of this term, TikTok's Privacy Policy does not include a definition of what constitutes a "biometric identifier" or "biometric information" that it may collect from users, except to note that this may include faceprints and voiceprints.
    - Please provide precise definitions for how TikTok defines the following terms for purposes of its Privacy Policy and otherwise: "biometric identifier" and "biometric information".
    - Please provide precise definitions for how TikTok defines the following terms for purposes of its Privacy Policy and otherwise: "faceprint", and "voiceprint".

As stated in our Privacy Policy, we use these terms as they are used under applicable laws, which in some cases have been interpreted broadly by litigants and others without necessary guidance from the courts.[6]

- Why did TikTok update its U.S. Privacy Policy in 2021 to include language relating to biometrics?
    - Why does similar language relating to biometrics, including faceprints and voiceprints, not appear in TikTok's European or "Other Regions" Privacy Policies?
    - Many other major social media platforms – including platforms that provide users with the ability to upload short-form videos and apply effects to those videos – have also faced lawsuits over biometric privacy practices. However, upon examination of certain competitors' privacy policies, there does not appear to be language regarding faceprints or voiceprints. Why does TikTok believe it needs to include language in its Privacy Policy that competitors offering similar services apparently do not feel compelled to include in their privacy policies?

TikTok periodically updates its Privacy Policy to, among other things, provide clarifications and address evolving legal standards and interpretations. TikTok's U.S. Privacy Policy differs from its Privacy Policies for other regions due to region-specific legal requirements and considerations. For example, TikTok's Privacy Policy in Europe has a section on legal bases to address legal requirements under the GDPR, whereas the U.S. Privacy Policy does not. TikTok is not in a position to comment on other companies' practices or why other platforms decide to include or not include particular disclosures in their privacy policies.

- In your Senate testimony, you stated "we do not use any sort of facial, voice, or audio, or . . . body recognition *that would identify an individual*" (emphasis added).

[6] TikTok's responses pertain to U.S. Privacy Policy and U.S. users.

TikTok

- What forms of facial, voice, audio, body, or other biometric recognition, collection, or processing does TikTok use, regardless of whether such practice would identify an individual?
  - Please specifically describe each practice and the contexts in which it is applied.
  - Some permitted practices for collected data pursuant to your Privacy Policy include "inform[ing] our algorithm," "infer[ing] additional information about you," and "prov[ing] your identity in order to use certain features . . . where verification may be required." The Privacy Policy also notes "aggregated or de-identified data is not subject to this Privacy Policy." Does TikTok use biometric information or provide it to any third party pursuant to the language quoted in this paragraph, regardless of whether it is aggregated or de-identified?
- What was meant by the term "body recognition" and does TikTok use such technology in any manner?

We use image and voice information for various non-identifying purposes, for example, the location of facial features within an image (e.g., to detect the location of a user's eyes for visual effects) or aspects of the audio (e.g., to alter the tone for voice effects). To confirm, we do not use any image or voice information to identify individuals. The reference in our Privacy Policy to aggregated or de-identified data is a general statement and not specific to this type of processing activity.

- In the aforementioned 2020 court filing, TikTok stated: "*The filters . . . allow users to add visual effects to videos (e.g., by altering a video's background or by changing a user's eye color). Such filters are possible because the App tracks the location of users' facial features based on facial landmarking, not any biometric identifying scan of users' faces. In other words, it marks the location of a user's eyes, nose, or mouth in a video frame so that the filter can be properly positioned. TikTok does not collect this landmarking and filter data from users' devices. It is stored locally on a user's device.*" This filing further stated *"such landmarking data is not unique to any user, nor does the filter specifically identify any individual user, and thus does not constitute a biometric indicator or biometric information*."
  - Does TikTok stand by each assertion relating to how it handles, processes, and stores biometric information contained in "Defendant TikTok's Statement of Support of Final Approval of Class Settlement and Opposition to Intervene" dated June 5, 2020?
    - If not, in what respect was such information inaccurate or how have TikTok's practices changed since such date?
    - Will TikTok restate all such representations with respect to "effects" as opposed to just "filters"?
  - In the course of conducting facial landmarking (or such other biometric process as TikTok may employ), what data is collected? In particular, how many facial "landmark points" are identified? Given that the spacing between each person's facial landmark points is unique and distinct, how do you reconcile this with your

statement before the Senate that TikTok does not employ "facial . . . recognition that would identify an individual"?

We do not comment on litigation matters. We do not use facial landmarking information to identify individuals. This information is processed on the user's device only and deleted from the device once the effect or filter is applied. We note that facial landmarking technology is widely used.

- This filing only discusses how TikTok purports to handle facial data. There has not been similar disclosure relating to how TikTok processes vocal data or the "voiceprints" mentioned in your Privacy Policy.
    - Does TikTok in fact collect voiceprints or other vocal data? If not, why is this term included in your Privacy Policy? If so, how is this data collected, processed, and secured?

Please see TikTok's responses to your prior questions above.

- At the hearing, I asked whether there is an "opportunity . . . during the time between the use of the . . . faceprint or voiceprint and [its] deletion . . . for anyone other than that device to access or capture that information?" You responded that "it's a technical area. . . [but] to the best of my knowledge, the data is stored on the devices, [and] deleted immediately once you post your video."
    - Were the substantive assertions in this testimony accurate in all respects? Is all biometric processing in fact performed locally on a user's device and all such data deleted immediately upon upload of a video?
        - If not, please describe how TikTok's actual practices differ from those described in the testimony.
    - Is it in fact technically impossible for this data to be accessed by anyone except someone with physical access to the user's device?
        - If so, please specifically elaborate, including with respect to why TikTok feels confident that there is no opportunity for TikTok or any other entity to obtain biometric information (including, but not limited to, facial landmarking data) from users applying filters or effects. If not, please elaborate.

This confirms that the hearing testimony was accurate in that the data is processed locally on the user's device and deleted from the device as soon as the effect is applied.

- Please describe any algorithms that TikTok uses to perform facial landmarking or other biometric functions (including any processing of faceprints or voiceprints).
    - Who designed each responsive algorithm, any they located in China, and has each such algorithm been submitted for vetting by an independent third party?

TikTok

- Some filters and/or effects are developed by third parties. Does TikTok vet third party filters and/or effects, including those developed using Effect House, from a biometric privacy perspective? If so, how?

Please see TikTok's responses to your previous questions above. All effects, including those created in Effect House, are based on a limited set of algorithms, such as the face landmarking algorithm. These algorithms undergo a legal review process to ensure compliance with applicable laws.

Algorithms in Effect House are a subset of TikTok's special effect algorithms, such as facial landmarking. As such, data processed in connection with the use of an Effect House algorithm is processed locally on the user's device and deleted from the device once the effect is applied.

- Please describe TikTok's stance on users applying any filter or effect, or otherwise uploading any content, that involves capturing closeup images of an individual's eyes generally or irises specifically.
  - Has TikTok deliberately promoted any filter, effect, or trend that involves users capturing closeup images of their irises (including, but not limited to, the "Euphoria" effect)? If so, why did TikTok choose to promote such filter, effect, or trend?
  - What technological process allows users to utilize such effects or filters? What, if any, data is collected or processed by TikTok or any third party in the process of enhancing or altering the image of an individual's eyes?
  - Is TikTok aware, or does TikTok suspect, that any third party has collected or attempted to collect images of users' irises from user-submitted content?
  - Has TikTok ever considered providing users with a warning regarding potential privacy implications before allowing users to upload closeups of their irises to the web?

The "Euphoria" effect is a low-exposure photographic filter that can be used on any image, including but not limited to close-up images of irises. We do not extract information from these images, nor do we promote filters, effects, or trends that encourage users to capture closeup images of their irises. We are not aware of any attempts by third parties to extract information of users' irises from their content posted on TikTok.

Content Moderation or Promotion

Another line of questioning I pursued at the hearing related to whether "TikTok ***ever*** altered its algorithm – or promoted or down-ranked content – based on the actual or perceived wishes of the Chinese government." You responded with an unequivocal "no." While I appreciate the clear and appropriate answer, I believe the American people would benefit from further questioning on this matter.

TikTok

- At a hearing before the British Parliament in 2020, a British TikTok executive stated: "In the early days of TikTok there [were] some policies in place that took what we call a 'blunt instrument' to the way in which content was censored. . . . At that time we took a decision . . . to not allow conflict on the platform, and so there was some incidents where content was not allowed on the platform, specifically with regard to the Uyghur situation." (sic).
    - Did this in fact occur on the U.K. version of TikTok? Additionally, were similar decisions ever made – without regard to whether such decisions were made at the request a foreign government – on the U.S. version of TikTok, particularly with respect to content discussing Xinjiang, other topics relating to China, or matters that the CCP deems sensitive? If so, when did this occur and why?

During the British Parliament hearing referenced in your question, TikTok's representative conflated two separate issues in her response and subsequently explained that she made an incorrect statement.

The first issue relates to TikTok's previous and outdated approach to content moderation. TikTok has previously acknowledged that in our very early days, we took a blunt approach to moderating content that promoted conflict, but we've also said we recognized this was the wrong approach and eliminated it. Even in those early policies, there was never a policy around the Uighur community.

The second issue is mismoderation of one piece of content that related to the Uighur community, which is explained in detail in a blogpost from November 2019 (https://newsroom.tiktok.com/en-us/an-update-on-recent-content-and-account-questions). This incorrect removal was the result of a human error; nothing in our Community Guidelines precluded the content in that video and it should not have been removed. We documented that the human error brought the video offline for 50 minutes before it was reinstated by a senior member of our moderation team who identified and overturned the error.

- In a 2019 article in *The Guardian*, it was reported that some content related to China on TikTok was marked as a "violation", leading to its deletion, and other China-related content was marked as "visible to self", limiting its distribution through TikTok's feed. For instance, it was reported that there were bans on "criticism/attack towards policies, social rules of any country, such as constitutional monarchy, monarchy, parliamentary system, separation of powers, **socialism system**, etc" (sic) (emphasis added) and "demonisation or distortion of local or other countries' history such as May 1998 riots of Indonesia, Cambodian genocide, **Tiananmen Square incidents**" (sic) (emphasis added).
    - Are these reports accurate and were such guidelines ever applied in the U.S.?

In TikTok's early days we took a blunt approach to minimizing conflict on the platform, and our moderation guidelines allowed penalties to be given for things like content that promotes conflict between religious sects or ethnic groups, spanning a number of regions around the world. As TikTok started taking off in new markets, we recognized that this was not the correct approach and began working to empower local teams that have a nuanced understanding of each market.

TikTok has a dedicated US Safety team that is responsible for US content and moderation policies. The old guidelines referenced in these 2019 Guardian articles are long outdated.

- In addition to TikTok, ByteDance used to operate a news app in the U.S. known as TopBuzz. According to a recent article, ByteDance instructed members of its staff to place specific pieces of pro-China messaging in the now-defunct TopBuzz app and "pin" certain China-related articles at the top of the app's feed. I understand ByteDance denies this claim.
    - Has ByteDance ever requested that TikTok promote or down-rank content relating to China, topics the CCP deems sensitive, or domestic U.S. political matters on TikTok, without regard to whether such requests were made at the request of a foreign government?

Our Community Guidelines apply equally to everyone and all content on TikTok, and we do not moderate or remove content based on political sensitivities. Furthermore, TikTok has a dedicated US Safety team that is responsible for US content and moderation policies. TikTok is transparent about the requests we receive from governments to restrict content and how TikTok responds and regularly publishes reports on disclosures (https://www.tiktok.com/transparency/en-us/government-removal-requests-2021-2/).

TikTok is led by its Singapore-based CEO and has full autonomy over its content moderation policies, and only TikTok's trust and safety teams can moderate TikTok content. Any attempt to go around these controls would be a gross violation of moderation procedures.

- Given the scrutiny TikTok faces on all matters relating to China or domestic U.S. political matters, I presume that any content moderation decisions involving China-related or U.S. election-related posts (such as a video that is critical of China or critical of a U.S. presidential candidate, but simultaneously violates one of your terms of service) are elevated for additional review prior to any action being taken.
    - What special procedures does TikTok have in place to deal with content moderation decisions involving China-related or U.S. election-related posts and who makes the ultimate decision regarding whether to remove the post?

TikTok has a dedicated team of policy, safety, security, and operations experts working to protect the integrity of the US midterm elections. Members of our US Safety team, which reports into TikTok COO Vanessa Pappas, are responsible for reviewing and removing election misinformation and other violations of our policies. Further, and as noted above, TikTok's Community Guidelines and election misinformation policies apply equally to everyone and all content on TikTok. We do not moderate or remove content based on political sensitivities. We do not have special policies or procedures in place for content related to elections in China.

- Going forward, I understand that Oracle will be conducting regular audits of TikTok's content moderation processes and its algorithms.

- What will these audits specifically entail? If Oracle identifies an issue, will TikTok be required to address it under the agreement?

Oracle is not performing these functions today and we have nothing to share at this time beyond our written testimony in relation to future plans.

Deepfakes

Deepfakes – AI-generated videos that make people appear to do or say things they didn't actually do or say – present serious homeland security risks. In fact, in other countries such as Gabon and Malaysia, we've already seen fake videos of leaders trigger political turmoil. In Ukraine, there was a deepfake of President Zelensky supposedly telling Ukrainians to lay down their arms against Russia. While platforms have increasingly adopted policies to ban many deepfakes, these policies only can be enforced if you know what is or is not real.

- What is TikTok's policy regarding the distribution of deepfake content on its platform?
- What is TikTok doing to improve deepfake detection? And in cases where deepfakes are too advanced to be detected, how do you go about determining a video's authenticity?

Regardless of whether content is synthetic or not, if content violates any one of our policies, we remove it from our platform when discovered. TikTok prohibits the use of synthetic media like deepfakes and cheapfakes in cases where they could mislead a person about the truth of an event and when they can be reasonably expected to cause harm to an individual or to society.

TikTok provides our moderators with the latest guidance on how to detect potential deepfakes as new research on the methods of creating that content is published. Several external vendors provide us with tips on deepfake content existing off-platform, so we can proactively search and remove any content we identify on platform. We are also sourcing vendors who can provide quick advanced detection and looking at new technology to help authenticate non-synthetic media. We have also worked with outside experts, such as our Content Advisory Council, to seek feedback on our deepfake policies.

In cases where we suspect but cannot prove that a piece of content is a deepfake, we err on the side of treating it as a deepfake and we would remove the content if we discern any harm to an individual or to society.

TikTok

Post-Hearing Questions for the Record
Submitted to Vanessa Pappas
From Senator Alex Padilla

"Social Media's Impact on Homeland Security"
September 14, 2022

1. TikTok users span the globe, and it makes the same commitments to Western English-speaking communities as it does to non-Western, non-English language speaking communities. There has been strong congressional interest in ensuring technology companies are investing in protecting every user, not just English language natives, or those from Western countries.

    a. What are the top five non-English languages for content that users within the United States encounter on TikTok and what percentage of users in the United States encounter content within each of those languages on a weekly basis?

    b. For each of these questions, please break down by employment status (full time in-house employees or contract-level).

        i. How many employees focus exclusively on trust and safety issues at TikTok?

        ii. Of the people working exclusively on trust and safety issues, how many focus on the content moderation of non-English language content, how many of them focus on each of those top five non-English languages, and how many of them focus on non-United States users?

        iii. Of the people working exclusively on trust and safety issues, how many of them focus on non-English language content in the United States? Of the people working exclusively on trust and safety issues, how many of them focus on non-English language content for users outside of the United States?

        iv. How many of them focus on each of those top five non-English languages?

        v. How many of them focus primarily on United States-based users, and how many of them focus primarily on non-United States based users?

    c. When making community standards enforcement decisions, is all content reviewed in the original language or are some or all subject to automated translation before being reviewed? What percentage of total content reviewed is subject to automated translation?

TikTok

d. Are TikTok's community standards available in the native language of all of your users?

e. What is TikTok doing to ensure that it has cultural competency for all the communities it serves?

TikTok is a community-powered entertainment platform. We believe people should be able to express themselves creatively and be entertained in a safe, secure, and welcoming environment. We're focused on building responsibly, equitably, and transparently for the long-term.

Approximately 12,700 employees globally focus exclusively on trust and safety issues for TikTok, which includes teams focused on safety policy, product, and operations, as well as in-house moderators. In addition, TikTok has approximately 30,000 contract moderators globally.

TikTok moderates content in more than 70 languages using a combination of people and technology. TikTok continues to invest in growing our team of multilingual content moderators, policy experts, and operations specialists to support our multilingual community. TikTok has several hundred content moderators who support Spanish language moderation, and we will continue to grow that support.

TikTok uses a combination of people and technology to enforce our Community Guidelines, which are available in 40 languages. We are constantly looking to improve the availability and accessibility of our Community Guidelines, which includes additional language support. To enforce these Community Guidelines effectively at scale, we continue to invest in technology-based flagging and moderation. We rely on automated moderation when our systems have a high degree of confidence that content is violative so that we can expeditiously remove violations of our policies. As a result, our overall protective detection efforts have improved, as is shown in our quarterly Community Guidelines Enforcement Reports that are available publicly on our website: https://www.tiktok.com/transparency/en-us/community-guidelines-enforcement/.

At TikTok, we prioritize safety, diversity, inclusion, and authenticity. We prize the global nature of our community and strive to take into account the breadth of cultural norms where we operate. In support of these efforts, our trust and safety teams partner with local experts, regional Content and Safety Advisory Councils, and civil society organizations to understand the unique cultures and experiences of communities.

2. In your testimony you say that 88.4% of removals under TikTok's violent extremism policy occurred within 24 hours of being posted.

   a. Where did that that 88.4% of content originate?

   b. What was the median amount of time it took to find and remove the remaining 11.6%? What was the global distribution of that 11.6% of content?

TikTok

c. What metrics does TikTok use to measure success around enforcement of its community standards?

d. How does TikTok assess its false positive rate—content you actioned for perceived violation of community standards, but that didn't actually violate the rules—and what is that rate? Please break down based on the top five languages of your users.

e. Internally, what metric do you use to evaluate the success of enforcement of your community standards?

TikTok publishes Community Guidelines Enforcement Reports (https://www.tiktok.com/transparency/en-us/reports/), which provide quarterly insights into the volume and nature of content and accounts removed from our platform. The 88.4 percent references the amount of content globally that was removed within 24 hours for violent extremism for 2022 Q1. TikTok also discloses the proactive removal rate, which means identifying and removing a video before it's reported, as well as the removal rate before any views. TikTok works swiftly to remove violative content, and users can report content using our in-app reporting tools.

TT employs multiple metrics to measure the success of our enforcement efforts, many of which are disclosed in our Community Guidelines Enforcement Reports. Over time, we have made improvements on proactive removals, removals at zero video views, and removals in under 24 hours. Internally, we look at Community Guidelines Violation Rates (CGVR) which is an estimate of the proportion of views on videos that violate our Community Guidelines. To derive this estimate, moderators review a sample of videos and tell us the percentage that were violative.

TikTok's reports provide additional insights about our efforts; for example, "Total videos removed/restored, by type and quarter" of our most recent quarterly report reflects videos initially removed but later restored to visibility on the platform (https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2022-2/). Additionally, TikTok has found that the false positive rate for automated removals is approximately 5% and requests to appeal a video's removal have remained consistent (https://newsroom.tiktok.com/en-us/advancing-our-approach-to-user-safety).

3. In August, it was reported that when a user opens up a link to a website in TikTok's iOS app, the in-app browser that is used tracks user's behavior on those external websites. In your testimony you mention that TikTok "solely uses this information for debugging, troubleshooting, and performance monitoring." The company is asking consumers to trust that you're not collecting sensitive information.

   a. How come nearly every other major iOS app is able to open up links without injecting tracking code but TikTok cannot?

TikTok

b. Is TikTok considering opening up links in its iOS app in the user's default browser?

While we have been clear that we have never collected keystroke or text content through the JavaScript code in our in-app browser, we recently removed the code in question to eliminate any doubts about the information we collect.

TikTok currently offers users the feature to open links in advertisements in the user's default browser and is considering expanding this feature to other links.

4. Election integrity is more important than ever. Platforms have a crucial role to play in ensuring that voters are able to access good, accurate election information, and in maintaining the basic integrity of our elections system. TikTok has made announcements about what it is doing around the midterm elections, though much of your strategy for the midterms is similar to steps taken two years ago.

   a. Have you studied how effective your efforts were during the last election to ensure that your users knew where, how, and when to vote, and if so, what were the results of those studies?

   b. What do you assess to be the biggest threats to election integrity on TikTok in advance of the 2022 elections? What do you assess to be the biggest threats to election integrity on TikTok in advance of the 2024 elections? What steps are you taking to mitigate these identified threats?

   c. How are you staying ahead of the evolving threats to election integrity?

   d. What is TikTok doing to keep election officials safe from doxxing and other online harassment campaigns?

Following the 2020 elections, TikTok conducted an after action case study, with an overview available at: https://newsroom.tiktok.com/en-us/tiktoks-h-2-2020-transparency-report. We regularly look to learn lessons from our experiences in elections globally so that we can continually strengthen our approach. Here are some of the lessons learned from our approach that we wanted to improve on in future elections:

What we think worked

1. Our proportionate focus on both foreign and domestic threats to our platform and overall elections integrity during the US 2020 elections was the right approach. We started our elections preparations in 2019 and built defenses based on industry learnings from the US 2016 elections, but we also prepared for more domestic activity based on trends we've observed on how misleading content is created and spread online.

TikTok

2. We made the correct tooling investments that allowed us to quickly and meaningfully reduce the discoverability of disinformation and terms of incitement. We moved to quickly redirect misleading hashtags to our Community Guidelines instead of showing results, such as #sharpiegate #stopthesteal #patriotparty. This approach has also helped us combat QAnon content, though we continually must update our safeguards as content and terminology evolves.

3. Prioritizing faster turnaround times for fact-checking helped us make informed and quick decisions on emerging content.

4. Our investment in building relationships with a range of experts improved our overall approach to platform integrity, from policies to enforcement strategies to product experiences in our app.

What we can improve

1. We will keep improving our systems to proactively detect and flag misleading content for review. For instance, we can immediately detect known disinformation using our disinformation hashbank, and we're working to advance our models so that we can better identify altered versions of known disinformation.

2. We will continue to develop our system that prevents repeat offenders from circumventing our enforcement decisions.

3. More investment is needed to educate creators and brands on disclosure requirements for paid influencer content. TikTok does not allow paid political ads, and that includes content influencers are paid to create, and we expect our community to abide by our policies and FTC guidelines.

4. We were proud of the in-app elections guide we developed with experts, and in the future we would launch it sooner in the elections process.

To protect the integrity of the 2022 US midterm elections, we have assembled a mission control center to centralize information gathering and content assessment for our cross-functional team made up of safety, security, policy, and operations experts. This cross-functional team has been working together for months and consulting with experts, including those at the National Association of Secretaries of State and National Association of State Election Directors. And, learning from 2020, we acted on one of our lessons learned by launching our in-app center 6 weeks earlier than in 2020 to capture more of the midterms conversation over the summer months.

Other efforts that we've taken to support the integrity of the TikTok platform include:

- Providing access to authoritative information is an important part of our overall strategy to counter election misinformation. That's why we rolled out an Elections Center to connect people who engage with election content to authoritative information and sources

TikTok

in more than 45 languages, including English and Spanish. For instance, people can learn how and where to vote through information provided by the National Association of Secretaries of State (NASS) and who and what is on their ballot from Ballotpedia. We're also collaborating with Center for Democracy in Deaf America to provide information on voting as a deaf person; as a student, with help from Campus Vote Project; and as a person with past convictions, with help from Restore Your Vote. For more information on voting as an overseas citizen or service member, you can visit the Federal Voting Assistance Program. As election results are reported, the latest results will be available in our app from the AP.

- We are committed to promoting digital literacy skills and education, and our in-app center will feature videos that encourage our community to think critically about content they see online, as well as information about voting in the election.

- To ensure that our Elections Center is visible and accessible, we add labels to content identified as being related to the 2022 midterm elections as well as content from accounts belonging to governments, politicians, and political parties in the US. These labels allow viewers to click through to our center and get information about the elections in their state. We also provide access on popular elections hashtags, like #elections2022 and #midtermelections, so that anyone searching for that content will be able to easily access the center. At any time, viewers can use our tools to automatically filter out videos with words or hashtags they don't want to see in their For You or Following feeds.

- To enforce our policies, we use a combination of people and technology. We also review content and accounts reported by community members. To bolster our response to emerging threats, TikTok partners with independent intelligence firms and regularly engages with others across the industry, civil society organizations, and other experts.
    - TikTok has launched a specialized queue to evaluate Spanish language mis and disinformation videos produced in the US.
    - In addition, TikTok partners with accredited fact-checking organizations who help assess the accuracy of content in more than 30 languages, and while they do not moderate content on our platform, their assessments provide valuable input which helps us take the appropriate action in line with our policies. Out of an abundance of caution, while content is being fact checked or when content can't be substantiated through fact-checking, it becomes ineligible for recommendation into For You feeds. We also inform viewers of unsubstantiated content and prompt them to reconsider before sharing potential misleading information.

- TikTok does not allow paid political ads, and that includes content influencers are paid to create. We work to educate creators about the responsibilities they have to abide by our Community Guidelines and Advertising policies as well as FTC guidelines. For sponsored content that is allowed, we've introduced a tool that makes it easy for creators to disclose paid relationships with brands and organizations. If we discover political content was paid for and not properly disclosed, it is promptly removed from the platform. Additionally, we recently announced updates for government, politician, and

political party accounts, including the testing of mandatory verification during the midterm elections.

We are constantly looking at how we can improve our practices and processes to protect the safety and integrity of our platform during elections.

TikTok

Post-Hearing Questions for the Record
Submitted to Vanessa Pappas
From Senator Jon Ossoff

"Social Media's Impact on Homeland Security"
September 14, 2022

1. Do either TikTok or its parent company, ByteDance, have the capability to monitor and log every time a TikTok or ByteDance employee accesses user data?

    a. If so, does either TikTok or ByteDance log each event where a TikTok or ByteDance employee accesses TikTok user data? If so, please describe in detail the information recorded in this log.

TikTok's data access approval policy, which sets out TikTok's data access controls and procedures, applies to all TikTok U.S. user data. Steps are taken to limit TikTok U.S. user data access, including restricting who has access to which dataset, implementing strong authentication measures, logging of access, limiting access periods, and encrypting data. To the extent an employee's authorized access to TikTok U.S. user data is logged, typically the information recorded in the log includes the employee's user name, the time and date of such access, and what type of data is accessed.

As has been reported in the press, for more than a year we have been pursuing a multi-pronged initiative called "Project Texas" to strengthen TikTok's U.S. data security program. In May 2022, TikTok announced the creation of a new division—U.S. Data Security ("USDS")—to bring heightened focus and governance to our ongoing efforts to strengthen our data protection policies and protocols, further protect our U.S. users, and build confidence in our systems and controls in the United States. This division has U.S.-based leadership. Access to U.S. user data by anyone outside of our new U.S. Data Security team will be limited by, and subject to, robust data access protocols that are being developed in close collaboration with Oracle and the U.S. government.

2. Please enumerate all types of TikTok user data which a TikTok or ByteDance employee can access (such as, but not limited to, history of use of the platform, browser history, geolocation history, or IP addresses)? Please differentiate levels or types of access by types of employee (seniority, function in the company, etc.).

TikTok employees may at times access certain types of user information that TikTok collects to perform their job function. This access approval is based on the employee's business needs, rather than seniority, and the access is subject to data access policy and procedures. The level of approval required is based on the sensitivity of the data according to the classification system.

TikTok

3. Since 2019, how many instances have there been of a TikTok or ByteDance employee accessing or requesting access to a TikTok user's account data?

    a. Of these instances, how many did TikTok or ByteDance find to be inappropriate or unauthorized based on the employee's job or function?

TikTok requires approvals for access to U.S. user data as described in our response to Question 2 above. Access approval is based on the employee's business needs, rather than job title or seniority. If it is determined that an employee does not need to access U.S. user data for her job responsibilities, the approval request will be denied. Once access is approved, the employee can access U.S. user data subject to the conditions set forth in the approval, which may include a limited time period for such access.

As has been reported in the press, for more than a year we have been pursuing a multi-pronged initiative called "Project Texas" to strengthen TikTok's U.S. data security program. In May 2022, TikTok announced the creation of a new division—U.S. Data Security ("USDS")—to bring heightened focus and governance to our ongoing efforts to strengthen our data protection policies and protocols, further protect our U.S. users, and build confidence in our systems and controls in the United States. This division has U.S.-based leadership. Access to U.S. user data by anyone outside of our new U.S. Data Security team will be limited by, and subject to, robust data access protocols that are being developed in close collaboration with Oracle and the U.S. government. In order to facilitate a global platform, our goal is to ensure non U.S.-based employees, including China-based employees, will only have access to a narrow set of TikTok U.S. user data, such as public videos and comments available to anyone on the TikTok platform, to ensure global interoperability.

4. Have TikTok, ByteDance or any ByteDance subsidiary received any requests from the Chinese Communist Party or any official, agency, or instrumentality of the government of the People's Republic of China to (i) alter or remove any content on TikTok's platform; or (ii) to provide them with any information regarding users of TikTok?

No, as TikTok has previously stated, TikTok has not been asked for user data by the Chinese government or the CCP. We have not provided U.S. user data to the Chinese government or CCP, nor would we if asked. More information about government requests for user data that we receive across the world is available in our Information Request Reports, available at https://www.tiktok.com/transparency/en-us/information-requests-2021-2/.

5. For the years 2020, 2021, and 2022, how many requests have TikTok, ByteDance, or any ByteDance subsidiary received from any foreign or U.S. government entity (including subnational entities of government, including but not limited to states, provinces, counties, or cities) to alter or remove any content on TikTok's platform or to access any information regarding users of TikTok? Please provide a list of each request made to TikTok, ByteDance, or any ByteDance subsidiary, including the name of the requesting entity, a short description of the nature of the request, and the date the request was made.

TikTok

a. For the years 2020, 2021, and 2022, of the requests received, with how many such requests has TikTok complied? Please provide a list of each request with which TikTok, ByteDance, or any ByteDance subsidiary has complied, including the name of the requesting entity, a short description of the nature of the request, and the date of the request.

TikTok publicly discloses information about government requests for user data that we receive across the world in our Information Request Reports, available at https://www.tiktok.com/transparency/en-us/information-requests-2021-2/.

TikTok

Post-Hearing Questions for the Record
Submitted to Ms. Vanessa Pappas
From Senator Rob Portman

"Social Media's Impact on Homeland Security"
September 14, 2022

1. You testified that under no circumstances would TikTok give data to the Chinese Communist Party. Yet, TikTok has employees and offices within China, thus making TikTok subject to China's national security law. Can you confirm with one hundred percent certainty that no data from American users has ever been accessed by the Chinese Communist Party?

As a global entertainment platform, TikTok spans most major markets except China. TikTok is provided in the United States by TikTok Inc., which is incorporated in California and subject to U.S. laws and regulations. Like many global technology companies, we have product development and engineering teams all over the world collaborating to deliver the best product experience for our community.

TikTok has not been asked for U.S. user data by the Chinese government or the CCP. We have not provided such data to the Chinese government or CCP, nor would we if asked. TikTok discloses on a regular basis in its Transparency Reports (https://www.tiktok.com/transparency/en-us/reports/) requests for user information that we receive from governments and law enforcement agencies.

2. In your testimony you mentioned that TikTok is now routing U.S.-based user data to Oracle servers in an alleged attempt to block China's access to this data. Last year, *The Intercept* released a report on Oracle's longstanding ties to the Chinese government and detailed how Oracle's software has been used by China's police, military, and other government entities to surveil its people and commit human rights atrocities. Were you aware of this affiliation between Oracle and the Chinese government when TikTok selected them to store U.S. user data, yes or no?
   a. If yes, then why did you choose Oracle to provide its servers for this data migration?
   b. How can you guarantee that data is not being stored within China – whether due to violations of company policy or is stored via screenshots and other formats?
   c. Oracle, like TikTok, has a presence in China, and is also therefore subject to China's national security laws. How then is it possible for Oracle to effectively safeguard U.S. user data from the Chinese Communist Party?

TikTok considered the major cloud vendors in the U.S. for the national security partnership. We selected Oracle based on their industry-leading technological capability

TikTok

and their demonstrated strength and experience with U.S. government and national security projects (e.g., https://www.oracle.com/news/announcement/dod-accredits-oracle-cloud-infrastructure-for-top-secret-missions-2022-02-15/).

TikTok is committed to implementing significant organizational, process, and technical changes related to the storage, access, and security of U.S. user data pursuant to this arrangement. The data is stored in Oracle Cloud Infrastructure with access limited to authorized personnel and additional access protocols being developed.

3. Oracle shared with Committee staff that any migration of U.S. user data to Oracle servers will not alleviate concerns held regarding the Chinese Communist Party's access to U.S. user data, because data storage does not impact where data may be accessed. Additionally, based on your testimony, TikTok will continue to share data with Chinese-based entities regardless of the outcome of the CFIUS process. Will TikTok commit to sharing with this Committee in a confidential environment all the data and metadata that has been accessed in China, yes or no?

Employees outside the U.S., including China-based employees, who work on TikTok can have access to TikTok U.S. user data subject to a series of robust cybersecurity controls and authorization approval protocols overseen by our U.S.-based security team. In addition, TikTok has an internal data classification system and approval process in place that assigns levels of access based on the data's classification and requires approvals for access to U.S. user data. The level of approval required is based on the sensitivity of the data according to the classification system.

Also, US user data is now stored in Oracle Cloud Infrastructure with access limited to authorized personnel and additional access protocols being developed.

Earlier this year, TikTok announced the creation of a new division—U.S. Data Security ("USDS")—to bring heightened focus and governance to our ongoing efforts to strengthen our data protection policies and protocols, further protect our U.S. users, and build confidence in our systems and controls in the United States. This division has U.S.-based leadership. Access to U.S. user data by anyone outside of our new U.S. Data Security team will be limited by, and subject to, robust data access protocols that are being developed in close collaboration with Oracle and the U.S. government. In order to facilitate a global platform, our goal is to ensure non-U.S.-based employees, including China-based employees, will only have access to a narrow set of TikTok U.S. user data, such as public videos and comments available to anyone on the TikTok platform, to ensure global interoperability.

4. In leaked TikTok PR documents, titled "TikTok Master Messaging" and "TikTok Key Messages," the company directs employees that speak publicly about the company to rely on a list of messages to convey to critics and skeptics of the company. One quote from the documents tells employees to "Downplay the parent company ByteDance, downplay the China association, downplay AI." Do

you agree that these statements confirm that a relationship between TikTok and ByteDance and China exists?

These messaging documents were created to support team members with external communications. We have always emphasized that TikTok is a global platform that does not operate in China.

5. In 2020, a TikTok spokesperson claimed your parent company, ByteDance, is not a Chinese company because it is incorporated in the Cayman Islands. However, the director listed in the Cayman Islands filing document is Liang Rubo, the CEO of ByteDance who oversees company operations in China. Does TikTok still maintain that it is not subject to any influence by a Chinese-based entity, even though we know the CEO of ByteDance is listed as a director on its Cayman Islands shell corporation filing documents, and is also responsible for all Chinse-based operations?

ByteDance Ltd., the ultimate parent entity of TikTok, is a global company, with employees around the world. It is incorporated in the Cayman Islands, and has a global board, including Bill Ford of General Atlantic, Arthur Dantchik of Susquehanna International Group, Philippe Laffont of Coatue, Neil Shen of Sequoia, and the company's co-founder and CEO Rubo Liang.

ByteDance's investors include global institutional funds such as Baillie Gifford, Blackrock, Coatue, Fidelity, General Atlantic, KKR, Sequoia, Softbank, Susquehanna International Group, and T. Rowe Price.

TikTok is led by its own CEO, Shou Chew, a Singaporean based in Singapore. The co-founder and CEO of ByteDance Ltd., Rubo Liang, is based in Singapore while he leads global expansion efforts.

6. Under any circumstances, have TikTok employees at any time taken direction from ByteDance employees?

TikTok is one of several product lines of ByteDance Ltd. The relationship between ByteDance and the entities focused on the TikTok business is typical of the relationship between parent corporations and subsidiaries in other global businesses with many lines of business. At the same time, because TikTok is not available in China, it is led by its own CEO, Shou Chew, a Singaporean based in Singapore, who leads the business.

As has been reported in the press, for more than a year we have been pursuing a multi-pronged initiative called "Project Texas" to strengthen TikTok's U.S. data security program. In May 2022, TikTok announced the creation of a new division—U.S. Data Security ("USDS")—to bring heightened focus and governance to our ongoing efforts to strengthen our data protection policies and protocols, further protect our U.S. users, and

TikTok

build confidence in our systems and controls in the United States. This division has U.S.-based leadership. Access to U.S. user data by anyone outside of our new U.S. Data Security team will be limited by, and subject to, robust data access protocols that are being developed in close collaboration with Oracle and the U.S. government.

7. Is it true that TikTok and ByteDance share the same internal auditor, yes or no?
   a. What data has been made available to TikTok's internal auditors?

Like many large companies, we share some corporate functions across divisions. Internal audits that ensure adherence to ByteDance policy are one such function. There are set processes for our Internal Audit team to be able to acquire the data and information they need to conduct their investigations. These investigations are conducted only to assess specific claims that involve the company or employees, not to undertake open-ended inquiries into users.

The Internal Audit team is responsible for objectively auditing and evaluating the company's and our employees' adherence to our codes of conduct. This team provides recommendations to the leadership team, but does not participate in disciplinary action.

8. Under any circumstances, has U.S. user and consumer data ever been shared or handled on TikTok's internal messaging platforms, yes or no?

As stated above, access to user data is on an as-needed basis subject to authorization approval protocols. In the performing of one's job duties, there may be times where it is necessary to internally communicate with other employees regarding a certain user, for example, in the context of user support, user safety, and legal matters.

9. Do China-based TikTok and ByteDance employees share any office space or technology, yes or no?

Yes. Like many global technology companies, we have teams all over the world collaborating to provide our community a safe and enjoyable experience on our platform.

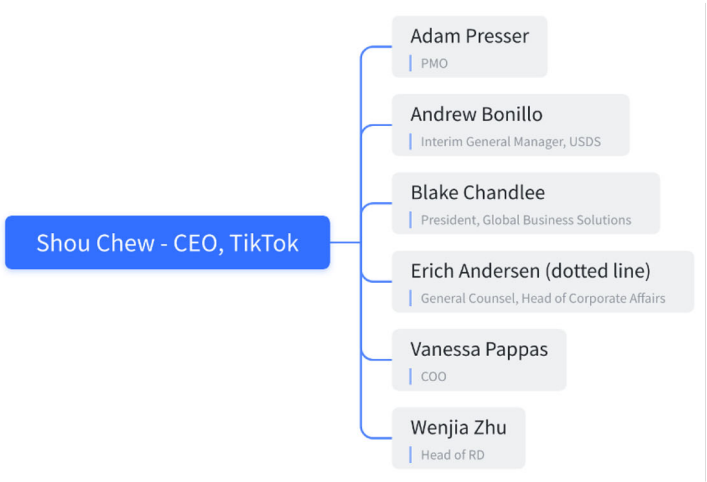
10. Who do TikTok's engineering teams report to?
    a. If your answer is Shou Chew or any other TikTok employee, can you guarantee that no engineering teams currently report to and, in the past, have ever reported to ByteDance employees in addition to TikTok employees?

TikTok's engineers report to many different leaders, by function, including teams such as security, trust and safety, and R&D.

TikTok

11. Please provide an organizational chart of TikTok's leadership.

Please see below for the TikTok CEO's leadership team as of 10/31/2022.



TikTok

Post-Hearing Questions for the Record
Submitted to Vanessa Pappas
From Senator James Lankford

"Social Media's Impact on Homeland Security"
September 29, 2022

Question 1: The *New York Times* published a report in November 2020 how the Mexican cartels are using your platform to recruit new members.[1] According to the report, this content was allowed to proliferate on TikTok. What actions, if any, did TikTok take following this report to eliminate this content? Does TikTok believe that such content violates its terms of service?

Such content is against TikTok's Community Guidelines and Terms of Service. TikTok works with external intelligence vendors, law enforcement, and our internal threat investigations team to help promptly detect and remove such content and accounts from our platform when we discover it.

Question 2: What are TikTok's content moderation policies around information related to crossing the U.S. border? Please describe these policies in detail.

TikTok's policies prohibit facilitation and solicitation of human smuggling services, including instructions on how to illegally cross or be smuggled into the United States. To that end, if an individual indicates that they are part of or participating in an illegal human smuggling operation, such content would be removed when we discover it.

Question 3: Were these policies developed in consultation with any entity outside of TikTok? If so, with whom?

TikTok consulted with several non-governmental organizations as well as an intelligence vendor to provide feedback for TikTok to consider when developing our content moderation policies around information related to crossing the U.S. border and to help operationalize these policies for our moderators.

Question 4: Has TikTok ever studied the prevalence of information on its platform related to crossing the U.S. border? If so, please share any such study with the Committee.

TikTok considers any content that depicts, glorifies, or promotes human trafficking to be a violation of the Guidelines and will remove it on the platform when discovered. TikTok seeks to remove any content that it discovers depicts, promotes, or speaks positively of human exploitation; that encourages, coordinates, or provides instructions to exploit others; that details methods for exploiting others or for illegally crossing international borders; or that provides contact information related to human exploitation. TikTok leverages our internal investigations teams to understand prevalence of content, along with rhetoric and signals (for example, known slogans).

---

[1] https://www.nytimes.com/2020/11/28/world/americas/mexico-drugs-cartel-tiktok.html

TikTok

Question 5: How do TikTok's algorithms recognize content which violates TikTok's terms of service related to human smuggling, drug smuggling, cartel activity, or crossing the U.S. southern or northern borders? What keywords, images, hashes, or other content and indicators regarding such subject matter do TikTok's algorithms recognize as violating TikTok's terms of service? In answering this question, please provide all keywords, images, hashes, or other content regarding such subject matter which TikTok would consider as violating its terms of service.

TikTok works diligently to ensure that its app is not being used to depict or enable criminal activity. For example, TikTok removes content that has been identified as involving human exploitation or smuggling and bans the user who posted the content from the platform when discovered. In addition, TikTok removes content that it discovers promotes illegal conduct, including gang, cartel, or other organized criminal activity, use of illicit drugs and other controlled substances, and adult prostitution.

We use a combination of image- and text-based AI to detect designated cartels (tools which we also use for drug and firearm detection); and we leverage text-based signals to detect facilitation of human smuggling. TikTok does not publicly disclose details about our strategies (such as keyword lists, images, or hashes), as these can be used by bad actors to circumvent our protections.

Question 6: How do TikTok's human content reviewers recognize content which violates TikTok's terms of service related to human smuggling, drug smuggling, cartel activity, or crossing the U.S. southern or northern borders? What keywords, images, hashes, or other content and indicators regarding such subject matter do TikTok's human content reviewers recognize as violating TikTok's terms of service? In answering this question, please provide all keywords, images, hashes, or other content regarding such subject matter which TikTok would consider as violating its terms of service.

TikTok's Community Guidelines prohibit content that normalizes, imitates, or facilitates human exploitation, trafficking, and smuggling. This includes the trade of humans for commercial exploitation, whether through sex trafficking, forced marriage, forced labor, slavery, the extraction and trade of human organs, or other activities conducted against a person's will, as well as the procurement or facilitation of illegal entry into a state across international borders. TikTok considers any content that depicts, glorifies, or promotes human trafficking to be a violation of the Guidelines and will remove it on the platform when discovered. TikTok seeks to remove any content that it discovers depicts, promotes, or speaks positively of human exploitation; that encourages, coordinates, or provides instructions to exploit others; that details methods for exploiting others or for illegally crossing international borders; or that provides contact information related to human exploitation.
TikTok does not publicly disclose details about our strategies (such as keyword lists, images, or hashes), as these can be used by bad actors to circumvent our protections. However, some examples of violative keywords include "tiro pegado," "busco choferes," and "tiren linea."

Question 7: Please provide a list of each piece of content on TikTok that TikTok's moderators have removed that regard human smuggling, drug smuggling, cartel activity, or crossing the U.S. southern or northern borders. In providing this list, please include how much each piece of content was viewed,

TikTok

liked, and shared. Please also include the time each piece of content was allowed to remain on TikTok's platform prior to being removed.

Content facilitating human smuggling, drug smuggling, cartel activity, or instructional content on how to illegally cross the border would violate TikTok's Community Guidelines. We do not track individual instances where moderators have removed this kind of content and so do not have that data.

As disclosed in our latest Community Guidelines Enforcement Report for 2022 Q2, around 1% of content on our platform is removed for Community Guidelines violations, and approximately 21.2% of this 1% of content takedowns fell under the category of "illegal activities and regulated goods." Of these takedowns relating to illegal activities and regulated goods, 97.6% of content is proactively removed (meaning identified and removed before it's reported); 93.1% of content is removed from our platform before any views; and 94.9% is removed within 24 hours. In addition to removing violative content, TikTok suspends or bans accounts involved in severe or repeated violations.

Question 8: In your testimony, you mentioned that TikTok works to ban all extremist groups and individuals from appearing on the app and that your organization relies up "expert" information from the Southern Poverty Law Center.

- Why do you consider the SPLC to be an expert?
- Are you aware that the SPLC designates conservative Christian organizations – like Alliance Defending Freedom and Family Research Council as hate groups?
- Does TikTok consider organizations that hold to a traditional view of marriage, believe that individual are either male or female or are prolife to be extremist hate groups?

The Southern Poverty Law Center is but one of many civil society organizations with which TikTok consults in setting standards around hate speech and hateful ideology. TikTok's policies are crafted using data-led research and consultations with a diversity of organizations and viewpoints. TikTok's policies do not rely on any one organization to determine whether content is or is not violative; instead we rely on our own policies crafted by internal experts with consultation from external experts.

Question 9: On September 19, TikTok banned CatholicVote from its platform after restricting the account for posts TikTok deemed to be "hateful behavior." Can you please explain how CatholicVote was in violation of TikTok's policies for sharing its position on gender? What steps does TikTok take to protect religious speech?

TikTok does not allow content that discriminates against anyone on the basis of protected attributes. We define hate speech and/or behavior as content that attacks, threatens, incites violence against, or otherwise dehumanizes an individual or a group on the basis of the following protected attributes: Race, Ethnicity, National origin, Religion, Caste, Sexual orientation, Sex, Gender, Gender identity, Serious disease, Disability and Immigration status. Content that dehumanizes people based on protected attributes such as sex and gender would violate our policies and be removed when detected.

TikTok

Question 10: In 2017, the National People's Congress passed the National Intelligence Law of 2017. This law requires all China-based companies to cooperate with the CCP's intelligence services by turning over user data when asked. Is TikTok's parent company ByteDance subject to the National Intelligence Law of 2017?

The TikTok platform is not available in mainland China and is not the focus of the Chinese national intelligence law. TikTok has never been asked or subpoenaed by the Chinese government to provide any U.S. user data, and the Chinese government has never asserted any legal rights over such data. Moreover, as we have stated publicly, we would not share if asked.

TikTok discloses on a regular basis in its Transparency Reports (https://www.tiktok.com/transparency/en-us/reports/) requests for user information that we receive from governments and law enforcement agencies.

Question 11: In 2020, China enacted the Data Security Law of 2020 which establishes a process to share user data throughout the Chinese government and compels companies to work with the CCP to take action against entities that threaten China's security. Is TikTok's parent company ByteDance subject to the Data Security Law?

We do not believe this law applies to TikTok. The TikTok platform is not available in mainland China and is not the focus of other Chinese laws that might compel production of data. TikTok has never been asked or subpoenaed by the Chinese government to provide any U.S. user data, and the Chinese government has never asserted any legal rights over such data.

Question 12: In 2020, China enacted the Cryptography Law which requires all encryption products to be approved by the Chinese government – giving China a backdoor into all "encrypted" and "secure" platforms. Is TikTok's parent company ByteDance subject to the Cryptography Law?

We do not believe this law applies to TikTok. The TikTok platform is not available in mainland China and is not the focus of this law.

Question 13: In 2020, China enacted the National Security Law, which eroded Hong Kong's autonomy and (among other things) threatened the security of data based in Hong Kong. In response, major tech companies including Facebook, Twitter, WhatsApp and Google stopped reviewing requests for user data from Hong Kong authorities. Is TikTok's parent company Bytedance subject to the National Security Law?

The TikTok platform is not available in mainland China and is not the focus of the Chinese national security law. TikTok has never been asked or subpoenaed by the Chinese government to provide any U.S. user data, and the Chinese government has never asserted any legal rights over such data.

TikTok

Question 14: In light of the strategic competition taking place between our countries, do you agree that the People's Republic of China has an interest in collecting the data of Americans? How does that assessment inform your security protocols and corporate structure?

TikTok is not in a position to speculate on the interests of the Chinese government.

As a general rule, we look to minimize the number of people who have access to user data and limit it to those who need the access in order to do their jobs. We employ access controls like encryption and security monitoring to secure data, and the access approval process is overseen by our U.S.-based security team.

TikTok is committed to protecting the security and integrity of our platform, which includes prohibiting unauthorized access to TikTok, as well as TikTok content, accounts, systems, or data. We are opening state-of-the-art cyber incident monitoring and investigative response centers in Washington DC, Dublin, and Singapore to enable follow-the-sun threat monitoring and intelligence gathering, as we continue working with industry-leading experts to test and enhance our defenses.

Question 15: Do you agree that in China, all industry is an extension of the Chinese state? To what extent, if any, do you assess Bytedance to be an extension of the Chinese state? How does that assessment inform your security protocols and corporate structure?

We disagree. Broadly speaking, there are two types of enterprises in China, state-owned enterprises ("SOEs") and private enterprises. ByteDance Ltd. is a private enterprise and none of its China subsidiaries are SOEs. TikTok is not available in mainland China.

ByteDance Ltd. is the ultimate parent entity that is incorporated in the Cayman Islands. The board is comprised of our co-founder and CEO Rubo Liang, Bill Ford of General Atlantic, Arthur Dantchick of Susquehanna International Group, Philippe Laffont of Coatue, and Neil Shen of Sequoia. The majority of ByteDance's investors are global institutional funds such as Baillie Gifford, Blackrock, Coatue, Fidelity, General Atlantic, KKR, Sequoia, Softbank, Susquehanna International Group, and T. Rowe Price.

Question 16: Since the TikTok application was engineered by developers based in the PRC, there are valid questions about whether the app was developed in such a way to ensure backdoor access to user data from outside the United States. What due diligence have you done to ensure that the developers of TikTok have not established backdoor connections to US-based data that would be accessible to a member of the Chinese Communist Party or any other unauthorized party?

TikTok has internal, U.S.-based teams that conduct testing to validate the integrity of the TikTok app. We have also engaged third parties to review our source code and for more than two years offered a bug bounty program through Hacker One (https://newsroom.tiktok.com/en-us/celebrating-the-cyber-community-on-tiktok), where we award external security researchers for responsibly identifying and disclosing vulnerabilities.

TikTok

TikTok is committed to implementing significant organizational, process, and technical changes related to the storage, access, and security of U.S. user data pursuant to an anticipated national security agreement, with a trusted technology partner.

Question 17: How long does it take TikTok to detect and take down child sexual abuse material, both previously detected material and new material?

TikTok has a zero tolerance approach to content that violates its minor safety policies. Any content, including animation or digitally created or manipulated media, that depicts abuse, exploitation, or endangerment of minors is a violation of TikTok's Community Guidelines and will be removed when detected.

TikTok's moderation system uses models to automatically identify videos, captions, and accounts containing CSAM and other violative content that is created and uploaded to the TikTok platform. Our moderation system also uses advanced technology to review comments for evidence of online grooming and other predatory behavior.

Each and every video uploaded to TikTok goes through automated moderation, and identified potential cases of CSAM are automatically removed or escalated for human review by a member of our moderation team. Regardless of whether content is new or known via a hashbank, our approach is the same – we take immediate action to remove it, terminate accounts, and report cases to NCMEC. To help identify CSAM, we've developed our own technology which we use, in addition to what's available from our partners - whether it's PhotoDNA or YouTube's CSAI match technology.

When TikTok identifies CSAM, it is removed from the platform and reported to NCMEC. In 2021, we made 154,618 reports to NCMEC and were alone among major platforms in not receiving any takedown requests from NCMEC. (https://www.missingkids.org/content/dam/missingkids/pdfs/2021-reports-by-esp.pdf; https://www.missingkids.org/content/dam/missingkids/pdfs/2021-notifications-by-ncmec-per-esp.pdf).

Question 18: How many employees do you have who are dedicated to ensuring child sexual abuse material is appropriately removed and reported?

Currently, TikTok's Child Safety Team (CST) has approximately 30 employees globally. This team is supported by TikTok's thousands of moderators who escalate content to them.

The CST's mission is to prevent new and ongoing sexual abuse of minors by detecting and reporting sexual abuse, predatory behavior, and eliminating the proliferation of sexual materials involving children. CST's primary function is to report incidents of child sexual abuse and exploitation on the platform to NCMEC, the National Center for Missing and Exploited Children (https://www.missingkids.org/HOME), a central repository that provides our disclosures to law enforcement for investigation and intervention in unlawful or harmful activity against children.

TikTok

In addition, the Minor Safety team within the Trust & Safety Product team aims to lead in the industry in creating safe and enriching experiences for minors. The team is comprised of product and program managers around the globe. The team manages the product roadmaps for Family Pairing and Digital Wellbeing, provides product design and moderation consultation to product teams inside and outside of Trust & Safety, and works cross-functionally to continuously improve TikTok's ability to detect and action against CSAM and other content and users that are harmful and/or exploitative of youth.

Question 19: Would you describe TikTok's efforts to take down and report child sexual abuse material as active or passive? In other words, is material passively scanned for known matches, or is there an active element to how you detect and take down this material?

TikTok takes a multi-pronged approach to identifying and removing CSAM, as described in our response to Question 17.

On top of our own technology, TikTok has integrated with NCMEC's Hash Sharing Web Services to enable the detection and removal of known CSAM at the point of upload to TikTok. The NCMEC hash list is applied against all videos uploaded to the platform and to users' avatars (profile pictures). This uses the hash format PhotoDNA. This is designed to identify real images as well as pseudo-images. TikTok also uses user GIF avatar frames for hashing and integrating with the NCMEC repository. In addition, TikTok uses Thorn's database for risky keywords to help block content or enqueue users for account review. Further, we use IWF's keyword list to supplement our existing sensitive words list, and we deploy IWF's URL list to block URLs containing CSAM that are shared through private messages or are present in a user's bio. Regardless of whether content is new or known via a hashbank, our approach is the same – we take immediate action to remove it, terminate accounts, and report cases to NCMEC. To help identify CSAM, we've developed our own technology which we use, in addition to what's available from our partners - whether it's PhotoDNA or YouTube's CSAI match technology.

TikTok recognizes that technology is ever-evolving and that we need to be prepared to address unexpected trends and challenges as they arise, and in addition to NCMEC and IWF, TikTok works with a variety of global organizations on minor safety efforts, such as the Technology Coalition, Thorn, ConnectSafely, the National Parent Teacher Association, and the Family Online Safety Institute.

Question 20: There has been an alarming uptick in sextortion (meaning a child is threatened or blackmailed with the potential release of sexual material in order to receive additional sexual content or money) and grooming on social media platforms. Between 2019 and 2021, the number of reports involving sextortion more than doubled. What has your platform done to address this alarming trend?

TikTok is deeply committed to child safety and has zero tolerance for predatory or grooming behavior toward minors. We prohibit activities that perpetuate the abuse, harm, endangerment, or exploitation of minors on TikTok. Any content, including animation or digitally created or manipulated media, that depicts abuse, exploitation, or endangerment of minors is a violation on our platform and will be removed when detected. Users must meet minimum age requirements to use TikTok, as stipulated in our Terms of Service. When underage account holders are identified, we remove their accounts from our platform and may direct them to a more appropriate app experience, if available in their region.

When it comes to safety there is no finish line, and to this end, TikTok engages with a variety of groups to stay alert to industry challenges. TikTok recently participated in an industry and law enforcement forum focused on sextortion. TikTok is also actively working with industry peers via the Technology Coalition (an alliance of global tech companies working together to combat child sexual exploitation and abuse online) (https://www.technologycoalition.org/), as well as directly with NCMEC, regarding this trend. TikTok reports to NCMEC text-based minor solicitation, grooming, and found cases of financial sextortion, the latter of which, per guidance from NCMEC, we escalate as exigent/time sensitive CyberTips.

Question 21: Please provide the total number of worldwide employees at TikTok's parent company ByteDance. Further, please provide the number of employees currently located in each respective country.

Approximately 160,000 employees globally work for ByteDance Ltd. and its subsidiaries and consolidated affiliates.

Question 22: Please provide the total number of employees at TikTok. Further, please provide the number of employees located in each respective country.

Approximately 30,000 employees globally work for TikTok, including approximately 5,200 employees in the United States.

Question 23: TikTok's parent company ByteDance's website states the company is based out of more than 200 cities globally. Please provide the number of properties the company, its holding companies, and subsidiaries leases or owns in each respective country.

In the United States, TikTok has offices in Austin, Chicago, Detroit, Irvine, Los Angeles, Mountain View, Nashville, New York, San Diego, San Francisco, San Jose, Santa Clara, Seattle (Bellevue), and Washington, DC. All locations are leased.

TikTok has offices globally in various locations including Dublin, London, Paris, Berlin, Dubai, Singapore, Jakarta, Seoul, Sydney, and Tokyo.

TikTok

Post-Hearing Questions for the Record
Submitted to Vanessa Pappas
From Senator Rick Scott

"Social Media's Impact on Homeland Security."

Wednesday, September 14, 2022

- How does TikTok ensure your China-based employees aren't members or associated with the CCP?

Consistent with standard industry practice, TikTok does not ask its employees to provide their political affiliations.

As has been reported in the press, for more than a year we have been pursuing a multi-pronged initiative called "Project Texas" to strengthen TikTok's U.S. data security program. In May 2022, TikTok announced the creation of a new division—U.S. Data Security ("USDS")—to bring heightened focus and governance to our ongoing efforts to strengthen our data protection policies and protocols, further protect our U.S. users, and build confidence in our systems and controls in the United States. This division has U.S.-based leadership. Access to U.S. user data by anyone outside of our new U.S. Data Security team will be limited by, and subject to, robust data access protocols that are being developed in close collaboration with Oracle and the U.S. government.

- TikTok's parent company, ByteDance, is a Chinese company that recently admitted certain employees outside the US can access information from American users, which raises concerns about the platform's data-sharing practices.
    - Do China-based employees have access to US users' data on the TikTok platform?
    - Is any of that information shared with the Chinese government/Chinese Communist Party?
    - Is TikTok, its employees and users subject to Chinese mainland data-sharing laws?

TikTok has long stored U.S. user data in data centers in the U.S. and Singapore, as well as in cloud-based services offered by AWS, the Google Cloud Platform, and Azure. 100% of U.S. user traffic is now being routed to Oracle Cloud Infrastructure. We are still using our U.S. and Singapore data centers for backup, but as we continue our work to deliver on U.S. data governance, we expect to delete U.S. users' protected data from our own systems and fully pivot to Oracle cloud servers located in the U.S.

TikTok

Employees outside the U.S., including China-based employees, can have access to TikTok U.S. user data subject to a series of robust cybersecurity controls and authorization approval protocols overseen by our U.S.-based security team. TikTok has an internal data classification system and approval process in place that assigns levels of access based on the data's classification and requires approvals for access to U.S. user data. The level of approval required is based on the sensitivity of the data according to the classification system.

The TikTok platform is not available in mainland China and is not the focus of other Chinese laws that might compel production of data. TikTok has never been asked or subpoenaed by the Chinese government to provide any U.S. user data, and the Chinese government has never asserted any legal rights over such data. We have not provided U.S. user data to the Chinese government or CCP, nor would we if asked. More information about government requests for user data that we receive across the world is available in our Information Request Reports, available at https://www.tiktok.com/transparency/en-us/information-requests-2021-2/.

TikTok continues to be in discussions with CFIUS, and we are confident that we are on a path to fully satisfy all reasonable U.S. national security concerns.

- How many accounts have you found advertising counterfeit pills on your platform?
    - How many of those accounts were reported to you by a user of your platform?
    - How many of those accounts did you find on your own?

As disclosed in our latest Community Guidelines Enforcement Report for 2022 Q2, around 1% of content on our platform is removed for Community Guidelines violations, and approximately 21.2% of this 1% of content takedowns fell under the category of "illegal activities and regulated goods." Of these takedowns relating to illegal activities and regulated goods, 97.6% of content is proactively removed (meaning identified and removed before it's reported); 93.1% of content is removed from our platform before any views; and 94.9% is removed within 24 hours. Of that 21.2%, approximately 13% of that content relates to drugs, controlled substances, alcohol, and tobacco. Of those takedowns relating to such content, 94.4% is proactively removed; 83.5% is removed before any views; and 86.9% is removed within 24 hours. In addition to removing violative content, TikTok suspends or bans accounts involved in severe or repeated violations, including violations for promotion or trade of drugs.

Furthermore, TikTok prohibits ads on the following (as disclosed on our public-facing Advertising Policies (https://ads.tiktok.com/help/article?aid=9550):

"Drugs and paraphernalia

- Promotion, sale, solicitation of, or facilitation of access to illegal drugs, controlled drugs, prescription drugs (prohibited in some markets), drugs for the purpose of recreation, homeopathy, enhancement, performance, including weight loss.

- Promotion, sale, solicitation of, or facilitation of access to drug paraphernalia, or accessories or supplies any of such.
- Promotion of or facilitation of access to unauthorized drugstores, pharmacies, or dispensaries.
- Depiction of or featuring drug use, drug abuse or prescription drug abuse.
- Depiction of or featuring drugs-related words, symbols, or images, whether in the form of visual or audio content, or any of such."

In sum, and as stated in previous answers, TikTok engages a range of tactics, and adapts them as needed, to ensure our policies in this realm are being enforced as effectively as possible.

- Do you proactively refer accounts mentioning counterfeit pills or fentanyl to law enforcement?
    - Do you notify the account holder that they are being investigated by law enforcement, potentially leading them to evade and continue selling their product and killing people?
    - Which law enforcement entities are you referring accounts suspected of drug sales and/or trafficking to?

TikTok's Law Enforcement Guidelines (https://www.tiktok.com/legal/law-enforcement) provide law enforcement with details on how to submit valid legal requests to TikTok as well as other useful information specific to our data disclosure practices. This set of Guidelines also reflects TikTok's commitments to cooperate with law enforcement while respecting our users' privacy in a fair, lawful, and transparent manner. We also provide a 24/7 team dedicated to responding to emergency requests from law enforcement. To obtain user information, law enforcement must provide the appropriate legal documents required for the type of information being sought, such as a subpoena, court order, search warrant, or localized equivalent. Any information request we receive is carefully reviewed for legal sufficiency. TikTok may share content or account information directly with law enforcement in the absence of a request when it believes in good faith that there is an emergency involving imminent harm or risk of death or serious physical injury to a person. Depending on the nature and type of case, referrals could include federal, state, and/or local law enforcement authorities.

- Are you committed to working with appropriate local, state, and federal law enforcement to create a Trusted Reporter program to help bring down these accounts?

TikTok

Our Law Enforcement Response Team (LERT) routinely discloses user data in response to valid legal requests from law enforcement agencies. Our Law Enforcement Outreach team facilitates and maintains relationships with these agencies to help ensure there is a direct line for communication, and we are continuously exploring additional ways to work with law enforcement while respecting our users' privacy in a fair, lawful, and transparent manner.

- Over the last year, on average, how many subpoenas have you received per month related to accounts being used to sell or distribute drugs?

TikTok generally receives a low volume of legal process explicitly associated with drug investigations. Thus far, during calendar year 2022, TikTok has received one or fewer subpoenas per month related to accounts believed to be used to sell or distribute drugs.

- What is the average turnaround time for responding to a subpoena issued by a law enforcement agency (Not including an automated or "received" style response)?

On average, TikTok responds to subpoenas within two weeks of submission, and responds to emergency disclosure requests, such as requests associated with an imminent threat to life, typically within approximately 20 minutes.

- Do your moderators search for accounts or posts that use the DEA's Emoji Drug Code, or various permutations of the Emoji Drug Code?
    - If no, why not?

Moderators maintain a comprehensive and frequently updated list of drug slang and emojis that may be used by bad actors on the platform in order to avoid detection in this realm. This list covers not only a wide swath of federally prohibited controlled substances, but also covers slang and emoji that drug traffickers may attempt to use. Our Trust & Safety team also regularly conducts on-platform internal investigations to identify and remove such content swiftly. As such, DEA's Emoji Drug Code is just one resource our policy team uses to ensure our moderators have the tools they need to successfully tackle violative controlled substance-related content on the platform.

- What has your company done to help spread awareness of the DEA's One Pill Can Kill campaign?

TikTok has engaged with the DEA on various topics, including on the Operation Overdrive initiative to combat drug-related violence and overdoses (https://www.dea.gov/press-

releases/2022/02/07/dea-launches-new-initiative-combat-drug-related-violence-and-overdoses-0).

TikTok also works with LegitScript (https://www.legitscript.com/), an organization dedicated to making internet ecosystems safer and more transparent. LegitScript provides us with intelligence on violative healthcare (including pharmacy and drug-related content) as well as tobacco and vaping content on TikTok. LegitScript also helps us identify sole-dedicated trade accounts, along with flagging accounts that show potential recidivist behavior.

TikTok has ongoing engagement with a variety of external partners, including non-government organizations and academic researchers, to bolster our efforts to keep illicit content off the TikTok platform and to increase awareness around the dangers of fentanyl and other illegal drugs. We have built out relationships with organizations such as Public Goods Project, Song for Charlie, Campaign for Tobacco Free Children, Talk to Frank, the Ad Council, and more.

CGVR

Year	Month	Week	AVERAGE of Global (include US)	AVERAGE of US	AVERAGE of Global (exclude US)
2021	2021/10	20210926 - 20211002	0.93%	1.10%	0.90%
		20211003 - 20211009	0.99%	1.20%	0.95%
		20211010 - 20211016	0.87%	0.80%	0.89%
		20211017 - 20211023	0.79%	0.90%	0.76%
		20211024 - 20211030	0.80%	0.80%	0.80%
	2021/10		0.88%	0.96%	0.86%
	2021/11	20211031 - 20211106	0.99%	1.60%	0.87%
		20211107 - 20211113	0.92%	1.10%	0.88%
		20211114 - 20211120	1.04%	1.10%	1.03%
		20211121 - 20211127	1.11%	1.10%	1.12%
	2021/11		1.02%	1.23%	0.98%
	2021/12	20211128 - 20211204	0.91%	0.80%	0.94%
		20211205 - 20211211	0.83%	0.80%	0.84%
		20211212 - 20211218	0.87%	1.00%	0.84%
		20211219 - 20211225	0.83%	0.80%	0.83%
		20211226 - 20220101	0.87%	0.80%	0.89%
	2021/12		0.86%	0.84%	0.87%
2021			0.91%	0.99%	0.90%
	2022/1	20220102 - 20220108	1.03%	1.01%	1.03%
		20220109 - 20220115	1.11%	1.16%	1.10%
		20220116 - 20220122	1.09%	1.17%	1.07%
		20220123 - 20220129	1.00%	1.10%	0.98%
	2022/1		1.06%	1.11%	1.05%
	2022/2	20220130 - 20220205	0.95%	0.87%	0.96%
		20220206 - 20220212	1.02%	0.95%	1.03%
		20220213 - 20220219	0.95%	0.91%	0.96%
		20220220 - 20220226	1.00%	0.93%	1.02%
	2022/2		0.98%	0.92%	0.99%
	2022/3	20220227 - 20220305	1.06%	0.96%	1.08%
		20220306 - 20220312	1.11%	0.99%	1.13%
		20220313 - 20220319	1.01%	0.95%	1.02%
		20220320 - 20220326	1.02%	0.95%	1.03%
		20220327 - 20220402	Null	0.64%	Null
	2022/3		1.05%	0.96%	1.07%
		20220403 - 20220409	0.92%	0.71%	0.96%

2022	2022/4	20220410 - 20220416	Null	Null	Null
		20220417 - 20220423	0.96%	1.22%	0.90%
		20220424 - 20220430	1.01%	1.46%	0.92%
		2022/4	0.96%	1.13%	0.93%
	2022/5	20220501 - 20220507	0.96%	0.99%	0.94%
		20220508 - 20220514	0.89%	0.82%	0.90%
		20220515 - 20220521	0.89%	1.02%	0.86%
		20220522 - 20220528	0.82%	0.92%	0.80%
		2022/5	0.89%	0.94%	0.88%
	2022/6	20220529 - 20220604	Null	Null	Null
		20220605 - 20220611	0.88%	0.91%	0.87%
		20220612 - 20220618	0.85%	0.97%	0.83%
		20220619 - 20220625	0.90%	1.11%	0.86%
		20220626 - 20220702	0.80%	0.88%	0.78%
		2022/6	0.86%	0.97%	0.84%
	2022/7	20220703 - 20220709	0.71%	0.76%	0.70%
		20220710 - 20220716	Null	Null	Null
		20220717 - 20220723	0.72%	0.63%	0.74%
		20220724 - 20220730	0.74%	0.70%	0.75%
		2022/7	0.72%	0.70%	0.75%
	2022/8	20220731 - 20220806	0.82%	0.80%	0.82%
		20220807 - 20220813	0.79%	0.77%	0.80%
		20220814 - 20220820	0.76%	0.63%	0.79%
		20220821 - 20220827	Null	0.94%	Null
		20220828 - 20220903	0.73%	0.74%	0.73%
		2022/8	18.84%	0.78%	0.79%
	2022/9	20220904 - 20220910	0.68%	0.59%	0.70%
		20220911 - 20220917	0.78%	0.84%	0.77%
		20220918 - 20220924	0.81%	1.03%	0.77%
		20220925 - 20221001	0.79%	1.18%	0.72%
		2022/9	0.77%	0.91%	0.74%
2022			3.02%	0.93%	0.89%

Global (excluding US) Issue Breakdown

Year	Month	Week	ANSA	M5	H&S	H6	VGC	MRG	SSD	H&A	VE	
		20210926 - 20211002		0.30%	0.12%	0.16%	0.04%	0.11%	0.13%	0.10%	0.01%	0.01%
		20211003 - 20211009		0.33%	0.11%	0.19%	0.04%	0.12%	0.13%	0.11%	0.01%	0.01%
	2021/10	20211010 - 20211016		0.31%	0.11%	0.18%	0.04%	0.10%	0.13%	0.10%	0.01%	0.01%
		20211017 - 20211023		0.27%	0.09%	0.15%	0.03%	0.09%	0.13%	0.09%	0.01%	0.01%
		20211024 - 20211030		0.29%	0.10%	0.16%	0.04%	0.07%	0.12%	0.09%	0.01%	0.01%
		2021/10		0.30%	0.10%	0.17%	0.04%	0.10%	0.12%	0.10%	0.01%	0.01%
		20211031 - 20211106		0.31%	0.10%	0.17%	0.04%	0.10%	0.13%	0.09%	0.01%	0.01%
2021	2021/11	20211107 - 20211113		0.30%	0.11%	0.17%	0.03%	0.09%	0.13%	0.10%	0.01%	0.01%
		20211114 - 20211120		0.29%	0.11%	0.22%	0.05%	0.10%	0.14%	0.12%	0.01%	0.01%
		20211121 - 20211127		0.40%	0.11%	0.22%	0.05%	0.12%	0.16%	0.13%	0.01%	0.01%
		2021/11		0.35%	0.11%	0.20%	0.04%	0.10%	0.14%	0.11%	0.01%	0.01%
		20211128 - 20211204		0.33%	0.11%	0.20%	0.04%	0.09%	0.12%	0.11%	0.01%	0.01%
		20211205 - 20211211		0.30%	0.10%	0.17%	0.04%	0.07%	0.12%	0.10%	0.01%	0.01%
	2021/12	20211212 - 20211218		0.31%	0.10%	0.16%	0.04%	0.10%	0.11%	0.11%	0.01%	0.01%
		20211219 - 20211225		0.33%	0.10%	0.16%	0.04%	0.09%	0.10%	0.09%	0.01%	0.01%
		20211226 - 20220101		0.34%	0.09%	0.15%	0.04%	0.09%	0.12%	0.12%	0.01%	0.00%
		2021/12		0.32%	0.10%	0.17%	0.04%	0.09%	0.11%	0.11%	0.01%	0.01%
	2021			0.32%	0.10%	0.18%	0.04%	0.10%	0.13%	0.10%	0.01%	0.01%
		20220102 - 20220108		0.35%	0.11%	0.23%	0.05%	0.11%	0.15%	0.13%	0.00%	0.01%
	2022/1	20220109 - 20220115		0.38%	0.11%	0.25%	0.05%	0.12%	0.14%	0.14%	0.01%	0.01%
		20220116 - 20220122		0.36%	0.12%	0.21%	0.05%	0.12%	0.15%	0.12%	0.01%	0.01%
		20220123 - 20220129		0.32%	0.11%	0.19%	0.05%	0.10%	0.14%	0.11%	0.01%	0.01%
		2022/1		0.36%	0.11%	0.22%	0.05%	0.11%	0.15%	0.13%	0.01%	0.01%
		20220130 - 20220205		0.33%	0.10%	0.19%	0.05%	0.10%	0.14%	0.12%	0.01%	0.01%
	2022/2	20220206 - 20220212		0.36%	0.11%	0.19%	0.06%	0.12%	0.16%	0.12%	0.01%	0.01%
		20220213 - 20220219		0.35%	0.10%	0.18%	0.05%	0.09%	0.16%	0.12%	0.01%	0.01%
		20220220 - 20220226		0.34%	0.10%	0.21%	0.07%	0.11%	0.17%	0.11%	0.02%	0.01%
		2022/2		0.35%	0.10%	0.19%	0.06%	0.11%	0.16%	0.12%	0.01%	0.01%
		20220227 - 20220305		0.37%	0.11%	0.21%	0.07%	0.11%	0.17%	0.12%	0.01%	0.02%
	2022/3	20220306 - 20220312		0.34%	0.14%	0.22%	0.06%	0.12%	0.16%	0.14%	0.01%	0.01%
		20220313 - 20220319		0.31%	0.11%	0.19%	0.06%	0.12%	0.17%	0.13%	0.01%	0.01%
		20220320 - 20220326		0.33%	0.14%	0.19%	0.05%	0.11%	0.17%	0.13%	0.01%	0.02%
		20220327 - 20220402	Null	Null	Null	Null	Null	Null	Null	Null	Null Null	
		2022/3		0.34%	0.13%	0.20%	0.06%	0.11%	0.17%	0.13%	0.01%	0.02%
		20220403 - 20220409		0.30%	0.13%	0.18%	0.05%	0.10%	0.14%	0.12%	0.01%	0.01%
	2022/4	20220410 - 20220416	Null	Null	Null	Null	Null	Null	Null	Null	Null	
		20220417 - 20220423		0.26%	0.13%	0.18%	0.04%	0.12%	0.14%	0.11%	0.01%	0.01%
		20220424 - 20220430		0.28%	0.13%	0.18%	0.04%	0.10%	0.14%	0.12%	0.01%	0.01%
		2022/4		0.28%	0.13%	0.18%	0.04%	0.11%	0.14%	0.12%	0.01%	0.01%
		20220501 - 20220507		0.27%	0.15%	0.17%	0.04%	0.10%	0.13%	0.14%	0.01%	0.01%
	2022/5	20220508 - 20220514		0.25%	0.12%	0.16%	0.04%	0.11%	0.13%	0.14%	0.01%	0.01%
2022		20220515 - 20220521		0.24%	0.13%	0.16%	0.04%	0.10%	0.13%	0.13%	0.01%	0.01%
		20220522 - 20220528		0.23%	0.12%	0.16%	0.05%	0.09%	0.11%	0.12%	0.01%	0.01%
		2022/5		0.25%	0.13%	0.17%	0.04%	0.10%	0.13%	0.13%	0.01%	0.01%
		20220529 - 20220604	Null	Null	Null	Null	Null	Null	Null	Null	Null	
	2022/6	20220605 - 20220611		0.24%	0.14%	0.18%	0.04%	0.09%	0.12%	0.13%	0.01%	0.01%
		20220612 - 20220618		0.25%	0.12%	0.18%	0.03%	0.08%	0.11%	0.13%	0.00%	0.01%
		20220619 - 20220625		0.24%	0.14%	0.17%	0.04%	0.09%	0.12%	0.13%	0.01%	0.01%
		20220626 - 20220702		0.24%	0.12%	0.16%	0.04%	0.09%	0.11%	0.10%	0.01%	0.01%
		2022/6		0.24%	0.13%	0.17%	0.04%	0.09%	0.12%	0.12%	0.01%	0.02%
		20220703 - 20220709		0.21%	0.11%	0.14%	0.03%	0.07%	0.09%	0.10%	0.00%	0.01%
	2022/7	20220710 - 20220716		0.23%	0.10%	0.17%	0.04%	0.09%	0.11%	0.11%	0.01%	0.01%
		20220717 - 20220723		0.21%	0.11%	0.16%	0.04%	0.08%	0.11%	0.10%	0.00%	0.01%
		20220724 - 20220730		0.21%	0.12%	0.15%	0.04%	0.08%	0.10%	0.09%	0.01%	0.01%
		2022/7		0.22%	0.11%	0.15%	0.04%	0.08%	0.10%	0.10%	0.01%	0.01%
		20220731 - 20220806		0.22%	0.13%	0.17%	0.04%	0.09%	0.12%	0.11%	0.00%	0.01%
	2022/8	20220807 - 20220813		0.22%	0.12%	0.19%	0.03%	0.07%	0.12%	0.11%	0.01%	0.01%
		20220814 - 20220820		0.21%	0.13%	0.17%	0.04%	0.07%	0.12%	0.09%	0.00%	0.02%
		20220821 - 20220827	Null	Null	Null	Null	Null	Null	Null	Null	Null Null	
		20220828 - 20220903		0.21%	0.12%	0.16%	0.03%	0.07%	0.11%	0.10%	0.00%	0.01%
		2022/8		0.22%	0.13%	0.17%	0.04%	0.08%	0.12%	0.10%	0.00%	0.01%
		20220904 - 20220910		0.20%	0.11%	0.15%	0.03%	0.07%	0.10%	0.09%	0.00%	0.01%
	2022/9	20220911 - 20220917		0.22%	0.13%	0.16%	0.04%	0.07%	0.11%	0.09%	0.01%	0.01%
		20220918 - 20220924		0.22%	0.12%	0.17%	0.03%	0.08%	0.10%	0.11%	0.01%	0.02%
		20220925 - 20221001		0.20%	0.11%	0.17%	0.04%	0.08%	0.10%	0.09%	0.00%	0.01%
		2022/9		0.21%	0.12%	0.16%	0.04%	0.08%	0.10%	0.10%	0.01%	0.01%
	2022			0.27%	0.12%	0.18%	0.04%	0.10%	0.13%	0.12%	0.01%	0.01%

US Issue Breakdown

Year	Month	Week	ANSA	MS	H&B	HB	VGC	IARG	SSD	I&A	VE	
		20210926 - 20211002		0.40%	0.19%	0.06%	0.09%	0.15%	0.21%	0.14%	0.01%	0.01%
		20211003 - 20211009		0.43%	0.18%	0.11%	0.10%	0.15%	0.19%	0.15%	0.01%	0.01%
	2021/10	20211010 - 20211016		0.28%	0.13%	0.07%	0.04%	0.11%	0.11%	0.09%	0.03%	0.01%
		20211017 - 20211023		0.18%	0.16%	0.15%	0.11%	0.05%	0.16%	0.04%	0.01%	0.02%
		20211024 - 20211030		0.17%	0.13%	0.09%	0.06%	0.10%	0.17%	0.15%	0.00%	0.01%
		2021/10		0.28%	0.16%	0.10%	0.08%	0.12%	0.17%	0.11%	0.01%	0.01%
		20211031 - 20211106		0.49%	0.26%	0.32%	0.14%	0.20%	0.29%	0.12%	0.01%	0.00%
	2021/11	20211107 - 20211113		0.27%	0.27%	0.20%	0.08%	0.13%	0.19%	0.09%	0.01%	0.00%
2021		20211114 - 20211120		0.30%	0.21%	0.26%	0.10%	0.13%	0.16%	0.03%	0.00%	0.00%
		20211121 - 20211127		0.32%	0.12%	0.21%	0.06%	0.18%	0.18%	0.11%	0.02%	0.00%
		2021/11		0.35%	0.22%	0.25%	0.10%	0.16%	0.21%	0.09%	0.01%	0.00%
		20211128 - 20211204		0.20%	0.20%	0.11%	0.06%	0.10%	0.14%	0.08%	0.00%	0.00%
		20211205 - 20211211		0.22%	0.13%	0.07%	0.11%	0.10%	0.07%	0.10%	0.00%	0.00%
	2021/12	20211212 - 20211218		0.24%	0.22%	0.15%	0.13%	0.14%	0.13%	0.08%	0.01%	0.01%
		20211219 - 20211225		0.16%	0.15%	0.11%	0.10%	0.11%	0.08%	0.01%	0.01%	0.02%
		20211226 - 20220101		0.17%	0.20%	0.11%	0.08%	0.07%	0.17%	0.07%	0.00%	0.01%
		2021/12		0.20%	0.19%	0.11%	0.10%	0.10%	0.13%	0.08%	0.00%	0.01%
	2021			0.28%	0.19%	0.15%	0.09%	0.13%	0.16%	0.10%	0.01%	0.01%
		20220102 - 20220108		0.26%	0.20%	0.16%	0.14%	0.09%	0.12%	0.12%	0.01%	0.00%
	2022/1	20220109 - 20220115		0.29%	0.29%	0.24%	0.15%	0.09%	0.08%	0.12%	0.01%	0.01%
		20220116 - 20220122		0.32%	0.15%	0.19%	0.13%	0.18%	0.08%	0.16%	0.00%	0.00%
		20220123 - 20220129		0.19%	0.18%	0.16%	0.15%	0.16%	0.15%	0.18%	0.00%	0.01%
		2022/1		0.27%	0.18%	0.19%	0.14%	0.13%	0.11%	0.16%	0.01%	0.01%
		20220130 - 20220205		0.21%	0.15%	0.17%	0.05%	0.09%	0.10%	0.14%	0.00%	0.00%
	2022/2	20220206 - 20220212		0.22%	0.04%	0.12%	0.11%	0.19%	0.17%	0.11%	0.03%	0.02%
		20220213 - 20220219		0.20%	0.18%	0.16%	0.09%	0.12%	0.07%	0.13%	0.00%	0.00%
		20220220 - 20220226		0.16%	0.15%	0.12%	0.09%	0.13%	0.08%	0.17%	0.01%	0.00%
		2022/2		0.20%	0.13%	0.14%	0.09%	0.11%	0.11%	0.14%	0.00%	0.01%
		20220227 - 20220305		0.35%	0.09%	0.07%	0.09%	0.16%	0.03%	0.14%	0.00%	0.01%
	2022/3	20220306 - 20220312		0.26%	0.15%	0.16%	0.11%	0.11%	0.12%	0.08%	0.02%	0.00%
		20220313 - 20220319		0.19%	0.22%	0.14%	0.08%	0.10%	0.07%	0.12%	0.00%	0.01%
		20220320 - 20220326		0.14%	0.24%	0.12%	0.10%	0.14%	0.10%	0.10%	0.00%	0.03%
		20220327 - 20220402		0.14%	0.16%	0.06%	0.05%	0.07%	0.07%	0.07%	0.00%	0.02%
		2022/3		0.22%	0.17%	0.11%	0.09%	0.12%	0.08%	0.10%	0.00%	0.02%
		20220403 - 20220409		0.10%	0.22%	0.09%	0.07%	0.08%	0.09%	0.07%	0.00%	0.00%
	2022/4	20220410 - 20220416		0.20%	0.26%	0.23%	0.12%	0.14%	0.06%	0.17%	0.02%	0.00%
		20220417 - 20220423		0.21%	0.13%	0.17%	0.11%	0.23%	0.16%	0.10%	0.00%	0.03%
		20220424 - 20220430		0.21%	0.30%	0.20%	0.09%	0.35%	0.16%	0.10%	0.00%	0.00%
		2022/4		0.19%	0.23%	0.17%	0.10%	0.20%	0.12%	0.11%	0.01%	0.01%
		20220501 - 20220507		0.14%	0.29%	0.11%	0.08%	0.16%	0.07%	0.14%	0.01%	0.03%
	2022/5	20220508 - 20220514		0.06%	0.21%	0.13%	0.04%	0.22%	0.10%	0.04%	0.00%	0.02%
2022		20220515 - 20220521		0.12%	0.26%	0.11%	0.06%	0.19%	0.13%	0.07%	0.00%	0.01%
		20220522 - 20220528		0.02%	0.13%	0.16%	0.05%	0.21%	0.06%	0.06%	0.01%	0.02%
		2022/5		0.09%	0.30%	0.13%	0.06%	0.20%	0.10%	0.08%	0.00%	0.02%
		20220529 - 20220604	Null	Null	Null	Null	Null	Null	Null	Null	Null	Null
	2022/6	20220605 - 20220611		0.09%	0.38%	0.11%	0.05%	0.12%	0.06%	0.12%	0.00%	0.00%
		20220612 - 20220618		0.07%	0.39%	0.18%	0.05%	0.12%	0.05%	0.10%	0.00%	0.01%
		20220619 - 20220625		0.17%	0.42%	0.16%	0.06%	0.12%	0.09%	0.12%	0.00%	0.01%
		20220626 - 20220702		0.05%	0.31%	0.09%	0.06%	0.16%	0.13%	0.09%	0.00%	0.00%
		2022/6		0.10%	0.38%	0.13%	0.06%	0.13%	0.08%	0.11%	0.00%	0.01%
		20220703 - 20220709		0.05%	0.31%	0.11%	0.05%	0.13%	0.07%	0.08%	0.00%	0.00%
	2022/7	20220710 - 20220716	Null	Null	Null	Null	Null	Null	Null	Null	Null	Null
		20220717 - 20220723		0.05%	0.17%	0.10%	0.04%	0.08%	0.10%	0.12%	0.00%	0.00%
		20220724 - 20220730		0.08%	0.28%	0.15%	0.03%	0.10%	0.05%	0.02%	0.00%	0.00%
		2022/7		0.06%	0.25%	0.12%	0.04%	0.10%	0.07%	0.07%	0.00%	0.00%
		20220731 - 20220806		0.11%	0.30%	0.10%	0.05%	0.10%	0.02%	0.05%	0.00%	0.01%
	2022/8	20220807 - 20220813		0.04%	0.17%	0.15%	0.07%	0.09%	0.11%	0.09%	0.00%	0.03%
		20220814 - 20220820		0.05%	0.19%	0.10%	0.08%	0.11%	0.08%	0.03%	0.00%	0.00%
		20220821 - 20220827		0.07%	0.20%	0.13%	0.15%	0.10%	0.11%	0.11%	0.00%	0.04%
		20220828 - 20220903		0.13%	0.19%	0.14%	0.07%	0.08%	0.06%	0.05%	0.01%	0.02%
		2022/8		0.08%	0.22%	0.12%	0.08%	0.10%	0.08%	0.06%	0.00%	0.02%
		20220904 - 20220910		0.06%	0.20%	0.11%	0.06%	0.07%	0.07%	0.04%	0.00%	0.00%
	2022/9	20220911 - 20220917		0.13%	0.21%	0.12%	0.13%	0.09%	0.07%	0.10%	0.00%	0.00%
		20220918 - 20220924		0.15%	0.16%	0.10%	0.08%	0.06%	0.15%	0.06%	0.00%	0.02%
		20220925 - 20221001		0.23%	0.18%	0.21%	0.06%	0.12%	0.22%	0.19%	0.00%	0.02%
		2022/9		0.14%	0.24%	0.16%	0.08%	0.09%	0.11%	0.10%	0.00%	0.01%
	2022			0.15%	0.23%	0.14%	0.08%	0.11%	0.10%	0.10%	0.00%	0.01%

Issue Verticals Description

MS	Minor safety
ANSA	Adult nudity and sexual activities
H&B	Harassment and bullying
SSD	Suicide, self-harm, and dangerous acts
IARG	Illegal activities and regulated goods
VGC	Violent and graphic content
HB	Hateful behavior
I&A	Integrity and authenticity
VE	Violent extremism

Committee on Homeland Security and Governmental Affairs
Hearing on
"Social Media's Impact on Homeland Security"
September 14, 2022

Mr. Jay Sullivan, General Manager of Bluebird, Twitter

Chairman Gary C. Peters

1. Please provide the following information regarding Twitter's current employees as of September 1, 2022: (i) the total number of full-time engineers at your company, (ii) the total number of full-time engineers working full time on ensuring trust and safety or integrity of your platform, and (iii) the total number of full-time engineers working full time on product development.

   RESPONSE

   Health, which captures trust and safety considerations and site integrity, is everyone's job at Twitter. Approximately 2,200 people work on enforcing Twitter Rules, specifically. Presently, approximately 1,500 people work in the "Bluebird" consumer product team, of which more than 800 are a part of the Bluebird engineering team. Additionally, engineers from other teams do cross-functional work that supports the consumer product and health teams. For example, our so-called Redbird team works on the fundamental and backend aspects of our service. The Redbird team complements the work of our product engineers, Health, and Trust and Safety teams, among others.

2. Please provide all metrics evaluated in your A/B testing for each of the last 2 years, including a description of each metric and the percentage of A/B tests in which the metric was evaluated.

   RESPONSE

   A/B testing is the general concept of comparing the outcomes of two variables, which could capture a variety of activities. We constantly conduct experiments to understand whether our products are working as intended and expected. For example, if we see a decrease in the number of user reports of abusive content, we might design an experiment to test whether there has been an actual decrease in abusive content or whether the design of our user reporting functionality could be improved. For each experiment, we might derive a unique variable or set of variables to measure. We do

this across the company for a variety of functions. To this end, we are not able to gather and provide exhaustive metrics.

3. Does Twitter use metrics to measure whether certain demographics see a disproportionate amount of harmful content on your platforms, such as, for example, the number of users who have repeated exposures to harmful content, or the amount of harmful content seen by the user in the 99th percentile of exposure to harmful content? If yes, please share a list of these metrics and their current values and values from last year.

RESPONSE

Twitter takes steps to minimize the exposure or amplification of content on our platform that violates the Twitter Rules, including our policies prohibiting violence, hateful conduct, terrorism or violent extremism, and abuse and harassment. As noted in our most recent Transparency Report, from July 1, 2021, through December 31, 2021, Twitter required users to remove 4 million Tweets that violated the Twitter Rules. Of these 4 million Tweets removed, 71% received fewer than 100 impressions prior to removal, with an additional 21% receiving between 100 and 1,000 impressions. Only 8% of removed Tweets had more than 1,000 impressions. In total, impressions on these violative Tweets accounted for less than 0.1% of all impressions for all Tweets during that time period.

4. What is the total number of views that violative content has received across your platform over the last year, month, and day? Please break down by category of violative content and platform.

RESPONSE

Please see the response to question #3. More relevant metrics, including information about trends over time, can be found here.

5. Are trust and safety metrics used to determine compensation of all product development employees in their individual goals? If no, why not? If yes, please provide which metrics are used and how they affect compensation.

RESPONSE

Twitter as a company is incentivized to keep our platform healthy and safe. The top two objectives within Twitter's product development organization are to develop

products to grow the number of Twitter users and to prioritize health and safety. These two priorities go hand-in-hand because if people are not protected from hate, abuse, and harassment, they will leave the service. We therefore build health and safety into the design of new features, but if we are not satisfied, we will pause, delay, or stop a product rollout because of health and safety concerns. Health is equally as important as user growth as they complement each other. Further, health is the job of everyone in product development, and is integral to product development.

At Twitter, employee compensation is determined according to a variety of factors, including tenure and experience, as well as individual employee performance. Our guiding principles are anchored on the goals of being able to attract, incentivize, and retain talented employees who can develop, implement, and deliver on long-term value creation strategies; promote a healthy approach to risk by reinforcing our values which serve to motivate our employees; and provide competitive compensation that is aligned with the market and is fair relative to our peers. Building healthy products is a core consideration with respect to individual employee performance for relevant teams, including Bluebird.

6. Are growth, engagement, or revenue metrics used to determine compensation of all product development employees in their individual goals? If yes, please provide which metrics are used and how specifically they affect compensation.

RESPONSE

Please see the response to #5.

7. How much has Twitter spent in the last year on trust and safety?

RESPONSE

Health is everyone's job at Twitter; therefore, we cannot accurately quantify this metric. Please see Twitter's latest Quarterly Earnings Report and Annual Report, which include metrics about Twitter's overall spending.

8. How much has Twitter spent in the last year on product development?

RESPONSE

Please see the response to #7.

9. Please provide the number of employees who specifically research extreme content on your platform and whether that research includes the extent to which your recommendation algorithms promote extreme content.

RESPONSE

Across Twitter, there are many employees who research and examine the effects of extreme content on the platform as part of their daily activities. These include team members from within Trust and Safety, Engineering, Public Policy, Research, and other elements of Twitter. The goal of this cross-functional research and examination is to proactively disrupt the posting of extreme content and to help fill gaps in preventing the platform from being exploited by those who wish to promote such content.

10. What restrictions, if any, does Twitter place on the type of research employees can perform or the type of data they can access?

RESPONSE

Twitter employs a variety of measures to ensure data is appropriately secured and protected. Basic access is granted automatically based on role, and additional access must be requested based on necessity. Twitter grants access only to the data that employees require to complete their work and when there are no better alternatives available.

There are rigorous checks and measures, including detection of suspicious access and activity, and additional permissions constrain what systems employees have access to within the production environment. Depending on how teams operate, the type of system or data at issue, and the risks presented, access controls befitting the use case are implemented, employees may need to undergo additional mandatory training, and additional background checks are performed where necessary.

11. What is the takedown rate of (i) hate speech, and (ii) all violative content over the last year, month, and day?

RESPONSE

Please see our latest Twitter Transparency Report for all information related to available content moderation metrics.

Ranking Member Rob Portman

1. Last week, Reuters reported that at least 30 brands had ads featured next to child exploitation content on Twitter. Additionally, internal teams identified the failure of Twitter to effectively moderate and combat the prevalence of child sexual abuse material on the platform.

a. What steps does Twitter plan to take to improve its ability to identify, moderate, and combat child sexual abuse material?

RESPONSE

Twitter has zero-tolerance for child sexual exploitation. We aggressively fight online child sexual abuse and invest in technology and tools to enforce our policy. Our dedicated teams work to stay ahead of bad-faith actors and to help ensure we are protecting minors from harm — both online and offline.

We work collaboratively with stakeholders, including the National Center for Missing and Exploited Children (NCMEC) and law enforcement, to address this societal challenge, and we continue to invest in resources, tools, and technology to aggressively fight this egregious and illegal activity.

Our approach to combating child sexual exploitation material is centered around proactive detection of this content. We have invested in tools and technology to help us surface violative content and remove it expeditiously. These tools and technology improve our efficiency and speed in removing this type of content, as well as removing the burden from users to report content.

Of the unique Twitter accounts that we suspended for child sexual exploitation content during the last reporting period in our Twitter Transparency Center, more than 91% were surfaced proactively and removed by a combination of technology solutions, including PhotoDNA and internal proprietary technical tools.

We use proprietary tools to detect patterns and behavior related to accounts that may be engaging with this type of content or involved in related networks. We use a combination of technologies, including PhotoDNA, to monitor broadcasts, private groups, avatar images, and bios and ensure that any violative content is being identified and removed. We also take measures to ensure this content is not surfaced in search as an additional safeguard.

Although we strive to proactively detect and remove this content, any individual can report potential violations of this policy, regardless of whether they have a Twitter account. Violations can be reported through in-app reporting or via our dedicated web form.

In 2021, Twitter removed 1,050,751 unique accounts for violations of our child sexual exploitation policies. As part of our commitment to increased transparency around our enforcement actions, we make information on our enforcement activities available at transparency.twitter.com.

Recent reports about Twitter provide an outdated, moment in time glance at just one aspect of our work in this space and are not an accurate reflection of where we are today. Our work here continues; we have sharpened our focus and expanded our resources dedicated to child safety. We are also hiring—underscoring our continued investment in this work.

b. How will you measure the effectiveness of these actions?

RESPONSE

We publish metrics about our enforcement of our policy in the Twitter Transparency Report. In our most recent Transparency Report, from July - December 2021, we suspended 596,997 unique accounts for child sexual exploitation – a 32% increase since our previous report. Of these, 91% of suspended accounts were identified proactively by employing internal proprietary tools and industry hash sharing initiatives. These tools and initiatives support our efforts in surfacing potentially violative content for further review and, if appropriate, removal.

c. Will you commit to providing ongoing updates on these efforts to the Committee?

RESPONSE

Yes. Please reach out to our teams at your convenience to set up a briefing with our experts to discuss Twitter's work to combat child sexual abuse material. We will also include the latest metrics in our next Transparency Report.

Senator Thomas R. Carper

1. Social media has fundamentally transformed the way we stay in touch with loved ones, create connections, and the way businesses large and small reach customers around the world. Unfortunately, it is also used to recruit, influence, and mobilize individuals to commit violent attacks.

   The rate at which information is shared using social media has significantly impacted the speed at which individuals may be radicalized and inspired to violence, which can narrow the window of opportunity for law enforcement to stop them before they commit violent attacks.

   a. What information does your platform proactively share with law enforcement to prevent or flag the spread of violent or hateful content on social media before an act of violence is committed? What more needs to be done?

      RESPONSE

      Twitter uses the information that we collect to provide for the safety and security of our users, products, services, and users' accounts. This includes verifying users' identities, authenticating users' accounts, and defending against fraud, unauthorized use, and illegal activity. We also use the information to evaluate and affect the safety and quality of content on Twitter—this includes investigating and enforcing our policies and terms, as well as applicable law.

      We may preserve, use, share, or disclose users' information if we believe that it is reasonably necessary to: comply with a law, regulation, legal process, or governmental request; protect the safety of any person; protect the safety or integrity of our platform, including to help prevent spam, abuse, or malicious actors on our services; explain why we have removed content or accounts from our services (e.g., for a violation of Twitter Rules); address fraud, security, or technical issues; or protect our rights or property, or the rights or property of those who use our services.

Senator Kyrsten Sinema

Cartel-Sponsored Content

In my capacity as Chair of the Border Management Subcommittee, I asked TikTok, Meta, Twitter, and YouTube to each commit to sharing cartel recruitment content with the Department of Homeland Security (DHS) as quickly as possible. You responded: "yes, with the appropriate privacy and oversight, I believe we could do that." I appreciate Twitter's willingness to work with DHS to help protect Arizonan youth from being targeted by cartels.

1. Please elaborate. Are there circumstances where known and identified cartel recruitment content would not be shared with DHS? If so, please describe in detail under what, if any, circumstances cartel-sponsored content on Twitter that seeks to recruit American minors would not be shared with DHS.

   RESPONSE

   Twitter takes enforcement action under our illegal or certain regulated goods or services policy and our violent organizations policy, which are most relevant to this topic. Our enforcement of these policies continues to evolve to meet the changing threat landscape and as such, have evolved to include drug cartels in recent years.

   Twitter uses the information that we collect to provide for the safety and security of our users, products, services, and users' accounts. This includes verifying users' identities, authenticating users' accounts, and defending against fraud, unauthorized use, and illegal activity. We also use the information to evaluate and affect the safety and quality of content on Twitter—this includes investigating and enforcing our policies and terms, as well as applicable law.

   We may preserve, use, share, or disclose users' information if we believe that it is reasonably necessary to: comply with a law, regulation, legal process, or governmental request; protect the safety of any person; protect the safety or integrity of our platform, including to help prevent spam, abuse, or malicious actors on our services; explain why we have removed content or accounts from our services (e.g., for a violation of Twitter Rules); address fraud, security, or technical issues; or protect our rights or property, or the rights or property of those who use our services.

2. Will you commit to engage with the Border Patrol and the sheriffs' departments in counties along the Arizona-Mexico border regarding these concerns and provide each of them with a designated point-of-contact at Twitter?

RESPONSE

Yes.

Deepfakes

Deepfakes – AI-generated videos that make people appear to do or say things they didn't actually do or say – present serious homeland security risks. In fact, in other countries such as Gabon and Malaysia, we've already seen fake videos of leaders trigger political turmoil. In Ukraine, there was a deepfake of President Zelensky supposedly telling Ukrainians to lay down their arms against Russia. While platforms have increasingly adopted policies to ban many deepfakes, these policies only can be enforced if you know what is or is not real.

1. What is Twitter doing to improve deepfake detection? And in cases where deepfakes are too advanced to be detected, how do you go about determining a video's authenticity?

   RESPONSE

   The Twitter Rules prohibit the sharing of synthetic, manipulated or out-of-context media (frequently called "deepfakes" or "cheapfakes") that may deceive or confuse people or lead to harm. In addition, the Twitter Rules provide for labeling media that violates our Synthetic and Manipulated Media (SAMM) policy. When we determine that content is eligible for labeling under our SAMM policies, we search our platform for further instances of that content and label further instances of it. When we label media under our SAMM policy, we include additional contextual information, such as a curated fact check, and we may reduce its visibility on our platform.

   Additionally, we stay up to date on trends in manipulated media through dialog with academics and civil society, and through multi-stakeholder convenings including our membership in the Coalition for Content Provenance and Authenticity.

Senator Rick Scott

1. How many accounts have you found advertising counterfeit pills on your platform?
    a. How many of those accounts were reported to you by a user of your platform?

    RESPONSE

    From January through September of this year, we removed approximately 121,357 accounts for violating our policies that address illegal or certain regulated goods or services. Of these accounts, 74% or 90,895 were spam. Our teams proactively identified and suspended 53.3% or 16,237 accounts, and 46.7% or 14,225 accounts were suspended reactively.

2. How many of those accounts did you find on your own? Are you committed to working with appropriate local, state, and federal law enforcement to create a Trusted Reporter program to help bring down these accounts?

    RESPONSE

    Please see the response to #1 for metrics.

    Yes, we are committed to working with law enforcement to receive reports for content that is violative of the Twitter Rules.

3. Do you proactively refer accounts mentioning counterfeit pills or fentanyl to law enforcement?

    a. Do you notify the account holder that they are being investigated by law enforcement, potentially leading them to evade and continue selling their product and killing people?

    b. Which law enforcement entities are you referring accounts suspected of drug sales and/or trafficking to?

    RESPONSE

    We provide information to law enforcement with valid legal requests. More information about our process can be found here.

For purposes of transparency and due process, Twitter's policy is to notify users (e.g., prior to disclosure of account information) of requests for their Twitter or Periscope account information, including a copy of the request, unless we are prohibited from doing so (e.g., an order under 18 U.S.C. § 2705(b)).

4. Over the last year, on average, how many subpoenas have you received per month related to accounts being used to sell or distribute drugs?

RESPONSE

Law enforcement is not required to inform us of what their investigation relates to, so we are not able to know how many subpoenas concern the selling or distribution of drugs.

5. What is the average turnaround time for responding to a subpoena issued by a law enforcement agency (Not including an automated or "received" style response)?

RESPONSE

Our response times vary depending on a number of factors, such as type of legal process, type of allegation, the number of accounts specified in the request, and the complexity of the request.

6. Do your moderators search for accounts or posts that use the DEA's Emoji Drug Code, or various permutations of the Emoji Drug Code?

    a. If no, why not?

RESPONSE

Twitter has received lists of terms from the United States Drug Enforcement Administration (DEA) to better understand trends to inform our detection efforts.

7. What has your company done to help spread awareness of the DEA's One Pill Can Kill campaign?

RESPONSE

For several years, Twitter has supported the DEA's #TakeBackDay, a biannual campaign that encourages Americans to dispose of their prescription drugs at a local drop-off site in order to prevent drug misuse.

Twitter provides a custom emoji for #TakeBackDay and coordinates with congressional leaders and Administration officials to raise awareness for this event on the platform.

The inaugural #TakeBackDay Twitter emoji helped to elevate the conversation, which helped to increase engagement multiple-fold compared to previous years.

We are open to exploring how we can support the One Pill Can Kill campaign.

Senator Alex Padilla

1. Twitter's users span the globe, and it makes the same commitments to Western, English-speaking communities as it does to non-Western, non-English language speaking communities. A recently disclosed third-party audit of Twitter's disinformation and misinformation work found that Twitter's integrity team lacked language expertise in countries it was serving even though 80% of Twitter users are outside of the United States. Here in the United States and around the globe the native language of millions of people is a language other than English.

    a. What are the top five non-English languages for content that users within the United States encounter on Twitter and what percentage of users in the United States encounter content within each of those languages on a weekly basis?

    RESPONSE

    The top five non-English languages for content that users within the United States encounter on Twitter are: Spanish, Japanese, Farsi, Chinese, and Arabic. Combined, content in these languages represents approximately 6% of Tweets in the United States. Approximately 78% of Tweets that people in the United States see are in English.

    b. For each of these questions, please break down by employment status (full time in-house employees or contract-level).

        i. How many employees focus exclusively on trust and safety issues at Twitter?

        RESPONSE

        We currently employ more than 2,200 full time employees and contractors across the globe who are directly involved in enforcing our rules on Twitter, all of whom work in different ways to improve the health of the platform. This number does not include support-related employees that augment our overall content moderation work, nor subject matter specialists or those working on safety product and technology investments.

        ii. Of the people working exclusively on trust and safety issues, how many focus on the content moderation of non-English language content, how

many of them focus on each of those top five non-English languages, and how many of them focus on non-United States users?

RESPONSE

Our policies are developed for a global audience and our enforcement is conducted globally to enable 24/7 coverage.

iii. Of the people working exclusively on trust and safety issues, how many of them focus on non-English language content in the United States? Of the people working exclusively on trust and safety issues, how many of them focus on non-English language content for users outside of the United States?

RESPONSE

Please see the response to question #3(ii).

iv. How many of them focus on each of those top five non-English languages?

RESPONSE

Please see the response to question #3(ii).

v. How many of them focus primarily on United States-based users, and how many of them focus primarily on non-United States based users?

RESPONSE

Please see the response to question #3(ii).

c. When making community standards enforcement decisions, is all content that is reviewed for conformance with your community standards reviewed in the original language or are some or all subject to automated translation before being reviewed? What percentage of total content reviewed is subject to automated translation? Please break this down across the top five languages used by Twitter users.

RESPONSE

Our teams utilize a variety of resources and tools, human and technical, to effectively enforce the Twitter Rules. Our automated tools can detect suspected misinformation in multiple languages, including English, Spanish, French and Portuguese. At present, more than half of Twitter Rules enforcement actions occur via automated review, though human review by native speakers supplements and extends automated review to ensure that content is actioned or not actioned appropriately. In some cases, subject matter experts on a given topic, such as civic misinformation, confer with native language speakers to ensure content is reviewed in the full context of local dialects. Additionally, we have also worked to improve our detection of potentially violative content in East Asian languages by continuing to improve our user reporting tools in Tagalog and Korean.

d. Are Twitter's community standards available in each of the languages of your users?

RESPONSE

The Twitter Rules are available in the following languages: English, Spanish, Japanese, Korean, Portuguese, Dutch, Turkish, French, Italian, Arabic, German, Bahasa Indonesia, Russian, Hindi, Hebrew, Chinese Simplified and Traditional, Thai, Vietnamese, Malay, Filipino, Farsi, Danish, Finnish, Swedish, Norwegian, Polish, Hungarian, Romanian, Ukrainian, Bulgarian, Catalan, Slovenian, Czech, Dari, Urdu, Kurdish, Oromo, Pashto, and Tigrinya.

e. What is Twitter doing to ensure that it has cultural competency for all the communities it serves?

RESPONSE

Twitter employs a truly global, multicultural workforce that powers our work across many functions that serve the public conversation around the world. The Twitter Trust and Safety Council, which is a group of independent expert organizations from around the world, also provides a broad range of perspectives. Together, they advocate for safety and advise us as we develop our products, programs, and rules. At the end of 2019, we expanded the Council to include even more global experts and diverse perspectives.

The Council is made up of several advisory groups, each dedicated to issues critical to the health of the public conversation. Areas of focus include Online

Safety and Harassment, Human and Digital Rights, Suicide Prevention, and Mental Health, Child Sexual Exploitation, and Dehumanization.

A list of advisory groups, organized by areas of focus, can be found here.

2. In Twitter's most recent transparency report, it says that impressions on Tweets that violated Twitter's rules and were actioned on accounted for less than 0.1% of all impressions.

   a. What about Tweets that were not actioned on? What retrospective analysis does Twitter have in place?

   RESPONSE

   We undertake a number of retrospective analyses to assess the effectiveness of our product and policy. For example, following the 2020 United States election, we published a retrospective blog that shared statistics on the impact of our civic integrity enforcement, including on our product interventions to limit the sharing of violative content.

   b. How does Twitter assess its false positive rate—content you actioned for perceived violation of community standards, but that didn't actually violate the rules—and what is that rate on Twitter? Please break this down across the top five languages used by Twitter users.

   RESPONSE

   While we strive to get things right, we sometimes make mistakes. To this end, when Twitter actions content, we provide an appeals process for users who believe that content was actioned mistakenly. We do not calculate the rate at which actions are later reversed following an appeals process. We publish data on enforcement of the Twitter Rules at the Twitter Transparency Center.

   c. Internally, what metric do you use to evaluate the success of enforcement of your community standards?

   RESPONSE

   We measure the enforcement of the Twitter Rules using "harmful impressions," or the number of times a violative Tweet has been seen before it has been actioned.

From July 1, 2021, through December 31, 2021, Twitter required users to remove 4 million Tweets that violated the Twitter Rules. Of the Tweets removed, 71% received fewer than 100 impressions prior to removal, with an additional 21% receiving between 100 and 1,000 impressions. Only 8% of removed Tweets had more than 1,000 impressions. In total, impressions on these violative Tweets accounted for less than 0.1% of all impressions for all Tweets during that time period. More detail is available at the Twitter Transparency Center.

3. In your testimony, you state that "before any major product or policy launch, a cross-functional group of people will work together to consider potential risks, unintended consequences, responses from bad actors, and risk of abuse." A whistleblower complaint that was filed earlier this year alleged that "[t]he Fleets product avoided undergoing security and privacy reviews before launch. When serious issues were identified at the last minute, Fleets was launched without addressing them."

   a. Was the Fleets product exempt from your testified Twitter policy with respect to product launches?

      RESPONSE

      Fleets was a product feature that Twitter launched in November 2020 as a lower-pressure, ephemeral way for people to share their fleeting thoughts. The company decided in July 2021 to discontinue the Fleets feature because the company was not seeing an increase in the number of new people joining the conversation with Fleets like it had hoped.

      As Twitter has explained, the whistleblower disclosure contained inconsistencies and inaccuracies. In addition, Fleets was launched before Mr. Zatko joined Twitter, so his knowledge of Fleets is limited. While Twitter's review of the whistleblower disclosure is ongoing, our current understanding is that a health review looking at, among other things, the product's potential risks and unintended consequences, consistent with Twitter's practices at the time, was conducted prior to launching Fleets.

   b. Have any other products been exempt from your testified Twitter policy? If so, which products were exempt?

      RESPONSE

Products at Twitter undergo review prior to launch in order to assess potential risks, unintended consequences, exploitation by bad actors, and the potential risk for abuse.

4. Election integrity is more important than ever. Platforms have a crucial role to play in ensuring that voters are able to access good, accurate election information, and in maintaining the basic integrity of our elections system. Twitter has made announcements about what it is doing around the midterm elections, though much of your strategy for the midterms are similar to steps taken two years ago.

   a. Have you studied how effective your efforts were during the last election to ensure that your users knew where, how, and when to vote, and if so, what were the results of those studies?

      RESPONSE

      In addition to the detail provided in response to question #2a above, we used a number of product features to promote the availability of reliable civic information in 2020, including:

      - Directing Twitter users to reliable voting information at vote.gov when they search for voting-related terms on our platform;
      - Creating a human-curated Elections Hub on our Explore tab, with reliable local information on the election in all 50 states; and
      - Using "pre-bunk" prompts to inform Twitter users that final election results might not be available on election night. Our pre-bunks were seen 389 million times, and were available in English and Spanish.

      Additionally, this year Twitter launched the industry-leading Twitter Moderation Research Consortium. It is open to global researchers from across academia, civil society, NGOs, and journalism. We prioritize transparency by sharing more data on more issues to those who are studying content moderation.

   b. What do you assess to be the biggest threats to election integrity on Twitter in advance of the 2022 elections? What do you assess to be the biggest threats to election integrity on Twitter in advance of the 2024 elections? What steps are you taking to mitigate these identified threats?

      RESPONSE

Our preparations for the 2022 elections have been ongoing for more than a year. We have frequent consultations with experts in civil society and academia to understand the evolving threat landscape. We also trained state and local elections administrators on how to use our platform safely, including how to enable two-factor authentication, and how to report suspected misinformation. More on our preparations for the upcoming elections are available on our blog.

c. How are you staying ahead of the evolving threats to election integrity?

RESPONSE

As part of our continual investments in improving and refining our approach to civic integrity, Twitter's teams take stock of their respective work in prior elections and adjust to better meet challenges.

We built on our learnings from the 2020 United States elections by, for example, redesigning the labels we apply to Tweets that we determine to be violative of the Twitter Rules. The new labels increased click-through rates by 17%, meaning that more people were clicking labels to read debunking content. We also saw notable decreases in engagement with Tweets labeled with the new design: -13% in replies, -10% in Retweets and -15% in likes.

We have also made changes to how we recommend content on our platform. Earlier this year, we tested ways to prevent misleading Tweets from being recommended through notifications. Early results show that impressions on misleading information dropped by 1.6 million per month, as a direct result of the experiment.

Twitter works with experts on platform manipulation to share data about misinformation, coordinated harmful activity, and safety, among other areas. Twitter hopes to create a model for the industry to follow, as well as share findings publicly as we learn.

While Twitter will continue to work on increasing proactive detection of content that violates the Twitter Rules in the short-term, Twitter knows that it is critical that it invests in projects that create decentralized networks and a broader system of checks and balances to help mitigate any systemic bias that may exist.

For additional detail, please see the response to #4b.

d. What is Twitter doing to keep election officials safe from doxxing and other online harassment campaigns?

RESPONSE

Please see the responses to #4a and #4b.

Senator James Lankford

1. Please provide a list of each piece of content on Twitter that Twitter's moderators have removed that regard human smuggling, drug smuggling, cartel activity, or crossing the U.S. southern or northern borders. In providing this list, please include how much each piece of content was viewed, liked, and shared. Please also include the time each piece of content was allowed to remain on Twitter prior to being removed.

   RESPONSE

   From January through September of this year, we removed approximately 121,357 accounts for violating our policies that address illegal or certain regulated goods or services. Of these accounts, 74% or 90,895 were spam. Our teams proactively identified and suspended 53.3% or 16,237 accounts and 46.7% or 14,225 accounts were suspended reactively.

   In our latest Twitter Transparency Report, we disclosed that violative Tweets accounted for less than 0.1% of all impressions for all Tweets during that time period.

2. Has Twitter assessed how Twitter is used to facilitate unlawful activity in relation to the border? Please answer yes or no. If yes, please share this study with the Committee. Additionally, has Twitter assessed how the platform is used to aid in individuals crossing the U.S. border illegally? Please answer yes or no. If yes, please share this study with the Committee.

   RESPONSE

   We maintain a good relationship with the DEA–their teams flag content to us that our teams review, independently, under the Twitter Rules. They have also provided us with information about trends to inform our detection measures.

3. Prior to action being taken against the account, the Sinaloa Cartel (**@carteidsinaloa)** was able to amass more than 88,000 followers on Twitter. Please share with the committee what delayed action being taken against the account of the Sinaloa Cartel. Additionally, please share with the committee the rationale behind the removal of the Sinaloa Cartel's account.

   RESPONSE

   This account was suspended because it violated our violent organizations policy.

4. What are Twitter's content moderation policies around information related to crossing the U.S. border? Please describe these policies in detail.

RESPONSE

Relevant to the issues concerning drug and human trafficking at the United States border, our illegal or certain regulated goods or services policy can be found here. With respect to cartel activity, there is no place on Twitter for violent organizations, including terrorist organizations, violent extremist groups, or individuals who affiliate with and promote their illicit activities. The violence that these groups engage in or promote jeopardizes the physical safety and well-being of those targeted. Our assessments under this policy are informed by national and international terrorism designations, as well as our violent extremist group and violent organizations criteria. More details regarding our violent organizations policy can be found here.

5. Were these policies developed in consultation with any entity outside of Twitter? If so, with whom?

RESPONSE

Creating a new policy or making a policy change requires in-depth research around trends in online behavior, developing clear external language that sets expectations around what is allowed, and creating enforcement guidance for reviewers that can be scaled across millions of Tweets.

While drafting policy language, we gather feedback from a variety of internal teams, as well as our Trust and Safety Council. This is vital to ensure we are considering global perspectives around the changing nature of online speech, including how our rules are applied and interpreted in different cultural and social contexts. Finally, we train our global review teams, update the Twitter Rules, and start enforcing the new policy. We outline our philosophy and development process here.

6. How have Twitter's policies on moderating content that includes human smuggling, drug smuggling, cartel activity, or crossing the U.S. southern or northern borders evolved since the platform's inception? Please provide the committee with a timeline that details the policy evolution of how Twitter has previously and now currently handles content that features human smuggling, drug smuggling, cartel activity, or crossing the U.S. southern or northern borders and details as to what motivated the changes (if any) to these policies?

RESPONSE

Please refer to our illegal or certain regulated goods or services policy and our violent organizations policy. Our enforcement of these policies continues to evolve to meet the changing threat landscape and as such, have evolved to include drug cartels in recent years.

7. Members of Cartels and other criminal organizations that partake in and post content related to: human smuggling, drug smuggling, cartel activity, or crossing the U.S. southern or northern borders frequently create new accounts to evade deplatforming. Has Twitter assessed and or studied instances of users creating new accounts after they've been banned from the platform or other forms of ban evasion? If yes, please provide the committee with the findings from these inquiries.

RESPONSE

As stated in our ban evasion policy, attempts by an individual to circumvent prior enforcement, including permanent suspensions, are against our rules. Creating a new account or repurposing an already-existing account, for example, violates our policies. In addition, we take enforcement action on accounts whose apparent intent is to replace or promote content affiliated with a suspended account.

8. How long does it take Twitter to detect and take down child sexual abuse material, both previously detected material and new material?

RESPONSE

When we are made aware of content depicting or promoting child sexual exploitation, including links to third-party sites where this content can be accessed, our policy is to remove it without further notice and report it to the NCMEC. This also applies to content which we identify proactively, or which is reported as part of an ongoing law enforcement investigation. Twitter is currently in compliance with all federal and international regulations regarding child sexual abuse material takedown windows.

9. How many employees do you have who are dedicated to ensuring child sexual abuse material is appropriately removed and reported?

RESPONSE

Approximately 2,200 people work on enforcing Twitter Rules, which includes our policy prohibiting child sexual exploitation.

10. Would you describe Twitter's efforts to take down and report child sexual abuse material as active or passive? In other words, is material passively scanned for known matches, or is there an active element to how you detect and take down this material?

RESPONSE

Twitter has zero-tolerance for child sexual exploitation. We aggressively fight online child sexual abuse and invest in technology and tools to enforce our policy. Our dedicated teams work to stay ahead of bad-faith actors and to help ensure we are protecting minors from harm — both online and offline.

We work collaboratively with stakeholders, including the NCMEC and law enforcement, to address this societal challenge, and we continue to invest in resources, tools, and technology to aggressively fight this egregious and illegal activity.

Our approach to combating child sexual exploitation material is centered around proactive detection of this content. We have invested in tools and technology to help us surface violative content and remove it expeditiously. These tools and technology improve our efficiency and speed in removing this type of content, as well as removing the burden from users to report content.

Of the unique Twitter accounts that we suspended for child sexual exploitation content during the last reporting period in our Twitter Transparency Center, more than 91% were surfaced proactively and removed by a combination of technology solutions, including PhotoDNA and internal proprietary technical tools.

We use proprietary tools to detect patterns and behavior related to accounts that may be engaging with this type of content or involved in related networks. We use a combination of technologies, including PhotoDNA, to monitor broadcasts, private groups, avatar images, and bios and ensure any violative content is being identified and removed. We also take measures to ensure this content is not surfaced in search as an additional safeguard.

Although we strive to proactively detect and remove this content, any individual can report potential violations of this policy, regardless of whether they have a Twitter account. Violations can be reported through in-app reporting or via our dedicated web form.

In 2021, Twitter removed 1,050,751 unique accounts for violations of our child sexual exploitation policies. As part of our commitment to increased transparency around our enforcement actions, we make information on our enforcement activities available at transparency.twitter.com.

11. There has been an alarming uptick in sextortion (meaning a child is threatened or blackmailed with the potential release of sexual material in order to receive additional sexual content or money) and grooming on social media platforms. Between 2019 and 2021, the number of reports involving sextortion more than doubled. What has your platform done to address this alarming trend?

RESPONSE

To address this challenge, we collaborate with stakeholders including the NCMEC and, through the NCMEC, with international and domestic law enforcement agencies, equipping them with timely and relevant information which we believe may help apprehend perpetrators and rescue victims from exploitative circumstances. We are committed to continued investment in resources, tools, and technology to aggressively fight this egregious activity.

In addition, we continue to build Twitter with the safety of users in mind. Through product features and settings we reduce the burden of dealing with unwelcome interactions and empower people to have control over their experience.

Senator Jon Ossoff

1. Does Twitter have the capability to monitor and log every time a Twitter employee accesses user data?

   RESPONSE

   Twitter's policy is that employees must have a valid business case to access a system or data. Under Twitter's policy, access to data sets or systems is given to individuals based on the business purpose that the individuals have for accessing such data sets or systems. This allows Twitter the capability to evaluate how and when teams around the world need to access its internal tools and customer data to provide account services and keep Twitter running. Similarly, while Twitter engineers have access to the Twitter code base to perform their jobs, they are only able to deploy code changes to the production environments that they support.

   Twitter employees authenticate to production systems via a variety of mechanisms. Twitter actively uses a variety of monitoring and logging mechanisms on systems and data to better ensure that employees are not accessing systems or data that they do not have a business justification to access. In all cases, Twitter aims to authenticate the specific employee, system, or service account that is connecting to a given service or system, and limits the actions the employee or system can take, and the data it can access, appropriately.

   a. If so, does Twitter log each event where a Twitter employee accesses user data?

      RESPONSE

      Please see the response to #1, above.

2. Please enumerate all the types of user data which a Twitter employee can access (such as, but not limited to, history of use of the platform, browser history, geolocation history, or IP addresses). Please differentiate levels or types of access by types of employee (seniority, function in the company, etc.).

   RESPONSE

   Please see the response to #1, above.

○