S. HRG. 118–432

OPEN HEARING: FOREIGN THREATS TO ELECTIONS IN 2024—ROLES AND RESPONSIBILITIES OF U.S. TECHNOLOGY PROVIDERS

HEARING

BEFORE THE

SELECT COMMITTEE ON INTELLIGENCE

OF THE

UNITED STATES SENATE

ONE HUNDRED EIGHTEENTH CONGRESS

SECOND SESSION

SEPTEMBER 18, 2024

Printed for the use of the Select Committee on Intelligence

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

U.S. GOVERNMENT PUBLISHING OFFICE

57–025 WASHINGTON : 2025

SELECT COMMITTEE ON INTELLIGENCE

(Established by S. Res. 400, 94th Cong. 2d Sess.)

MARK R. WARNER, Virginia, *Chairman*
MARCO RUBIO, Florida, *Vice Chairman*

RON WYDEN, Oregon
MARTIN HEINRICH, New Mexico
ANGUS S. KING, JR., Maine
MICHAEL F. BENNET, Colorado
ROBERT P. CASEY, JR., Pennsylvania
KIRSTEN E. GILLIBRAND, New York
JON OSSOFF, Georgia
MARK KELLY, Arizona

JAMES E. RISCH, Idaho
SUSAN M. COLLINS, Maine
TOM COTTON, Arkansas
JOHN CORNYN, Texas
JERRY MORAN, Kansas
JAMES LANKFORD, Oklahoma
MIKE ROUNDS, South Dakota

CHARLES E. SCHUMER, New York, *Ex Officio*
MITCH MCCONNELL, Kentucky, *Ex Officio*
JACK REED, Rhode Island, *Ex Officio*
ROGER F. WICKER, Mississippi, *Ex Officio*

WILLIAM WU, *Staff Director*
BRIAN WALSH, *Minority Staff Director*
KELSEY STROUD BAILEY, *Chief Clerk*

CONTENTS

SEPTEMBER 18, 2024

OPENING STATEMENTS

WITNESSES

SUPPLEMENTAL MATERIAL

QUESTIONS FOR THE RECORD

OPEN HEARING: FOREIGN THREATS TO ELECTIONS IN 2024—ROLES AND RESPONSIBILITIES OF U.S. TECHNOLOGY PROVIDERS

WEDNESDAY, SEPTEMBER 18, 2024

U.S. SENATE,
SELECT COMMITTEE ON INTELLIGENCE,
Washington, DC.

The Committee met, pursuant to notice, at 2:35 p.m., in Room SH–216 in the Hart Senate Office Building, Hon. Mark R. Warner, Chairman of the Committee, presiding.

Present: Senators Warner (presiding), Rubio, Heinrich, King, Bennet, Gillibrand, Ossoff, Kelly, Risch, Collins, Cotton, Cornyn, Lankford.

OPENING STATEMENT OF HON. MARK R. WARNER, A U.S. SENATOR FROM VIRGINIA

Chairman WARNER. I am going to call this hearing to order. And I want to welcome today's witnesses: Mr. Kent Walker, President of Global Affairs and Chief Legal Officer, Alphabet; Mr. Nick Clegg, President, Global Affairs, Meta; and Mr. Brad Smith, Vice Chair and President of Microsoft.

Today's hearing builds on this Committee's longstanding practice of educating the public about the intentions and practices of foreign adversaries seeking to manipulate our country's electoral process. I do know we have all come a long way since 2017, and as many folks may remember, there was a lot of skepticism that our adversaries might have utilized America's social media platforms for intelligence activities.

It was almost seven years ago that in response to inquiries from this Committee that Facebook shared first evidence of what would become an expansive discovery documenting Russia's use of tens of thousands of inauthentic accounts across Facebook, Instagram, YouTube, Twitter, Reddit, LinkedIn, and even smaller platforms like Gab and Tumblr and Medium and Pinterest, all to try to divide Americans and influence their votes.

Through this Committee's bipartisan investigation into the Russian interference in the 2016 election, we learned that Russia had devoted millions to wide-ranging influence campaigns that literally generated hundreds of millions of online impressions which sowed political division, racial division, and impersonated social, political, and faith groups of all stripes across all ends of the political spectrum to infiltrate and manipulate our debate.

Our committee's bipartisan efforts also resulted in a set of recommendations for government, for the private sector, and for political campaigns recommendations for which I hope today's hearing will serve as a status check.

These recommendations included greater information sharing between the U.S. Government and the private sector about foreign malicious activity—not domestic—foreign malicious activity; greater transparency measures by platforms to inform users about that malicious activity; as well as more information on the origin and authenticity of information that was presented to them.

This is something that didn't get a lot of attention: the facilitation of open-source research by academics and civil society organizations to better assist platforms here and others and the public in identifying malicious use of social media, again, by foreign actors.

On the government side we have also seen some significant progress. Let me state right now that the 2020 election I think was the most secure in the United States' history, and that is verified by election security experts, and I want to commend the Trump administration for helping that come about.

Now it came about because the progress has been made through a combination of the bipartisan appropriation of funding for election upgrades, things that folks on both sides now have been calling for for a long time; paper records, implementing audits to verify results, a better postured, frankly, national security community that we have oversight on to track and expose and disrupt foreign adversarial election threats and I think a pretty successful effort to share threat information about foreign influence activity with the private sector.

U.S. tech companies as well have made progress, although as I warned all of our witnesses, albeit uneven, since 2016. These include, and I want to cite because many of you were present, when the three companies in front of us and literally 24 other companies, including companies where a lot, unfortunately, a lot of this is taking place right now X, formerly known as Twitter, which wouldn't even send a representative today, where 27 companies signed in Munich what was called the Tech Accord to Combat Deceptive Use of AI in 2024 Elections—not just in America, but around the world.

While I appreciate the voluntary commitments that were made there, you know, I think it has been uneven about "where is the beef?" and how much has actually been done.

Recently, I sent letters to all 27 of those companies. Some came back with specificity some of you. Unfortunately, others simply ignored even responding. And why we are doing this, and I think on a bipartisan basis is, there are four new factors that I think have raised my concerns dramatically.

The first is, I'm certain our adversaries realize this is effective and cheap. Putin clearly understands if he wants to try to undermine support—American support for Ukraine, weighing in and frankly putting up fake information can help him in that matter.

Similarly, we have seen since the conflict between Israel and Hamas post-October 7, this has also been a ripe area for foreign misinformation and disinformation. Again, we have seen Iran dramatically increase their efforts to stoke social discord in the U.S. while, again, potentially seeking to shape elections.

We have seen less from China, but there have been some efforts by China, not at the national level, but on down ballot races where candidates may not be taking a pro-CCP position.

Recently, literally in the last eight weeks, we have seen a covert influence project led by RT to bankroll unwitting U.S. political influencers on YouTube. We have seen a wide-ranging Russian campaign that frankly has not gotten much media attention because I think they focused on the guys in Tennessee and not some of the slides that we are going to put up later in our questioning part where major institutions like the Washington Post and Fox News—the bad guys have basically put out false information under those banners with the goal of spreading what sounds like credible sounding narratives to really shape American voters' perceptions of candidates and campaigns.

And we have seen—and this Committee has called this out—efforts to infiltrate American protests over the conflict in Gaza by Iranian influence operatives, who, again, seek to stoke division and in many cases in terms of these efforts to denigrate former President Trump.

I do want to acknowledge in these recent efforts you all have played a positive role. I want to thank Meta, and I hope our committee's interest in this subject helped move you yesterday when you guys decided to take down RT and related Russian influence operations.

I want to thank Microsoft for being forward leaning and publicly sharing information on, again, some of the Russian activity. And I want to thank Alphabet—and I want to call you guys Facebook and Google for the less informed, when you were one of the first ones to come forward on the sources on the Iranian hacks. So, compliments to all of you on that.

On an overall basis, we have also seen the scale and sophistication of the kind of attacks be escalated. When we think about AI tools, we all know about that. I think we originally thought this would be in the form of deep fakes, video and audio alteration. You are going to see AI-type tools being used to create what appears and virtually any American voter would think is a real Fox News or Washington Post site when in reality, it isn't.

And unfortunately, Congress has not been able to take on this issue. But I would point out, and it is a pretty broad swath of individual States, and they range across the political spectrum that have really put some pretty significant guardrails in place at least in terms of deep fake manipulation in their States' elections, and that is Alabama, Texas, Michigan, Florida, and California. I wish we could take some of the best ideas of some of those States and bring them to the national level.

Most of you have indicated that you have not seen, and I think the good news is so far, we have not seen the kind of massive AI interference that we might have expected particularly in the British or French elections, but as we know that from past times, the real time this will gear up will be closer to the election.

And third, the truth is, way back in 2016, Russia had to create fake personas to spin wild stories. Unfortunately, we now have a case where too many Americans, frankly, don't trust key U.S. institutions from Federal agencies to local law enforcement to tradi-

tional media. There is an increased reliance on the internet. I think most of us would try to tell our kids, "Just because you saw it on the internet doesn't mean it's true." But the job of the adversary to amplify things that are stated by Americans goes up—goes up dramatically.

Finally, we have seen a concerted litigation campaign that has sought to undermine the Federal Government's ability to share this vital threat information between you guys and the government and vice versa. And frankly, a lot of those independent academic third-party checkers have really been bullied in some cases or litigated into silence. For instance, we have seen the shuttering of the election disinformation work at Stanford's Internet Observatory as well as the termination of a key research project at Harvard Shorenstein Center. We need those academic researchers in the game as that independent source.

And again, this is a question that really bothers me, and we will—I know, we may litigate this a bit—but too many of the companies have dramatically cut back on their own efforts to prohibit false information. Again, we are talking about foreign sources. And we have seen the rise—and Senator Rubio and I have been in the lead on this—of a foreign-owned platform that has a huge reach in the case of TikTok, that has huge national security concerns. I'm very, very glad that over 80 percent of both the House and the Senate voted to say that a creative platform shouldn't be ultimately controlled by the CCP.

Now, in the last open hearing we had on this topic we heard about what the Federal Government is doing to disrupt. We are going to continue to get with law enforcement and the IC before election day. But this is really our effort to try to urge you guys to do more to kind of alert the public that this problem has not gone away. Lord knows, we have enough differences between Americans that those differences don't need to be exacerbated by our foreign adversaries.

Again, we are not cherry-picking these adversaries. These are nation-states that are in the law of our country—China, Russia, Iran, North Korea and others that have been designated as foreign adversaries.

The truth is, we are 48 days away from the election. And the final point I want to make clear is that we need to do all we can before the election, but I also think it is not like at the end of election night particularly assuming how close this election will be, that this will be over. One of my greatest concerns is that the level of misinformation, disinformation that may come from our adversaries after the polls close could actually be as significant as anything that happens up to closing of the polls on election night.

With that, I appreciate you are here.

Let me just, before I go to Senator Rubio, when we do the open hearings and I appreciate Senators Cornyn, Cotton, and a lot of our colleagues getting here early. We are going to do by seniority rather than at the gavel.

With that, Senator Rubio.

OPENING STATEMENT OF HON. MARCO RUBIO, A U.S. SENATOR FROM FLORIDA

Vice Chairman RUBIO. Thank you for holding this hearing. Thank you all for agreeing to be here. This is important. It is actually a tricky and difficult topic, because I think there are two kinds of things we are trying to address.

The first is generated disinformation. And I think you are going to describe some of those efforts today. But that is some foreign adversary—Iran, China, Russia—they create or make something up and then they amplify it. They make it up, they push it out there and they hope people believe it.

It is actually something—I remember giving a speech back in 2018 or 2019, warning about AI-generated videos that were going to be a wave of the future in terms of trying to influence what people think and see, and we have seen some of that already play out. That is pretty straightforward.

Let me tell you where it gets complicated. Where it gets complicated is, there is a preexisting view that people have in American politics. I use this as an example, not because I generally agree with it, but because this is an important example. There are people in the United States who believe that perhaps we shouldn't have gotten involved with Ukraine or shouldn't have gotten involved in the conflict in Europe. Vladimir Putin also happens to believe and hope that view will conclude.

Now there is someone out there saying something that whether you agree with them or not is a legitimate political view that is preexisting, and now some Russian bot decides to amplify the views of an American citizen who happens to hold those views. And the question becomes: Is that disinformation, is that misinformation, is that an influence operation because an existing view is being amplified?

Now, it is easy to say, well, just take down the amplifiers. But the problem is it stigmatizes the person whose view it is. Now the accusation is that that person isn't simply holding a view, they are holding the same view that Vladimir Putin has on the same topic or something similar to what he has, and as a result, they themselves must be an asset. That is problematic and it is complicated. And as we try to manage all this, we recall that in 2020—and this is now well known. Obviously, it has been well discussed. There was a laptop—Hunter Biden's laptop. There was a story in the New York Post and 51 former—and I say "former" because I have people comment all the time saying, "intelligence officers." These are former intelligence officials, went out and said: This has all the attributes of a Russian disinformation campaign. And as a result, the New York Post who posted the original story had their story censored and taken down, their account locked. There was a concerted effort on the basis of that letter to silence a media outlet in the United States on something that actually turned out not to be a Russian disinformation. Even though, I imagine maybe the Russians wanted to spread that story. They might have amplified it, but it also happened to be factual.

We know, based on the letter from the CEO of Meta, that the government pressured him during the COVID pandemic to censor

certain views, and he expressed regret about agreeing to some of that.

So, there are people in this country that had their accounts locked or even got in some cases canceled out because they questioned the efficacy of masks—something that we now know that Dr. Fauci agreed that masks were not a solution to all the problems.

The question whether there was a lab leak put out the lab leak theory that at one time was considered a conspiracy and a flat out lie, and now our own intelligence agencies are saying it is 50 percent likely, just as likely as naturally occurring.

So, this is a tricky minefield. And it is even trickier now because Russia is still doing it more than anybody else. But the others—you don't need to have a big expensive operation to pursue some of this. I think we should anticipate that in years to come—and it is happening already—the Iranians are going to get into this business. They already are. The Chinese are going to get into this business. They already are. And you see them using that in other countries to sow discord and division. It is coming. It is also North Korea, multiple—and maybe even friendly States who have a preference on how American public opinion turns.

So, I do think it is important to understand what our policies are today in terms of identifying what is disinformation, what is actually generated by a foreign adversary versus the amplification of a preexisting belief in America which has left a lot of people in a position of being labeled collaborators when, in fact, they just hold views that on that one issue happen to align with what some other country hopes we believe as well.

I am very interested to learn what our internal policies are in these companies, because I think it is a minefield that we may end up sowing in an effort to prevent discord. I don't want to sow discord, and that is one of the dangers that we are now flirting with.

Thank you for being here. I look forward to hearing your testimony.

Chairman WARNER. And before I go, I just want to reemphasize and I agree with Senator Rubio, Americans have got the right to say whatever, their First Amendment right to say that we agree or disagree, no matter how crazy.

I do think there is difference when foreign intelligence services cherry-pick information and amplify it that in many ways stokes division. And that is again where the core of this debate is, and we are anxious to hear your testimony.

I am not sure who drew the short straw to go first.

OPENING STATEMENT OF KENT WALKER

PRESIDENT, GLOBAL AFFAIRS, AND CHIEF LEGAL OFFICER, ALPHABET

Mr. WALKER. Happy to launch.

Chair Warner, Vice Chair Rubio, Members of the Committee: Thank you all for the opportunity to be with you today.

Google, Alphabet, is in the business of earning the trust of our users. We take seriously the importance of protecting free expression and access to a range of viewpoints while also maintaining and enforcing responsible policy frameworks. A critical aspect of

that responsibility is doing our part to protect the integrity of democratic processes around the world. That is why we have long invested in significant new capabilities, updated our policies, and introduced new tools to address threats to election integrity.

We recognize the importance of enabling people who use our services in America and abroad to speak freely about the political issues that are most important to them. At the same time, we continually take steps to prevent the misuse of our tools and our platforms, particularly attempts by foreign state actors to undermine democratic elections.

To help advance this work, we created the Google Threat Intelligence Group which combines our Threat Analysis Group or TAG and Mandiant intelligence. Google Threat Intelligence identifies, monitors, and tackles threats, including coordinated influence operations and cyber espionage campaigns. We disrupt activity on a regular basis, and we publish our findings, and we provide expert analysis on threats originating from the kinds of countries we are talking about: Russia, China, Iran, and North Korea, as well as from the criminal underground.

This year alone, we have seen a variety of malicious activity, including cyberattacks, efforts to compromise personal email accounts of high-profile political actors, and influence operations both on and off our platforms that are seeking to sow discord among Americans the way you were both discussing.

We remain on the lookout for new tactics and techniques in both cybersecurity and disinformation campaigns. We are seeing some foreign state actors experimenting with generative AI to improve existing cyberattacks like probing for vulnerabilities or creating spear phishing emails. Similarly, we see generative AI being used to more efficiently create fake websites, misleading news articles, and robotic social media posts.

We have not yet seen AI bring about a sea change in these attacks, but we do remain alert to new attack vectors.

To help us all stay ahead, we continue to invest in state-of-the-art capabilities to identify AI generated content.

We have launched since_id, an industry leading tool that watermarks and identifies AI generated content in texts, in audio, in images, and in video. We were also the first tech company to acquire election advertisers to prominently disclose ads that include realistic looking content that is synthetic or digitally altered.

On YouTube, when creators upload content, we now require them to indicate whether it contains material that appears realistic which we then label appropriately. And we will soon begin to use content credentials that is a new form of tamper evident metadata coming out if the C2PA program that we will discuss, I'm sure, to identify the provenance of content across ads, search, and YouTube, and to help our users identify AI generated material.

We, our users, industry, law enforcement, and civil society all play important roles in safeguarding election integrity. We encourage our high-risk users, including elected officials and candidates, to protect their personal and official email accounts, and we offer them our strongest cyber protections, our Advanced Protection Program.

We also work across the tech industry, including through the Tech Accord that you mentioned, Chair Warner, and the Coalition for Content Provenance and Authenticity, the C2PA group I mentioned, to identify emerging challenges and to counter abuse.

We are committed to doing our part to keep the digital ecosystem safe, reliable, and open to free expression.

We appreciate the Committee convening this important hearing, and we look forward to your questions.

[The prepared statement of the witness follows:]

U.S. Senate Select Committee on Intelligence
September 18, 2024

Written Testimony
of Kent Walker
President of Global Affairs, Google & Alphabet

Chair Warner, Vice Chair Rubio, and Members of the Committee, thank you for the opportunity to appear before you today.

My name is Kent Walker, and I serve as President of Global Affairs for Google and Alphabet.

Our business relies on earning the trust of our users. And we take seriously the importance of protecting free expression and access to a range of viewpoints, while also maintaining and enforcing responsible policies.

A critical part of that responsibility is doing our part to protect the integrity of democratic processes around the world. That's why we have long invested in cutting-edge capabilities, strengthened our policies, and introduced new tools to address threats to election integrity.

We recognize the importance of enabling the people who use our services – in America and abroad – to speak freely about the political issues most important to them. At the same time, we continue to take steps to prevent the misuse of our tools and platforms, particularly attempts by foreign state actors to undermine democratic elections.

I. 2024 Election: Protecting Our Users by Disrupting Foreign Threats

This year, more than 50 national elections — including the U.S. Presidential election — are taking place around the world. With each election cycle, we apply new learnings to improve our protections against harmful content and create trustworthy experiences.

Furthering our commitment to protect elections, we created the Google Threat Intelligence group, which builds on our Threat Analysis Group, or TAG, and Mandiant Intelligence. Google Threat Intelligence helps identify, monitor, and tackle threats ranging from coordinated influence operations to cyber espionage campaigns across the Internet. TAG tracks and works to disrupt more than 270 government-backed attacker groups from more than 50 countries and publishes its findings each quarter. Mandiant similarly shares its findings on a regular basis, and has published more than 50 blogs to date this year alone analyzing threats from Russia, China, Iran, North Korea, and the criminal underground.

In the current election cycle, we have seen a variety of malicious activity, including cyber-attacks, efforts to compromise personal email accounts of high-profile political actors, and influence operations both on and off our platforms —many of which seek to sow discord among Americans. I describe some of these efforts in more detail below.

A. Combating Threats Posed by APT42

Associated with Iran's Islamic Revolutionary Guard Corps (IRGC), the group known as Advanced Persistent Threat (APT) 42 consistently targets high-profile users, including current and former government officials, political campaigns, and diplomats, as well as think tanks, NGOs, and academic institutions that contribute to foreign policy conversations. In the past six months, roughly 60 percent of APT42's known attacks have been directed against U.S. and Israeli targets, including former senior Israeli military officials and individuals affiliated with both U.S. presidential campaigns. During the 2020 U.S. presidential election cycle, we disrupted APT42 attempts to target accounts associated with the Biden and Trump presidential campaigns. These activities reflect the group's aggressive, multi-pronged effort to quickly alter its operational focus in support of Iran's shifting political and military priorities.

In the current U.S. presidential election cycle, TAG has detected and disrupted a small but steady cadence of APT42's Cluster C credential phishing activity. In May and June of 2024, APT42 targets included the personal email accounts of approximately a dozen individuals variously affiliated with President Biden or former President Trump, including current and former officials in the U.S. government and individuals associated with the two campaigns. We blocked numerous APT42 attempts to log in to personal email accounts of targeted individuals.

Recent public reporting shows that APT42 has breached accounts across multiple email providers, and we saw that the group successfully gained access to the personal Gmail account of a high-profile political consultant in June 2024. In addition to quickly securing the compromised account and sending government-backed attacker warnings to all of the targeted accounts, we proactively referred this malicious activity to law enforcement in early July, and we are continuing to cooperate with them on this matter. At the same time, we informed officials from both campaigns that we were seeing heightened malicious activity originating from foreign state actors and underscored the importance of using enhanced account security protections on personal email accounts.

TAG continues to observe unsuccessful attempts from APT42 to compromise the personal accounts of individuals affiliated with President Biden, Vice President Harris, and former President Trump, including current and former government officials and individuals associated with the campaigns.

APT42's efforts to target the U.S. presidential election are of course not limited to Google products. As documented in recent public reporting, the group has successfully breached accounts across multiple email providers and we believe this activity is ongoing. TAG has notified other service providers of this malicious activity so that they can take appropriate action on their platforms. We will continue to monitor developments and share findings with industry peers as we uncover additional activity.

B. Countering Russian-State-Controlled Influence Operations

YouTube offers many millions of channels of content and billions of videos. Since Russia invaded Ukraine, YouTube has blocked thousands of channels and millions of videos from Russian state-sponsored organizations, including channels directly tied to RT (formerly Russia Today) and Sputnik. In 2024, we terminated more than 11,000 YouTube channels linked to coordinated influence operations with ties to Russia. We also continue to terminate channels belonging to Russian entities and individuals subject to U.S. government sanctions.

Most recently, following a Department of Justice indictment issued on September 4 regarding covert Russian support to Tenet Media, and after careful review to verify violations of YouTube's Community Guidelines, we terminated Tenet Media's channel and other channels owned or operated by its owners.

Last week, the U.S. Department of State sanctioned RT for engaging in both direct disinformation and covert influence operations. These recent developments highlight the importance of receiving information from law enforcement, government, and other trusted flaggers, which add to the signals we can observe about activity on our platforms.

Finally, over the last two years, the Russian government has periodically throttled access to YouTube. In the last two months, we have seen more frequent efforts to throttle and even block YouTube in Russia. YouTube has long been one of the last remaining sources of independent media inside Russia, and has refused to comply with a number of Russian government demands to remove political speech and similar content.

C. Disrupting DRAGONBRIDGE Activity

DRAGONBRIDGE, also known as "Spamouflage Dragon," is an influence network linked to the People's Republic of China that has a presence across multiple platforms. While the majority of DRAGONBRIDGE activity remains low-quality content without a political message, a small fraction of DRAGONBRIDGE accounts post about current events with messaging that supports pro-PRC views. DRAGONBRIDGE content has also featured U.S political issues and figures, particularly in the lead-up to elections, and is often presented as short "news" clips.

In the lead-up to the 2022 U.S midterm elections, Google terminated channels in which DRAGONBRIDGE attempted to spread narratives highlighting U.S. political divisions, potential for political violence, and threats to democracy. For example, one video attempted to portray voting in the U.S. as ineffective and a waste of time. The activity extended across platforms, with DRAGONBRIDGE posting similar messages via tweets and identical video content on Twitter. In 2023, we observed pro-PRC campaigns conducting operations that targeted the upcoming 2024 U.S. presidential election. These campaigns used inauthentic accounts (positioned as coming from U.S. citizens or social movements) to promote partisan content on polarizing issues in American and global politics and to engage with U.S. voters.

In 2024 DRAGONBRIDGE continues to spread narratives highlighting U.S. political divisions and portraying the U.S. government, society, and democracy in a negative light, cycling through political and social narratives that evolve with the headlines. In May 2024, for example, DRAGONBRIDGE began uploading videos and commenting on the student protests over the Israel-Hamas war on U.S. university campuses. DRAGONBRIDGE content appeared in English, was generally pro-Palestine in its themes, and used the student protests to frame the U.S. and Western media as hypocritical.

This year, we terminated more than 22,000 YouTube channels linked to Chinese coordinated influence operations, as we publicly shared in the first quarter, second quarter, and third quarter of 2024. Though it is evident that substantial resources are being expended around pro-PRC operations, these efforts do not appear to be gaining significant traction. When we have observed them spinning up activity across platforms, we have been able to stop them relatively quickly. Google Threat Intelligence is actively monitoring DRAGONBRIDGE activity for any shifts in tone or focus related to the U.S. presidential election.

II. Securing Our High-Risk Users and Election Infrastructure

Understanding the patterns and trends of threat actors informs our approach to keeping all users and their personal information safe — and this is especially important for high-risk users during election cycles. We recommend our Advanced Protection Program — our strongest set of cyber protections — for all high-risk individuals, including elected officials, candidates, campaign workers, journalists, and election workers.

We have also expanded our longstanding partnership with Defending Digital Campaigns (DDC) to give U.S. campaigns the security tools they need to stay safe online, including tools to rapidly configure Google Workspace's security features. We encourage campaigns that are Workspace customers to enroll in Workspace for Campaigns, a free one-click feature to immediately configure 26 core security settings for an entire team. This feature is available to all campaigns eligible for support from Defending Digital Campaigns.

In 2023, through partners like DDC, we distributed 100,000 free Titan Security Keys to high-risk users. This year we have committed to providing an additional 100,000 updated versions of these security keys. Additionally our Campaign Security Project has helped train more than 9,000 campaign and election officials across the American political spectrum in digital security best practices. In the EU, we are proud to partner with PUBLIC, The International Foundation for Electoral Systems (IFES), and Deutschland sicher im Netz (DSIN) to scale account-security training and to provide security tools.

Additionally, we encourage all eligible websites supporting the election to sign up for Project Shield to increase stability during the election cycle. Project Shield helps protect against both distributed denial of service (DDoS) attacks and legitimate traffic surges, and provides free protection for websites that host information on political candidates, voting, poll monitoring, and any other websites supporting the election process.

Further, since 2014 Mandiant has provided trusted cybersecurity capacity, capability, and expertise to state and local governments through its professional and managed services. It is an active partner in CISA's Joint Cyber Defense Assistance Collaborative (JCDC) 2024 Election Cyber Defense Plan and is also supporting election security webinars for state and local U.S. election officials to help them understand their cyber threat landscape and improve their awareness of the tools and resources available to help harden election infrastructure from cyber attacks. Mandiant provides various services helping stakeholders harden and test their defenses and monitor, respond to, and recover from cyber threats, including:

- **Cyber Threat Diagnostic:** Mandiant helps customers understand their threat profile – who is targeting them, why they are being targeted, what assets the threats are targeting and how – by analyzing evidence of threat activity within their environment.
- **Exercises and Red Teaming:** Mandiant provides intelligence-informed tabletop exercises, simulations, and red teaming services to help customers validate their incident response readiness and ability to respond to real-world attacks.
- **Managed Defense and Incident Response:** Mandiant offers solutions for 24/7 overwatch and incident response expertise on-demand or on-site.
- **Proactive Threat Hunting:** Mandiant uses tailored intelligence to identify indicators of compromise as well as advanced anomalous precursors to attacks.
- **Critical Asset Protection and Attack Surface Management:** Mandiant services help customers identify their critical assets, map their entire environment, and ensure the integrity of their critical systems.
- **After-Action Reviews and Gap Identification:** After elections have concluded, Mandiant offers customers a summary of observed activities, tailored recommendations for cybersecurity improvements, and a catalog of prioritized technology gaps to remediate before the next election cycle.

III. Mitigating Risks Posed by Generative AI in the 2024 Elections

We remain on the look-out for new tactics and techniques in both cyber-security and disinformation campaigns. We are seeing some foreign state actors experimenting with generative AI to improve existing tactics, like more efficiently creating fake websites, misleading news articles, and robotic social media posts. We have not yet seen AI bring about a sea change in these tactics, but we may not always be able to see the full scope of nefarious activity, and we remain alert to new vectors of attack.

A. Empowering Users to Navigate AI-Generated Content

To combat the risks posed by AI-generated content in the context of elections, we have put in place new tools and policies and entered into partnerships with key global stakeholders.

- **AI Prohibited Use Policy:** Drawing on our experience in policy development and technical enforcement, we have created generative AI prohibited use policies outlining the types of harmful, inappropriate, misleading, or illegal content that is not allowed on our systems. We then use our extensive system of classifiers to detect and remove content that violates these policies.

- **AI Ads Disclosures:** We have long had policies against deceptively manipulated media. And last year, we were the first tech company to launch new disclosure requirements for election ads containing synthetic content. We require that federal election advertisers prominently disclose when their ads contain synthetic content that inauthentically depicts real or realistic-looking people or events. This disclosure must be clear and conspicuous, and placed in a location where users are likely to see it. This policy applies to image, video, and audio content. Ads that contain synthetic content altered or generated in such a way that is inconsequential to the claims made in the ad are exempt from these disclosure requirements. This includes editing techniques such as image resizing, cropping, color or brightening corrections, defect correction (for example, "red eye" removal), or background edits that do not create realistic depictions of actual events.

- **YouTube AI Content Labels:** We seek to give viewers relevant context about the content they watch. In mid-March, YouTube also began requiring YouTube creators to disclose when they upload realistic content – content a viewer could easily mistake for a real person, place, or event – made with altered or synthetic media, including with generative AI. We apply transparency labels to signal to users that they are watching this type of content. We apply these labels automatically for content created with certain YouTube generative AI features, like Dream Screen. For most videos, a label will appear in the expanded description, but for videos that touch on more sensitive topics — like elections, health, news, or finance — we will also show a more prominent label on the video itself.

- **Election Responsibility and Generative AI**: Last December we announced that our Gemini AI App and Search products would not provide responses for election-related prompts during the 2024 elections. As we integrate Gen AI into more consumer experiences, we are also applying election-related restrictions to many of these products, including Search AI Overviews, YouTube AI-generated summaries for Live Chat, Gems, and image generation in Gemini. Our users often use Google to get reliable and up-to-date information on topics like current candidates, voting processes, and election results — and this new technology can make mistakes as it learns or as news breaks, so we want to implement it cautiously. For many queries and prompts on Gemini, we also provide a link connecting users directly to Google Search for links to the latest and most accurate information.

- **Additional Context Features:** The About this Image feature in Search helps people assess the credibility and context of images they see online, and we recently expanded this feature to cover even more surfaces and languages where users might encounter content about which they have questions. And our double-check feature in Gemini evaluates whether there is content across the web to substantiate its responses to user prompts.

- **Digital Watermarking:** Last year we introduced SynthID, a tool that adds imperceptible watermarks to our AI-generated images and audio so that they are easier to identify. This year, we expanded SynthID to two new modalities: text and video. We are also expanding our work on identifying the provenance of AI-generated content created on other platforms through the Coalition on Content Provenance and Authenticity, as described below.

B. Working Across Industry and the U.S. Government to Address Risks Posed by GenAI in Elections

Election integrity is a shared challenge. Although we design and enforce our policies independently, we have received information for many year from national security agencies and federal, state, and local law enforcement, as well as from a range of trusted flaggers, who may have access to information and intelligence about malicious activity, including from foreign adversaries, that we do not.

Further, we have long taken a principled and responsible approach to introducing Generative AI products. And we recognize the importance of collaborating across the tech industry – including through the Tech Accord and the Coalition for Content Provenance and Authenticity – to identify emerging challenges and counter abuse.

i. Tech Accord

Earlier this year, we were proud to sign on to the Tech Accord to Combat Deceptive Use of AI in 2024 Elections, a set of commitments to deploy technology countering harmful AI-generated content meant to deceive voters. We pledged to help prevent deceptive AI-generated image, audio, or video content from interfering with this year's global elections.

As described in greater detail above and in a recent update on the Tech Accord website, we have taken a number of steps across our products to reduce the risks that intentional, undisclosed, and deceptive AI-generated imagery, audio, or video may pose to the integrity of electoral processes. We have taken steps to develop technologies, assess models, detect distribution, and appropriately address deceptive AI election content.

In line with our Tech Accord Commitments, we are also continuing our efforts to foster cross-industry resilience, provide transparency to the public, and engage with civil society. We actively share our learnings and expertise with researchers and others in the industry. These efforts include increasing public awareness by, for example, actively publishing and updating our approach to AI, our research into provenance solutions, and our approach to content labeling.

Artificial intelligence innovation raises complex questions that neither Google, nor any other single company, can answer alone. Google continues to engage and collaborate with a diverse set of partners including the Partnership on AI, ML Commons, and is a founding member of the Frontier Model Forum, an initiative to help share safety best practices and inform collective work on AI safety. We look forward to continuing to engage with stakeholders and doing our part to advance the AI ecosystem.

ii. Coalition for Content Provenance and Authenticity, or C2PA

In addition to our Tech Accord commitments, we joined the Coalition for Content Provenance and Authenticity (C2PA) as a steering committee member. The C2PA is a cross-industry effort to help provide more transparency and context regarding AI-generated content. Google has worked alongside the other members to develop and advance the technology used to attach provenance information to content.

Through the first half of the year, we collaborated on the newest version (2.1) of the technical standard, Content Credentials. This version is more secure against a wider range of tampering attacks due to stricter technical requirements for validating the history of the content's provenance, which will help ensure the data attached is not altered or misleading. We will soon bring the latest version of Content Credentials to certain key products like Search and Ads, and we will continue to expand its application to more products over time. We also encourage more services and hardware providers to adopt the C2PA's Content Credentials standard.

* * *

We are committed to doing our part to keep the digital ecosystem safe and reliable. We appreciate the Committee convening this important hearing, and we look forward to answering your questions.

OPENING STATEMENT OF BRAD SMITH VICE CHAIR AND PRESIDENT, MICROSOFT

Mr. SMITH. Thank you, Chairman Warner, and thank you, Vice Chairman Rubio. It is a pleasure to be here.

I first want to say, many days, we are competitors, but I think when it comes to protecting the American public, all three of us and all of us across the tech sector are and need to be colleagues committed to a common cause of protecting our elections.

I think we have to start by recognizing that there are real and serious threats, including in this election. We all have all been reporting on them, we have been seeing them, and you have talked about them.

Every day, we know that there is a Presidential race between Donald Trump and Kamala Harris; but this has also become an election of Iran versus Trump and Russia versus Harris. It is an election where Russia, Iran, and China are united with a common interest in discrediting democracy in the eyes of our own voters and even more so in the eyes of the world.

So, what do we do?

What is the role and responsibility of the tech sector? That is the fundamental question you have put to us.

First, I think we should always adhere to two principles. The first is to preserve the fundamental right to free expression that is enshrined in our Constitution that Vice Chairman Rubio spoke about. That is and needs to be our North Star.

The second is to defend the American electorate from foreign nation states who are seeking to deceive the American public.

How do we do this?

I think we have three roles. The first is really to prevent foreign nation state adversaries from exploiting American products and platforms to deceive our public. We do that with guardrails, especially around AI-generated content; but we also do it by identifying and addressing content on our platform—especially AI-generated content created by foreign States.

I think our second role is to protect candidates the people who are putting themselves out there to run for office, their campaign staffs, the political parties, the county and State election officials, on which we all rely. And we do that in part by providing them with technology and knowhow. Google, Microsoft, we all do that, and we do it by getting out there and working with them.

At Microsoft, we have now worked across 23 countries this year. We have had more than 150 training sessions reaching more than 4,700 people. And we do it by responding immediately in real-time when incidents arise, as we do, to work with campaigns to help protect them.

And the third role we play, quite possibly the most important, is to build on your leadership in having this hearing to prepare the American public for the risks ahead.

We do that by informing them, encouraging them to check what they see, to recheck it before they vote. And we do it by I think recognizing that there is a potential moment of peril ahead.

Today, we are 48 days away from this election, as you said, Chairman Warner. The most perilous moment will come I think 48

hours before the election. That is the lesson to be learned from, say, the Slovakian election last fall and other races we have seen.

I think above all else, even in a country that has so many divisions, I do hope we can all remember one thing: If Google and Microsoft and Meta can get together, if Republicans and Democrats and Independents can work together, then I think we have an opportunity as a country to stand together to ensure that we, the people of the United States, will choose the people who lead us and we will protect ourselves from foreign interference and deception.

Thanks very much.

[The prepared statement of the witness follows:]

Securing US Elections from Nation-State Adversaries

Written Testimony of Brad Smith
Vice Chair and President, Microsoft Corporation

U.S. Senate Select Committee on Intelligence

September 18, 2024

Chairman Warner, Vice Chairman Rubio, Members of the Committee, I appreciate the opportunity to join you and other technology leaders today to discuss the timely and critical issue of protecting US elections from nation-state interference.

Today we are 48 days away from the general election; in some states like Pennsylvania, voters have already begun casting ballots, and three days from now all 50 states will send ballots to military and overseas voters. The election is here, and our adversaries have wasted no time in attempting to interfere. Earlier this week, Microsoft's Threat Analysis Center (MTAC) reported efforts by our adversaries to interfere in our elections leveraging both old and new tactics. Earlier this month the United States Government sanctioned[1] Russian actors for their attempts to influence the election.[2]

The threats to our democracy from abroad are sophisticated and persistent. We must stand together as a tech community, as leaders, and as a nation to protect the integrity of our elections. We pursue this work guided by two key principles:

1. We must uphold the foundational principle of free expression for our citizens.
2. We must protect the American electorate from foreign nation-state cyber interference.

Our adversaries target our democracy in part because they fear the open and free expression it promotes and the success it has brought our country.

Current State of Nation-State Interference

Among Microsoft's vast team of security professionals, dozens are part of Microsoft's Threat Analysis Center (MTAC), a team whose mission is to detect, assess, and disrupt cyber influence threats to Microsoft, its customers, and democracies worldwide. Part of MTAC's mission is protecting elections from nation-state adversaries who seek to use online operations to distort the information going to voters, change the outcome of an election, or interfere in electoral processes.

As MTAC has observed and reported, foreign adversaries are using cyber influence operations to target both political parties in the 2024 U.S. presidential election. In the last two years, Microsoft has detected and analyzed cyber-attacks and cyber-enabled influence operations stemming from Russia, Iran, and China, many of which pertain to elections and elections infrastructure.

---

[1] Treasury Takes Action as Part of a U.S. Government Response to Russia's Foreign Malign Influence Operations | U.S. Department of the Treasury

[2] Office of Public Affairs | Justice Department Disrupts Covert Russian Government-Sponsored Foreign Malign Influence Operation Targeting Audiences in the United States and Elsewhere | United States Department of Justice

This follows similar activity Microsoft has observed in several other countries that recently have held national elections. This includes the 2023 elections in the Netherlands and Slovakia and, in 2024, the Taiwanese, EU, UK and French elections (as well as the 2024 Paris Summer Olympics). Since the beginning of this year, we have been working directly with elected government officials and often with the public to combat these threats. We have used our findings to better understand adversarial behavior and intentions leading into the upcoming 2024 U.S. election, including with respect to nation states' malicious employment of generative AI, of which we have detected and analyzed many such instances.

Today, we see Iran, Russia, and China using cyber operations to target the U.S. election in November. Iranian operations have targeted candidates of both parties but are inclined to denigrate former President Trump's campaign, which indicates a preference for a Harris victory. Russian operations, meanwhile, are inclined to denigrate Vice President Harris's campaign, indicating a preference for a Trump victory. China, for its part, has aimed to collect intelligence and to stoke discord, while to date not showing a clear preference for a specific candidate.

Let me share more about the details in what we have detected so far this year:

Iran

So far in 2024, Iranian election interference mirrors what we observed from Iran in 2020 in tempo, timing, and targets. As we reported in an August 8 report,[3] an Iranian actor we track as Sefid Flood, known for impersonating social and political activist groups, started in March to lay the groundwork for U.S. election cyber-enabled operations. Additionally, Iranian-linked covert propaganda sites and social media networks began and have continued to aim to amplify divisions among Americans across ethnic and religious lines.

In June 2024, Microsoft observed an Iranian actor tracked as Mint Sandstorm compromised a personal account linked to a U.S. political operative. Mint Sandstorm used this access to the political operative's account to conduct a spear phishing attack on a staff member at a U.S. presidential campaign. Microsoft products automatically detected and blocked this phishing email. Microsoft took additional steps to notify the political operative and the campaign of this activity. Last month, Microsoft detected that Mint Sandstorm compromised additional personal accounts belonging to individuals linked to a U.S. presidential candidate. Microsoft quickly took action to notify these users and assist them in securing their accounts. We expect the pace and persistence of Iran's cyberattacks and social media provocations will quicken as Election Day approaches in November.

Iran has a history of targeting voters in U.S. swing states. In 2020, an IRGC-directed group, Cotton Sandstorm, posed as the right-wing "Proud Boys" to stoke discord in the U.S. over purportedly fake votes. Using a Proud Boys-named email, Cotton Sandstorm sent emails to Florida residents warning them to "vote for Trump or else!"[1] Cotton Sandstorm's cyber activity ahead of the operation included scanning of at least one government organization in Florida.

In 2022, ahead of the midterm elections, Microsoft detected Mint Sandstorm targeting county-level government organizations in a few states, a pair of which were tightly contested states in 2020. Similarly, in 2024, we've observed another group operating on the IRGC's behalf, Peach Sandstorm, successfully access an account at a county government in a tightly contested swing state.

---

[3] Iran Targeting 2024 US Election - Microsoft On the Issues

We do not know if the IRGC's targeting of swing states in 2022 or 2024 was election related; in fact, Peach Sandstorm's targeting was part of a large-scale password spray operation. That said, Iran appears to have demonstrated an interest in U.S. swing states for potential follow-on operations similar to the one ahead of the 2020 elections that sought to sow discord on our electoral process.

Russia

Russian threat actors, the most notable adversary in previous U.S. election cycles, currently are spoofing reputable media outlets and distributing staged videos to spread the Kremlin's preferred messages to U.S. voters online. In some cases, these campaigns gain a significant number of views and sizeable reach among U.S. and international audiences.

For example, in early May, Microsoft observed a Russia-affiliated influence actor we track as Storm-1516 disseminate a staged video that claimed to show Ukrainian soldiers burning an effigy of former President Trump. The fake video received some international press, which inaccurately covered the video as genuinely originating from Ukraine. The video was reposted across social media and received several million impressions.

Later, after Vice President Harris joined the presidential race, our team saw Storm-1516 pivot its campaigns. In a second video staged in a Storm-1516 operation in late August, several people who are depicted as Harris supporters are shown assaulting an alleged supporter of former President Trump. This video received at least five million impressions. In a third staged video released earlier this month, Storm-1516 falsely claimed that Harris was involved in a hit-and-run incident. This video similarly gained significant engagement, the original video reportedly receiving more than two million views in the week following its release.[4]

We also anticipate that Russian cyber proxies, which disrupted U.S. election websites during the 2022 midterms,[5] may seek to use similar tactics on Election Day in November 2024. In addition to the Russian cyber proxy "RaHDit," which the U.S. State Department recently revealed as led by Russian intelligence,[6] Microsoft tracks nearly a dozen Russian cyber proxies that regularly use rudimentary cyberattacks to stoke fear in election and government security on social media.

In our August 9 elections report, we revealed a Russian actor that we track as Volga Flood (also known as Rybar) and their efforts to infiltrate U.S. audiences by posing as local activists.[7] Volga Flood created multiple social media accounts called "TexasvsUSA." The accounts post inflammatory content about immigration at the Southern border and call for mobilization and violence. This month, we've seen Volga Flood shift its focus to the Harris-Walz campaign, posting two deceptively edited videos of Vice President Harris on social media.

Volga Flood is publicly positioned as an anonymous military blogger covering the war in Ukraine. In reality, however, Volga Flood is a media enterprise employing dozens of people and headed by EU-sanctioned Russian national Mikhail Zvinchuk. Volga Flood's media enterprise is divided across multiple teams that include monitoring, regional analytics, illustration, video, foreign languages,

---

[4] https://uk.news.yahoo.com/russia-spread-fake-rumour-kamala-153333198.html
[5] https://www.usatoday.com/story/news/politics/elections/2022/11/08/2022-midterm-websites-mississippi-hit-cyber-attack/8308615001/
[6] https://www.state.gov/u-s-department-of-state-takes-actions-to-counter-russian-influence-and-interference-in-u-s-elections
[7] https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/5bc57431-a7a9-49ad-944d-b93b7d35d0fc.pdf

and geospatial mapping—all to fulfill its mission statement of waging information warfare on behalf of the Kremlin. Volga Flood publishes analyses through dozens of social media brands and establishes and runs covert social media accounts.

Two additional Russian actors MTAC tracks have largely focused on European audiences but at times shift to U.S. electoral influence. Since March 2022, we have seen the Russian threat actor we track as Ruza Flood, known internationally as "Doppelganger," attempt to undermine U.S. politics. Ruza Flood receives significant resourcing and direction from the Russian Presidential Administration.[8] The U.S. Justice Department, in its September 4 announcements, revealed Ruza Flood's efforts to influence the U.S. citizenry through projects like the "Good Old USA Project," "The Guerilla Media Campaign," and the "U.S. Social Media Influencers Network Project."[9]

Finally, Storm-1679, a Russian influence actor previously focused on malign influence operations targeting the 2024 Paris Olympic Games, has recently shifted its focus to the U.S. presidential election.[10] Storm-1679 routinely creates videos masquerading as reputable news services or impersonating international intelligence agencies, including France's DGSI and the U.S.'s CIA. Storm-1679 recently pivoted to creating videos sowing conspiracies about Vice President Harris, which the actor distributes across a range of social media platforms.

Microsoft's current tracking of current Russian influence operations targeting elections extends beyond the U.S. presidential election. We are also seeing efforts to influence the upcoming Moldovan presidential election and EU referendum on October 20, 2024. In Moldova, a longstanding target of Russian strategic influence campaigns, we currently observe pro-Kremlin proxy activity aimed at achieving Moscow's goal of destabilizing democratic institutions and undermining pro-EU sentiment. We and others expect Russia will leverage an array of techniques in Moldova: political influence, electoral interference, cyberattacks, sabotage, and cyber-enabled influence campaigns that promote pro-Kremlin political parties and denigrate the current Moldovan leadership.

Microsoft is working in collaboration with the Moldovan government and others to assist in identifying and defending against Russian cyber and influence operations seeking to influence the outcome of these two elections.

China

Chinese actors' election efforts are more extensive in 2024 than in previous U.S. election cycles. We observe Chinese influence actors spreading politically charged content over covert social media networks, pretending to be U.S. voters and polling Americans on divisive social issues. Chinese actors have also posed as student protestors online, seeking to stoke division over conflict in the Middle East. These fake accounts—masquerading largely as U.S. conservative voters but also a handful of progressive personas as well—frequently ask their followers whether they agree with a political topic or political candidate. This tactic may be for reconnaissance purposes to better understand how Americans view nuanced political issues.

This messaging style may also be part of a broader engagement strategy: Over the past year, these China-linked personas have conducted more tailored audience engagement than observed

---

[8] https://www.justice.gov/opa/media/1366261/dl

[9] https://www.justice.gov/opa/pr/justice-department-disrupts-covert-russian-government-sponsored-foreign-malign-influence

[10] https://blogs.microsoft.com/on-the-issues/2024/06/02/russia-cyber-bots-disinformation-2024-paris-olympics/

previously, replying to comments, tagging politicians and political commentators, and creating online groups with likeminded voters. Their content strategy has shifted as well. Rather than producing original infographics and memes that largely failed to resonate with U.S. voters in the past cycle, these personas are creating simple short-form videos taken and edited from mainstream news media. Clips denigrating the Biden administration have successfully reached hundreds of thousands of views.

In July 2024, Microsoft responded to a cyberattack on an organization supporting the upcoming U.S. presidential election. Microsoft worked to remediate and secure the organization's infrastructure. Subsequent investigation and analysis has attributed this attack to a state affiliated actor based in China.

These examples, as well as others, underscore the ways in which Iranian, Russian, and Chinese influence actors may seek in the next two months to use social divisions and digital technology to further divide Americans and sow discord ahead of this November's election. We also need to be vigilant in combatting the risk that nation-state adversaries will seek to conduct cyberattacks directly on key American entities that play critical roles in these elections. More information on these actors can be found in our most recent MTAC report.

Deceptive use of synthetic media (deepfakes)

AI is a tool among many tools that adversaries may opt to leverage as part of a broader cyber influence campaign. As we have navigated through the numerous global elections this year, the emergence of AI as a means for interference has presented itself so far this year as less impactful than many had feared. We recognize, however, that determined and advanced actors will continue to explore new tactics and techniques to target democratic countries, which will include additional and improved use of AI over time.

Though we have not, to date, seen impactful use of AI to influence or interfere in the U.S. election cycle, we do not know what is planned for the coming 48 days, and therefore we will continue to be vigilant in our protections and mitigations, against threats both traditional and novel.

As a leading technology company heavily invested in AI, we recognize our important responsibility to implement proactive measures to counter these risks. This includes developing robust frameworks for detecting and responding to deceptive AI election content, enhancing transparency within AI applications, and fostering international collaboration to protect the democratic process. The future of our elections depends on our collective ability to utilize AI responsibly and ethically.

In response to these challenges, we have taken significant steps, including joining together with twenty-seven of the world's largest technology companies this year to sign the Tech Accord to Combat Deceptive Use of AI in 2024 Elections.[11] This accord addresses abusive AI-generated content through eight specific commitments, categorized into three pillars: Addressing Deepfake Creation, Detecting and Responding to Deepfakes, and Transparency and Resilience. It represents one of the important steps the tech sector has taken this year to protect our elections, and we appreciate the encouragement and support of this Committee to be more proactive, including through Chairman Warner's presence and voice at the launch of this accord at the Munich Security Conference in February.

[11] AI Elections accord - A Tech accord to Combat Deceptive Use of AI in 2024 Elections

Here are some updates on how Microsoft is directly responding to these threats and upholding our joint commitments:

Addressing Deepfake Creation

We recognize that companies whose products are used to create AI generated content have a responsibility to ensure images and videos generated from their systems include indicators of their origin. One way to accomplish this is through content provenance, enabled by an open standard created by the Coalition for Content Provenance and Authenticity (C2PA).[12] Microsoft is a founding member of C2PA and has leveraged this standard ("content credentials") across several of our products, ensuring that AI generated content is marked and readable.

Specifically, Microsoft has added content credentials to all images created with our most popular consumer facing AI image generation tools, including Bing Image Creator, Microsoft Designer, Copilot, and in our enterprise API image generation tools via Azure OpenAI. We recently started testing a content credentials display in Word. When images with content credentials are inserted into Word documents, future viewers will be able to right click and see the credits and author information of these images. In addition, C2PA tagged content is starting to be automatically labeled on LinkedIn.[13] The first place you'll see the content credentials icon is on the LinkedIn feed, and we'll work to expand our coverage to additional surfaces.

As important as it is to mark content as AI generated, a healthy information ecosystem relies on other indicators of authenticity as well. This is why in April we announced[14] the creation of a pilot program that allows political campaigns in the U.S. and the EU, as well as elections authorities and select news media organizations globally, to access a tool that enables them to easily apply content provenance standards to their own authentic images and videos.

We also joined forces with fellow Tech Accord signatory, TruePic[15] to release an app that simplifies the process for participants in the pilot. This app has now launched for both Android and Apple devices and can be used by those enrolled in Content Credentials program.

Detecting and Responding to Deepfakes

Microsoft is harnessing the data science and technical capabilities of our AI for Good Lab and MTAC teams to better assess whether abusive content—including that created and disseminated by foreign actors—is synthetic or not. Microsoft's AI for Good lab has developed and is using detection models (image, video) to assess whether media was generated or manipulated by AI. The model is trained on approximately 200,000 examples of AI and real content. The Lab continues to invest in creating sample datasets representing the latest generative AI technology. When appropriate, we call on the expertise of Microsoft's Digital Crimes Unit to operationalize the early detection of AI-powered criminal activity and respond fittingly, including through the filing of affirmative civil actions to disrupt and deter that activity and through threat intelligence programs and data sharing with customers and governments.

---

[12] Overview - C2PA
[13] (1) LinkedIn Adopts C2PA Standard | LinkedIn
[14] Expanding our Content Integrity tools to support global elections - Microsoft On the Issues
[15] Truepic's Secure Camera Enhances Microsoft's Content Integrity Tools - Truepic

To build on the work of our AI for Good lab, in April we announced[16] that we were joining up with AI researcher, Oren Etzioni[17] and his new non-profit, True Media.[18] True Media provides governments, civil society and journalists with access to free tools that enable them to check whether an image or video was AI generated and/or manipulated. Microsoft's contribution includes providing True Media with access to Microsoft classifiers, tools, personnel, and data. These contributions will enable True Media to train AI detection models, share relevant data, evaluate and refine new detection models as well as provide feedback on quality and classification methodologies.

We are also empowering candidates, campaigns and election authorities to help us detect and respond to deceptive AI that is targeting elections. In February we launched the Microsoft-2024 Elections site[19] where candidates in a national or federal election can directly report deceptive AI election content on Microsoft consumer services. This reporting tool allows for 24/7 reporting by impacted election entities who have been targeted by deceptive AI found on Microsoft platforms.

Transparency and Resilience

In advance of the EU elections this summer, we kicked off a global effort to engage campaigns and elections authorities. This enabled us to deepen understanding of the possible risks of deceptive AI in elections and empower those campaigns and election officials to speak directly to their voters about these risks and the steps they can take to build resilience and increase confidence in the election. So far this year we have conducted more than 150 training sessions for political stakeholders in 23 countries, reaching more than 4,700 participants. This included training and public educations sessions at the Republican and Democratic National Conventions, as well as with state party chairpersons for both major political parties in the United States.

Building on this training, Microsoft also ran public awareness campaigns in the EU ahead of the EU Parliamentary elections,[20] as well as in France[21] and the UK[22] ahead of their national elections. We are now pursuing similar work in the United States ahead of the November general election. This campaign, which is entitled "Check, Recheck, Vote," educates voters of the possible risks posed by deepfakes and empowers them to take steps to identify trusted sources of election information, look for indicators of trust like content provenance, and pause before they link to or share election content. This includes our 'Real or Not?' Quiz, developed by our AI for Good lab to expose users to the challenges of detecting a possible deepfake. So far, individuals from 177 countries have taken the quiz.

Overall, our public awareness campaigns outside the United States have reached more than 350 million people, driving almost three million engagements worldwide. Our U.S. Public Awareness campaign[23] has just begun and already has reached over six million people with over 30,000 engagements.

---

[16] TrueMedia.org to Enhance Deepfake Detection Capabilities - TrueMedia
[17] An A.I. Researcher Takes On Election Deepfakes - The New York Times (nytimes.com)
[18] TrueMedia.org
[19] Microsoft-2024 Elections
[20] Addressing the deepfake challenge ahead of the European elections - EU Policy Blog (microsoft.com)
[21] Microsoft s'engage dans la préservation de la sincérité des élections législatives en France – News Centre
[22] Combating the deceptive use of AI in elections (microsoft.com)
[23] Combating the deceptive use of AI in US elections (microsoft.com)

In May, we announced a series of societal resilience grants in partnership with OpenAI.[24] Grants delivered from the partnership have equipped several organizations, including Older Adults Technology Services (OATS) from AARP, International Institute for Democracy and Electoral Assistance (International IDEA), C2PA, and Partnership on AI (PAI) to deliver AI education and trainings that illuminate the potential of AI while also teaching how to use AI safely and mitigate against the harms of deceptive AI-content.

Protecting Campaign and Election Infrastructure

Since the 2016 election, adversaries have regularly targeted essential systems that support elections and campaigns in the U.S. to advance their cyber enabled influence operations. As mentioned earlier, recent Iranian hacking incidents involved attempts by these actors to provide stolen or allegedly stolen material to the media to propagate narratives of dissent and distrust. This underscores why we continue to invest in efforts that focus on safeguarding the critical infrastructure that underpins our elections.

Our efforts include several initiatives designed to support election officials and political organizations. First, we offer AccountGuard, a no-cost cybersecurity service available to our cloud email customers in 35 countries. This service provides advanced threat detection and notifications against nation-state adversaries for high-risk customers, including those involved in elections. AccountGuard extends beyond commercial customers to individuals at election organizations, their affiliates, and immediate family members who may use personal Microsoft accounts for email. We have observed that sophisticated adversaries often target both professional and personal accounts, amplifying the need for comprehensive protection. More than 5.4 million mailboxes of high-risk users are now protected by AccountGuard globally.

Additionally, our Election Security Advisors program provides federal political campaigns and state election departments with expert security consultation. This includes proactive security assessments or forensic investigations in the event of a cyber incident. Our goal is to ensure that these entities have the necessary support to maintain the integrity of their operations.

For critical election-adjacent systems, such as voter registration databases and voter information portals, we provide our Azure for Election service. This service provides proactive security reviews, resilience assessments, and load analysis. During the election week, we offer our highest tier of reactive support to address any security or availability issues that may arise. Since offering this service from 2018 to today, we have assisted more than half of U.S. states, including many counties and cities, in reviewing their election IT infrastructure.

In preparation for the election this November, we are also establishing a situation room staffed by our team to provide constant management and triage of any election-sensitive issues and maintain real-time communications with other situations rooms across our industry partners. This ensures that any incidents receive the highest level of priority and executive support.

While we continue to provide robust security services, we recognize that collaboration is essential. Public-private partnerships are crucial in strengthening the entire ecosystem. Our Democracy Forward team actively participates in tabletop cybersecurity training exercises with U.S. election officials at both national and state/county levels.

[24] Microsoft and OpenAI launch Societal Resilience Fund - Microsoft On the Issues

Microsoft remains steadfast in its commitment to supporting the security and integrity of democratic processes. Through our comprehensive programs and collaborative efforts, we aim to protect democracy from the evolving threats posed by nation-state actors.

Policy Recommendations

Finally, we find ourselves at a moment in history when anyone with access to the Internet can use AI tools to create a highly realistic piece of synthetic media that can be used to deceive: a voice clone of a family member, a deepfake image of a political candidate, or even a doctored government document. AI has made manipulating media significantly easier, quicker, more accessible, and requiring little skill. As swiftly as AI technology has become a tool, it has become a weapon.

I want to acknowledge and thank this Committee for its longstanding leadership on these important issues. We particularly commend the efforts reflected in section 511 of the SSCI FY 25 Intelligence Authorization Act (IAA), which focuses on protecting technological measures designed to verify the authenticity and provenance of machine-manipulated media. These protections are essential as technology companies strive to provide users with reliable information about the origins of AI generated content.

We are also encouraged and supportive of the recent agreement by the Federal Election Commission (FEC)[25] applying existing restrictions regarding fraudulent misrepresentation to campaigns use of AI technology. Existing robocall provisions are another means of addressing the fraudulent use of synthetic content. These provisions have historically restricted the use of artificial or prerecorded voices and allow for enforcement actions when these rules are violated.

Along those lines, it is worth mentioning three ideas that may have an outsized impact in the future fights against deceptive and abusive AI-generated content.

- First, Congress should enact a new federal "deepfake fraud statute." We need to give law enforcement officials, including state attorneys general, a standalone legal framework to prosecute AI-generated fraud and scams as they proliferate in speed and complexity.
- Second, Congress should require AI system providers to use state-of-the-art provenance tooling to label synthetic content. This is essential to build trust in the information ecosystem and will help the public better understand whether content is AI-generated or manipulated.
- Third, Congress should pass the bipartisan Protect Elections from Deceptive AI Act, sponsored by Senators Klobuchar, Hawley, Coons, and Collins. This important piece of legislation prohibits the use of AI to generate materially deceptive content falsely depicting federal candidates in political ads to influence federal elections, with important exceptions for parody, satire, and the use of AI-generated content by newsrooms. Such legislation is needed to ensure that bad actors cannot exploit ambiguities in current law to create and distribute deceptive content, and to ensure that candidates for federal office have meaningful recourse if they are the victim of such attacks. Several states have proposed or passed legislation similar to this federal proposal. While the language in these bills varies, we recommend states adopt prohibitions or disclosure requirements on "materially deceptive" AI-generated ads or something akin to that language and that the bills contain exceptions for First Amendment purposes.

[25] showpdf.htm (fec.gov)

Conclusion

In conclusion, we recognize that the protection of electoral integrity and public trust is a shared responsibility and a common good that transcends partisan interests and national borders.

This must be our guiding principle.

Looking ahead, we believe that new forms of multistakeholder action are essential. Initiatives like the Paris Call and Christchurch Call have demonstrated positive global impacts by uniting representatives from governments, the tech sector, and civil society. In addressing the challenges posed by deepfakes and other technological issues, it is evident that no single sector of society can solve these complex problems in isolation. Collaboration is crucial to preserving our timeless values and democratic principles amidst rapid technological change.

Thank you for your time and consideration. I look forward to answering any questions you may have.

OPENING STATEMENT OF NICK CLEGG, PRESIDENT OF GLOBAL AFFAIRS, META

Mr. CLEGG. Chairman Warner, Vice Chairman Rubio, distinguished Members of the Committee:

Thank you for the opportunity to appear before you today. At Meta we are committed to free expression. Each day, more than 3 billion people around the world use our apps to make their voices heard. By the end of this year, more than two billion people will vote in elections around the world, and we are proud that our apps help people participate in the civic process.

No tech company delves or invests more to protect elections online than does Meta, not just during peak election seasons, but at all times. We have around 40,000 people overall working on safety and security, and we have invested more than $20 billion on safety and security since 2016.

Meta has developed a comprehensive approach to protect the integrity of elections based on several key principles. First, we have strong policies designed to prevent voter interference and intimidation. Second, we connect people to reliable voting information. Third, we work tirelessly to combat foreign interference and the spread of misinformation. And finally, we lead the industry in transparency for political advertisements.

Our approach reflects the knowledge gained from prior elections and we continue to adapt to stay ahead of emerging challenges. One of the most pressing challenges for the industry is people seeking to interfere with elections to undermine the democratic process. We constantly work to find and stop these campaigns across our platforms. This is an adversarial space, and we are often responding to urgent situations with imperfect information. We may not always get it right, so we need to be cautious, and in each case, we need to conduct our own independent investigation to identify what is and is not interference.

Where we identify coordinated inauthentic behavior, we remove the networks at issue. In fact, we have removed over 200 such networks since 2017, including networks from Russia, Iran, and China. We remain committed to stopping these threats and we are constantly improving and evolving our defenses to stay ahead of our adversaries.

I am pleased to appear beside other industry leaders today, and it underscores an important point. People trying to interfere in elections rarely target a single platform. Cross-industry collaboration and transparency in reporting are essential to tackle these networks across the internet. That is why we publicize our takedowns for all to see and share the relevant information we learn with researchers, academics, and others including, of course, Congress.

This year elections are also taking place as more people are using AI tools. To date we have not seen generative-AI enabled tactics used to subvert elections in ways that have impeded, so far, our ability to disrupt them. However, we remain vigilant and will continue to adapt as the technology does as well.

We know that AI progress and responsibility can and must go hand in hand. That is why we are working internally and externally to address the risks of AI. We have implemented industry

leading efforts to label AI generated content, giving people greater context to what we are seeing. And of course we are working across industry to develop common AI standards.

We are proud to have signed on to the White House's voluntary AI commitments and the Tech Accord to combat deceptive use of AI in 2024 elections, both of which will help guide the industry towards safer, more secure, and more transparent development of AI.

Every election brings its own challenges and complexities. We are confident our comprehensive approach can help protect of not only this year's elections in the United States, but elections everywhere.

Thank you. I look forward to your questions.

[The prepared statement of the witness follows:]

HEARING BEFORE
THE UNITED STATES SENATE
SELECT COMMITTEE ON INTELLIGENCE

September 18, 2024

Testimony of Nick Clegg

I. Introduction

Chairman Warner, Vice Chairman Rubio, and distinguished members of the Committee, thank you for the opportunity to appear before you today. My name is Nick Clegg, and I am the President, Global Affairs at Meta.

At Meta, we are committed to free expression. Each day, more than three billion people around the world use our apps to express themselves and make their voices heard. We want people to be able to talk openly about the issues that matter to them, whether through written comments, photos, music, or other creative media. By the end of this year, more than two billion people will have voted in elections across some of the world's largest democracies, including in the United States, and we are proud that our apps help people participate in the civic process.

We also recognize that adversaries are simultaneously working to interfere with elections through coordinated campaigns across the industry in an effort to undermine the democratic process. No tech company does more or invests more to protect elections online than Meta—not just during peak election seasons, but at all times. We have around 40,000 people working on our overall safety and security, and we have invested more than $20 billion in teams and technology in this area since 2016. Over the years, we have developed a comprehensive approach that sets out policies and safeguards for elections, including identifying threats and fighting manipulation and deception on our platforms.

This year, elections are taking place as more and more people are using artificial intelligence (AI) tools. As a company that has been at the forefront of AI development for more than a decade, we believe that progress and vigilance go hand in hand. We take seriously the concern of generative AI tools being misused during elections. We are committed to transparency in the use of AI, and we are making significant investments to further its responsible use. We are also working externally to address risks, including by signing on to both the White House's voluntary commitments and the Tech Accord to Combat Deceptive Use of AI in 2024 Elections. These commitments will help guide the industry toward safer, more secure, and more transparent development of AI technology, while helping to prevent and address harmful AI-generated content from interfering with elections.

While much of our approach to this year's elections reflects the knowledge we have gained from prior election cycles, we are committed to adapting where needed so that we can stay ahead of new challenges, including those presented by AI.

II. Our Election Integrity Efforts

Providing Access to Reliable Information

We think it is critical for people to be able to find reliable election information from trusted sources. We build free tools to support civic engagement, including ones to encourage eligible voters to register to vote, to remind them of deadlines, and to connect them with non-partisan resources.

We also provide people with election information from their state and local election officials, including during primaries. In the United States, when people search for terms related to the 2024 elections on Facebook and Instagram, they will see links to official information from state and local election officials about how, when, and where to vote.

The success of these efforts is a result of close communication with state and local election authorities, who provide important feedback that enables us to provide up-to-date and nonpartisan information in our decentralized election system. For example, based on feedback from the broader elections community, we developed Voting Alerts, a free tool that allows state and local election offices to broadcast key information to everyone in their jurisdictions. Since launching the initiative in 2020, state and local election officials have sent more than 650 million notifications through Voting Alerts on Facebook.

Prohibiting Harmful Content

In addition to connecting people with reliable voting information, we also prohibit misinformation that is likely to interfere directly with people's ability to vote, including misinformation about the dates, locations, times, and methods for voting; voter registration; and who is eligible to vote. Our policies prohibit calls for voter fraud and coordinated election interference. And we have invested in proactive threat detection and expanded our policies to combat harassment against election officials and poll workers online. In the United States and a number of other countries, we prohibit ads that discourage people from voting, call the legitimacy of an upcoming election into question, or contain premature claims of election victory. We have and will continually review and update these policies with election security in mind.

Promoting Transparency in Advertisements

We provide industry-leading transparency into political advertising on our services. Anyone who places an ad about politics, elections, or social issues must complete an authorization process and disclose who is paying for the ad. We add "paid for by" disclaimers and identify the owner and locations for political Pages and Groups. All political ads are archived in a publicly searchable Ad Library for 7 years, so anyone can see exactly what candidates are saying, who they are targeting, and who paid for it. Today, there are more than 15 million U.S. entries in our Ad Library.

We also require advertisers to disclose when they use AI or other digital techniques to create or alter a political or social issue ad that contains a photorealistic image or video, or

realistic sounding audio, that was digitally created or altered to depict a real person as saying or doing something they did not say or do. It also applies if an ad depicts a realistic-looking person that does not exist or a realistic-looking event that did not happen, alters footage of a real event, or depicts a realistic event that allegedly occurred but is not a true image, video, or audio recording of the event. If we determine that an advertiser has not disclosed the required information, we will reject the ad. Repeated failure to disclose required information may result in penalties against the advertiser. On both Instagram and Facebook, we give people the choice to adjust their Ad Preferences if they want to see fewer ads about social issues, elections, or politics. And in the United States, we prohibit new political, electoral, and social issue ads during the final week of an election.

III. Combating Manipulation and Deception

We know that foreign adversaries try to reach people on our platforms and others before elections, and we remain vigilant in our fight against their evolving tactics. We have made important investments to improve our ability to detect and stop foreign election interference and strengthen the security of our platforms. And we build increasingly sophisticated AI systems so we can proactively and successfully identify these abuses, for example by:

- **Preventing Coordinated Inauthentic Behavior.** We are constantly working to find and stop coordinated campaigns that seek to manipulate public debate across our platforms.

  Our Community Standards prohibit coordinated inauthentic behavior, which is when multiple accounts—including fake and authentic accounts—work together to mislead people. We do not want organizations or individuals creating networks of accounts that mislead people about who they are or what they are doing.

  This is an adversarial space, and we are often acting with imperfect information. We may not always get it right. So we need to be cautious and in each case, we need to conduct our own independent investigation to identify what is—and isn't—foreign interference.

  When we take down these accounts, it is because our investigation has identified deceptive behavior (like using networks of fake accounts to conceal their identity); it is not based on the identity of those behind the account or what they say. We've removed over 200 networks of coordinated inauthentic behavior since 2017, including networks from Russia, Iran, and China. Still, people continue to look for new ways to mislead people, which is why we continue to take steps to make it harder for them to do so.

- **Removing Fake Accounts and Banned Organizations.** One of the ways we identify and stop foreign interference is by proactively detecting and removing fake accounts. We also remove accounts that violate our policies and are not allowed to have a presence on our platform, such as foreign terrorist organizations and those designed under our Dangerous Organizations and Individuals policy.

- **Tackling Misinformation.** We are constantly working to stop the spread of misinformation and disinformation. We have built the largest independent

> fact-checking network of any platform, with nearly 100 partners around the world to review and rate viral misinformation in more than 60 languages. Stories they rate as false are shown lower in Feed. If Pages repeatedly create or share misinformation, we significantly reduce their distribution and remove their advertising rights. We make these efforts regardless of the viewpoint of the content or its author.

I appreciate the opportunity to appear before you today with Microsoft and Google because it underscores an important point about addressing the threats we are facing: people who want to interfere in elections rarely target a single service or platform. Cross-industry collaboration, transparency, and reporting are essential to preventing and discouraging these networks from engaging in harmful conduct across the internet. That is why we publicize our takedowns of coordinated inauthentic behavior for all to see, provide information about them to third parties for their review, and share relevant information with researchers, academics, and others, including the Congress. In 2017, we started publishing detailed reporting on our work to detect and counter security threats on our platforms, known today as our Adversarial Threat Reports. We also publicly release threat indicators we identify on our GitHub platform. Today, we have compiled the largest repository of threat indicators, including more than 6,000 threat indicators of cross-internet activity by Doppelganger, the most persistent Russian foreign influence campaign.

As another recent example, we published our insights on a small cluster of malicious activity on WhatsApp that originated in Iran and appeared to have focused on political and diplomatic officials and other public figures. Our research suggests that these efforts were unsuccessful, and our security teams blocked the behavior after investigating user reports. In an abundance of caution, and given the heightened threat environment ahead of the US election, we also shared information about this malicious activity with law enforcement and the relevant presidential campaigns to encourage them to guard against potential adversarial behavior.

We continually adapt our platforms to make this kind of deception much more difficult and costly. When we conduct a takedown, we identify the tactics used and we build tools into our platforms to make those tactics more difficult at scale. By continuing to develop smarter technologies, enhance our defenses, improve transparency, and build strong partnerships, we are making the constant improvements we need to stay ahead of our adversaries and to protect the integrity of our platforms. We have also learned that we need to be cautious about seeing foreign interference where it is not. As we recently indicated, knowing what we know now, we would have taken different actions on certain issues in 2020.

IV. The Impact of Artificial Intelligence

Meta has been a pioneer in AI development for more than a decade, using machine learning to proactively identify and remove violating content across our services. As with election security, we know that AI progress and responsibility can and must go hand in hand. Generative AI tools offer huge opportunities, and we believe that it is possible and necessary for these technologies to be developed in a transparent and accountable way, while also working to minimize potential risks.

As detailed in a recent Adversarial Threat Report, we have not seen attempts on our apps to

use new generative AI tactics to subvert elections in ways that we could not address through our existing safeguards, specifically by disrupting adversarial networks behind them. However, this does not mean that people are not using AI to try to interfere in elections. To the contrary, adversaries have used different tools, such as AI-generated photos for profile photos on fake accounts, or AI to publish a large volume of fake articles resembling reputable news sources. We recently disrupted a campaign from Russia that was publishing a large volume of stories on fictitious news websites, which our investigation found were likely AI-generated summaries of original news articles. The same campaign also created fictitious journalist personas with generative adversarial network-created profile photos. Our teams found and removed many of these campaigns early, before they were able to build audiences and communities on our services. This shows that our industry's existing defenses already apply to novel generative AI, and are proving effective thus far.

However, we know that we must continue to monitor and assess risks with new technology. That is why we are continually adapting to address new challenges, including by advancing efforts to detect and label AI-generated media. We believe that providing transparency and additional context is the best way to address AI-generated content.

Earlier this year, we announced changes to our approach to identifying and labeling AI-generated organic content. This includes labeling a wider range of video, audio, and image content as "AI info" when we detect industry-standard AI image indicators or when people disclose that they are uploading AI-generated content. If we determine that digitally created or altered image, video, or audio content creates a particularly high risk of materially deceiving the public on a matter of importance, we may add a more prominent label. This approach gives people more information about the content so they can better assess it and appreciate the context if they see the same content elsewhere.

When photorealistic images are created using Meta's AI feature, we take several steps so that people know AI is involved, including putting visible markers on the images, applying "Imagined with AI" labels, and embedding both invisible watermarks and metadata within the image files. Using both invisible watermarking and metadata improves the effectiveness of these markers and helps other platforms identify them. We have been working with others in our industry to develop common standards for identifying AI-generated content through forums like the Partnership on AI (PAI) and the Coalition for Content Provenance and Authenticity. The invisible markers we use are in line with PAI's best practices.

We believe that our current approach represents the cutting edge of what is technically possible right now; however, we continue to pursue a range of options to improve our AI detection capabilities. This work is especially important as this is likely to become an increasingly adversarial space in years to come. People and organizations that actively want to deceive people with AI-generated content will look for ways around the safeguards that are put in place to detect it. Across our industry and society more generally, we will need to keep looking for ways to stay one step ahead.

Importantly, this issue is not unique to Meta and will require a whole-of-industry approach. We have collaborated with experts from technical, policy, media, legal, civic, and academic backgrounds to inform our policy development and processes. We also work closely with companies, such as Adobe, to develop technologies that make it possible for us and other platforms to share with people when they see content that has been AI-generated.

V. Conclusion

While we are conscious that every election brings its own challenges and complexities, we are confident that our comprehensive approach puts us in a strong position to do our part to help protect the integrity of not only this year's elections in the United States, but elections around the globe at all times. We look forward to continuing our work, as well as our collaboration with others in the industry, to drive transparency and counter potentially harmful threats to our democratic process.

Thank you, and I look forward to your questions.

Chairman WARNER. Thank you, gentlemen.

I am going to put up the first two presentations.

Let me add to what Mr. Smith said. I concur that the 48 hours before the election, but I would argue the 48 hours after the polls close, particularly if we have as close an election as we anticipate, could be equally if not more significant in terms of spreading false information, disinformation, and literally undermining the tenets of our democracy.

Now there was a lot of press attention recently on the Department of Justice indictments of the Canadians in Tennessee who were using—paying off influencers, knowingly or unknowingly.

What didn't get much attention is the first slide here, where under the banner of Fox News and the Washington Post—These look exactly like the Washington Post and Fox News. As a matter of fact, it may not be what we thought of as AI, but these are kind of AI techniques to make this so real.

As a matter of fact, they have even got real authors' bylines, and the balance of the ads and other things are totally reflective. This came out of this DOJ indictment. I guess the question in these are—Nick, you mentioned "comprehensive." They appeared on your site. They also appeared on Twitter's site—X's site.

I think it is a real shame that in the previous investigations Twitter was a very collaborative entity. Under X they are absent and some of the most egregious activities are taking place.

But I am not sure any American, even a technology savvy American, is going to figure out that these are fake. So where does that responsibility lie?

Shouldn't your efforts have been able to spot that, and how do we make sure—because only after the fact in 2016, we didn't have real-time numbers of how many Americans were viewing the fake sites, and they literally ended up with hundreds of millions.

I still remember both the Tennessee Republican Party and the Black Lives Matter site. The real sites had less viewership than did the Russian based sites.

How does this get through? How do we know how extensive this is?

And we have many, many more of these.

What are we going do about these in the next 48 hours to make sure Americans are informed to be aware?

Mr. Clegg.

Mr. CLEGG. Well, firstly, Senator, you are absolutely right that it is a hallmark of Russian foreign interference in the democratic process to generate AI stories resembling real media. As it happens, since those appeared on our site, we have just over the last 48 hours banned the organization that spawned a lot of this activity, the disinformation.

Rosia Sovodnia [sic], not least after the editor-in-chief gave an interview where she said publicly—and this is in effect a media organization owned and run out of the Kremlin—that she, and I quote, at least this is the translation, is conducting—her and her team are conducting what she calls "guerrilla projects" in half of American democracy, and the panel behind you is a manifestation of that. That is one of the reasons why—

Chairman WARNER. All right. I want to make sure I get this in. I need to know how many Americans viewed this and other Russian generated Facebook sites that appear to be on your sites. I hope you get that information as soon as possible.

I also want to indicate, there is still an effort, and this is more over here in terms of targeting by the Russians, towards specific groups. In this case it was the Doppelganger gang, and it was both Jewish Americans and then it was targeted towards the Latino community. They are very sophisticated. I guess it is not—wouldn't be jaw dropping that they have focused most of their efforts on the same six States that everybody else is focused on. This again goes more to both Mr. Clegg and Mr. Walker. You know, they are still targeting paid advertising.

We remember in 2016, when we didn't have controls when Russians were paying with rubles for paid advertising on sites. I would have thought 8 years later, we would be better in at least screening the advertising. Again, in the case of YouTube and in the case of Facebook, what are we doing to stop the paid advertising targeting by these adversaries?

Mr. WALKER. You took the last one. I can start on this one.

We have an extensive series of checks and balances in our advertising networks that are designed to identify problematic accounts, particularly around election ads. We require election ads to have registration, effectively.

In the 2016 situation, I remember we did an extensive forensic review of our systems and found that less than $4,000 had been spent on those.

Chairman WARNER. Respectfully, sir.

Mr. WALKER. Senator.

Chairman WARNER. As recently as January, I note the Treasury Department has said that both of your companies have still repeatedly allowed Russian influence actors including sanctioned entities to use your ad tools. We will get that specific information to you.

We are going to really need as soon as possible the content, the bad actors, how much content have they purchased on both of your sites and, frankly, others, and we are going to need that extraordinarily fast because I think they are getting through in many, many more ways than has been represented up here.

Mr. WALKER. I certainly appreciate the concern. And we have taken down, as we indicated earlier, something like 11,000 different efforts by Russian associated entities to post content on YouTube and the like.

Chairman WARNER. We are just going to need this as quickly as possible.

Mr. WALKER. Happy to provide that.

Chairman WARNER. The number of Americans viewing Fox News—what they think is Fox News or Washington Post, or advertisements. We need that data to make sure, again, that we inform the public.

OK. Thank you.

Vice Chairman RUBIO. The area I want to focus on is where political speech is involved, and it is sort of the area I talked about in my opening statement, which is and really in particular, I want to understand what the current policies and practices are as we speak

regarding to content moderation. Just as I am reading and I'm not reading the opening statement from Meta:

We are constantly working to stop the spread of misinformation and disinformation. We have built the largest independent fact-checking network of any platform with nearly 100 partners from around the world, to review and rate viral misinformation in more than 60 languages. Stories that they—this platform or these group of people rate as false are shown lower in feed, and if some page repeatedly creates or shares misinformation, we significantly reduce their distribution and remove their advertising rights.

Let me explain. We are not talking about the stuff that was up here. That is fake content. That is just purely fake content. It is generated to look like Fox News or Wall Street Journal or New York Times. No one is arguing that. That is fake. That should be taken down. Those companies should want them taken down. That is their copyright and their logo and their letterhead.

I'm talking about this. So, you have got a group of people that I think are your fact checkers from all over the world to determine whether something is true or not.

So let me take you back to the real-world scenario which ties into what the CEO of the company said, and that is, there were people at one point saying, "maybe I believe that the pandemic began in a lab. I believe there was an accident in a lab, and it leaked out."

And at one time that was considered not factual. In fact, there was pressure from government officials on companies not to report on that.

How would that work today, a story like that? Who determines whether that is true or not, because it wasn't true then but all of a sudden now it is 50 percent maybe likely. How would something like that—because there were people that were caught up in that. I would imagine that under the policies that you described, if I was out there or someone was out there raising the specter of a potential lab leak, it would run through these fact-checkers, from 100 partners all over the world. They would decide whether it is true or not, and you could have your page diminished, potentially de platformed if I write too much about it.

So how does this policy deal with that problem that I just described, which is a real world one?

Mr. CLEGG. Senator, yes, indeed it is. And as I said in my opening statement, we all—obviously, we all inhabit a world of imperfect information. And crucially, the pandemic was a very good example of that information which changes. And obviously, with the benefit of hindsight, we now understand the epidemiology of the pandemic which we didn't at the time.

When we were in the middle of the pandemic, prior to the vaccines being rolled out, when people were dying, when really no one knew what the trajectory was of this global pandemic. We as an engineering tech firm, of course, we are not specialists in epidemiology.

Vice Chairman RUBIO. Yes, but I'm not asking what happened. I understand what happened. I want to know how this policy today would prevent that from happening, because if the government is telling you this is a lie, "We have proof that it's a lie. Take it down"

and your fact-checkers say it's a lie, then my account gets blocked, gets diminished.

Today, is that happening today, right now?

Mr. CLEGG. So, two things. Firstly, we do continue to rely on these independent fact-checkers. We don't employ them. They are not part of Meta. They are independently vetted by a third-party organization.

Vice Chairman RUBIO. Who are they?

Mr. CLEGG. Oh, there is a variety of organizations which—which specialize in examining what they think is a reliable way of asserting whether something is mis——

Vice Chairman RUBIO. Is there a way to know who those vetters are?

Mr. CLEGG. Oh, yes, absolutely.

Vice Chairman RUBIO. Is there a list somewhere, a roster?

Mr. CLEGG. Yes, We have a full list. Absolutely. We can provide them to you, and they obviously work in multiple languages, in fact, including the United States. I think there are 11 fact checkers in the United States, and we can provide you with all the information of them. That is the first thing.

And the second thing is, and Mark Zuckerberg did indeed explain this in his recent letter to the House Judiciary Committee. I think we learned our lesson, certainly as Meta is concerned, that in the heat of the moment when governments, and it is governments around the world, exert particular pressure on us on particular classes of content which they are particularly focused on, we need to act always—and we strive to do this, but, of course, we make mistakes—we need to act independently; and we need to be resistant to the sort of passing moods and passions around particular bits of content, which was particularly the case in the pandemic. People were, in effect, in a panic.

Vice Chairman RUBIO. Well, let me give you a different context, the exact same system. A laptop appears and 51 people sign a letter saying: We used to work in the intelligence community. This is Russian disinformation. And your fact checkers say: We got to listen to the experts. They would know.

Does anybody—does the New York Post get their account taken down again?

Mr. CLEGG. To be very clear, we did not take down the account or the content. I think X—they are not here but they did.

Vice Chairman RUBIO. But under this policy, if you deem it to not be true because it is disinformation because some guys signed a letter saying that it was, it will lower them in the feed and potentially reduce their distribution, and if they post that story too many times, you may actually lock them out. So that is policy.

Mr. CLEGG. So, in this instance, Senator, you are correct that that story was demoted. I mean, it was always available. Millions of people saw it. But its prominence on our services was temporarily reduced. And we used to do that to allow the fact-checkers to give them the space and time to choose to examine that content.

In this instance, the Hunter Biden story, they didn't do so. So that temporary demotion of a few days was then released, and it was circulated back to normal.

Vice Chairman RUBIO. Did the fact-checkers reduce or demote the 51 people who signed the letter or the letter they signed, because that turned out to be not true?

Mr. CLEGG. I don't believe they did so at the time, no.

Vice Chairman RUBIO. All right. Thank you.

Chairman WARNER. Senator Heinrich.

Senator HEINRICH. All right. So, I want to stay on this same topic of the sort of fraudulent news sites that look like something people would recognize from their own news preferences.

Do each of your companies have a policy of removal once you become aware of something that is clearly a fraudulent version of a legitimate site?

Mr. SMITH. I think the answer is yes, and Vice Chairman Rubio, I thought, captured it very well.

It actually, in my view, does not depend on whether the topic had anything to do with politics. Those are counterfeit sites. Those are people using the trademarks of Fox News and the Washington Post without their permission and in a way that deceives the public and diminishes the value of those companies.

And so, yes. And I think you see pretty universally across the industry in the terms of use that prohibit that.

(Vice Chairman Rubio is now Presiding.)

Senator HEINRICH. Why does it seem to take as long as it does for those sites to be identified and removed?

I think they remain up sometimes longer than I think most of us would hope or expect. And then, have you been able to use AI proactively to identify some of those fake news outlets?

Mr. SMITH. I think increasingly we are using AI to detect these kinds of problems, and I think AI is especially good at detecting the use of AI to create content. That is one of the things we do, and we are able to see things faster. You always have to be in a race.

For example, just this morning, we saw a Russian group put online an AI-enhanced video putting into Vice President Harris's words, at a rally, words she never spoke. So, I think that is one of the goals for all of us to keep pursuing to identify these things faster and then where appropriate take action.

Senator HEINRICH. I am encouraged, because obviously AI is being used offensively and we need to be on our game and responding with those same tools to be able to identify and appropriately deal with these things at a much faster rate.

At a hearing of the U.S. House Committee on House Administration, last week, New Mexico Secretary of State testified that, quote:

"[Y]ears of false election claims and ideological attempts to discredit our voting systems and processes . . . [have] led to . . . increased threats and harassment to election workers."

How have you sought to improve your platform's ability to detect and remove content that actually threatens or harasses people who are part of the democratic process and apparatus for fairly administrating elections?

Mr. WALKER. I will take that, and I suspect the same is true for all of us. There are two elements of that. One is making sure that we are trying to safeguard our election officials against threats that may be posted online. And I am confident that all of our companies have policies against incitements to violence, direct threats, bul-

lying, cyberattacks, et cetera. So that kind of material would come down.

The second half is helping our election officials be more protected themselves through the use of some of the tools that we have spoken about like the Advanced Protection Program, so their information is not being hacked or doxed, et cetera—their personal information is not being made public and the like.

So, between the various companies here, including, I know, our Mandiant Group has worked with a number of election officials and agencies to make them more cyber resilient, if you will—more robust against cyberattack.

Senator HEINRICH. Mr. Clegg.

Mr. CLEGG. Senator, again, I'm sure this is incumbent for all of us represented here, but we also encourage local election officials to use our platforms to communicate with voters. So, we established a system called voting alerts. I think since we established that program in 2020, around 650 million voting alerts have been issued by local and State officials on Facebook's apps and services so that voters are properly informed about where and when to vote.

Mr. HEINRICH. I am going to give the rest of my time back, very uncharacteristic for this body, but nonetheless.

Senator COLLINS. I will take it.

Vice Chairman RUBIO. Senator Collins.

Senator COLLINS. Thank you, Mr. Chairman.

Mr. Clegg, we have received briefings from the intelligence community that indicate that China is not focused on the Presidential election race but rather on down ballot races at the State level, county level, local level. That concerns me because officials at those levels are far less likely to receive the kinds of briefings that we receive or to get information from Homeland Security or the FBI on how to be on alert.

In addition, China is attempting to build relationships with State and local officials. We see the sister city programs. We see the Confucius Institutes at educational institutions.

So how are your platforms attempting to help safeguard the down ballot races? The presidential race, I think, everybody is aware of the risk there, but the down ballot is what really concerns me.

Mr. CLEGG. Senator, I think you are right to be concerned, and that is why our vigilance needs to be constant. It can't just sort of peak at the time of the Presidential elections. It is something in which we need to deploy policies and enforcement around the world and around the clock.

And also, you are right, Senator, to point out that what we have seen—what we have at least seen, I know my colleagues have witnessed what we call the coordinated inauthentic behavior networks conducted by China. Some are quite specifically targeted at particular communities.

So, for instance, quite recently we disabled dozens of Facebook and Instagram accounts which were targeting the Sikh community in the United States. That is one of the reasons why the central signals that we look for aren't related to the content or even the person, but the behavioral patterns that we see. And the telltale patterns are most especially the use of a network of fake accounts.

And that of course manifests itself in lots of different ways and is targeted at different communities, but the underlying analysis that our team has conducted is about the behavior rather than the individual bit of content. Because as Vice Chairman Rubio said, sometimes the content can be actually consistent with things that are circulated by kind of ordinary folk in kind of, you know, the normal, day-to-day business.

Senator COLLINS. Thank you.

Mr. Smith, you talked about the need for the American people to be prepared and to be on the alert. Why isn't part of the answer so that we are not getting into suppressing dissenting views or criticism of public officials, for example, why isn't the answer to watermark posts to indicate not whether they are AI generated, but rather where they originate?

Like, why couldn't you do an "R" if it came from Russia? Then the person who is looking at the post can make his or her own determination, but they would be on alert that this isn't Joe, down the street, who has posted this. This is someone who is in Russia.

Mr. SMITH. I do think that is a really interesting idea and it is one that across the industry people have been giving a lot of thought to.

I would say a couple things. First, I think actually it starts with also picking up on the idea you just described and putting Americans and American organizations in a position to put what is called metadata, in effect, to put the credentials in place so it is clear, where their content has come from.

We worked, for example, with the Republican National Convention, and they used that on more than 4,000 images that were released in Milwaukee so that it would protect their content from being distorted.

I do think one can then go farther, and it is an important question, as you raised, if we find something that is coming from somewhere else, how and when should we identify it.

I frankly think the most important thing is that we address content where that kind of protection has been removed. And that has been the subject of legislation being proposed, including from Members of this Committee, to protect against tampering. And then we can think about other forms of identification for the public.

Senator COLLINS. Thank you.

Vice Chairman RUBIO. Senator Kelly.

Senator KELLY. Thank you, Mr. Chairman.

Thank you all for being here for this very important hearing.

I just got back from visiting our allies in the Baltics who all border Russia, also to Finland. And they have been targeted by disinformation attacks at a pretty high level and come pretty quickly. And they have efforts in place to try to equip their citizens and their institutions to counter disinformation campaigns, they feel, somewhat successfully, though it is a big problem for them. But I do think we can learn something from our partners in the Baltics.

Malicious actors, as you know, use social media and internet platforms as a key vector for these campaigns that they have against us and are increasingly employing tools. We've talked about this bots, generative AI. So, It is my hope that we can also

count on the partnership of the American tech industry to aggressively counter these threats.

I want to turn to a specific problem that is of great concern to me, and as my constituents learn about this, I am sure it will be to them as well.

Behind me you can see a screen capture of Russian made web pages designed to look like major American outlets Fox News and the Washington Post but showing fabricated headlines. I went through these the other day.

I think the Chairman showed something very similar, so apologies for being a little bit redundant here.

But these pages were created by Russians or Russian cyber operatives to distribute Russian messages by co-opting the brand of a real news website that Americans trust, both Fox News and the Washington Post, but there are others, as well.

These are really well done. I mean, it would be hard, unless you were looking specifically at the URL and noticed that something was not exactly right, where there was no dot-com, there was a dot-pm or dot something else at the end you wouldn't otherwise know and you would think this was a legitimate news source.

They've also spoofed the official NATO website as well. And they use these sites to push messages that cast doubt on Russian atrocities that we know are real. They lie about NATO suppressing peaceful protests. They stoke controversies or even invent them where they don't exist.

So, an additional concern is that they specifically targeted swing State voters—so, my constituents in Arizona and others—and they seek to influence the outcome of these elections. This is absolutely beyond the pale. We have got to do something about it.

So, I am curious from each of you, and I have about two minutes here. Just what are you doing about it? And specifically with these websites, if we were to go and look for them now, have they been taken down—the Fox News website, the Washington Post? Would we still—is there a way to—

Let's say we start with you, Mr. Walker. If we search on Google and tried to find this through a Google search engine or search for the Washington Post, could we navigate from your website to these fake websites?

Mr. WALKER. We are obviously concerned about the larger problem. I haven't searched for these specific sites, but I can tell you, we have launched tools called "about this image" and "about this result" which tells you the first time we saw an image appear on the internet. So, in many cases disinformation may not be AI-generated, it may be a repurposed photo.

Most of the disinformation we see coming out of Gaza is not AI-generated, it is pictures from a different war.

So, providing that kind of context is valuable. Then just quickly, to say that if content is AI generated, increasingly the ability to watermark it or understand its provenance through the C2PA cross industry group that I mentioned before will help all of us do a better job of identifying and removing this type of content.

Senator KELLY. Once you find the content and you know it is fake, at that point, can you take action to make sure that your customers cannot navigate to that content?

Mr. WALKER. The search context is somewhat different than other contexts where we are hosting information. So, let's say YouTube, which would be our hosted content example here. If something is demonstrably false and harmful, we will remove it, in addition to all of our other policies. And that has been consistent for many years.

We also have a general manipulated media policy, whether it is AI manipulation, or you may remember the cheap fakes that went around some time ago, which were slowing down videos to make a politician look as though they were intoxicated. We will remove that kind of content, yes.

Senator KELLY. You said if it is false or harmful. How about if it is just them co-opting somebody else's website like Fox News or Washington Post?

Mr. WALKER. I go back to Brad's earlier comments with regard to the notion of trademark infringement, copyright infringement. As we get complaints about that, we will remove that content, yes.

Senator KELLY. All right. Thank you.

Chairman WARNER. I would quickly note, I think most of your companies do a pretty good job on trademark protection. I just feel like Fox News and Washington Post should have gotten that same level of protection, and, frankly, they should be weighing in as well.

Senator Cotton.

Senator COTTON. Thank you. Gentlemen, thanks for appearing. I mean, I want to bring a little perspective to the topic today.

I think this committee's own report of more than 1,000 pages, said that Twitter users alone produce more election related content in about three hours in 2016 then all of the Russian agents working together.

I have no doubt that Russia and China and Iran and North Korea are all doing these things, up to no good. And if you don't know what they are doing, it is probably no good. And there is a lot of things they could do that are very bad to influence American politics.

You know, Russian intelligence spent millions of dollars in the early 1980s to promote the nuclear freeze movement which Joe Biden bought hook, line and sinker. And Russian intelligence under Vladimir Putin has spent millions of dollars to oppose fracking which Kamala Harris has bought hook, line and sinker, trying to ban fracking.

And there is plenty of things they could do in our election infrastructure as well. They could hack into campaigns, leak their strategy, or steal their voter contact information. Even worse, they could hack into county clerk's offices or Secretary of State's offices and delete voter registration files or try to manipulate votes.

They don't even have to get into the election machinery. They can turn off the electricity in a major American city on election day and wreak havoc there.

So, there is a lot of threats that our adversaries could pose to us in our elections. I just don't think that memes and YouTube videos are among the top, especially when we have an example of election interference here in America that was so egregious.

Some of your companies' efforts, in collusion with Joe Biden's campaign, led by the current Secretary of State to suppress the factual reporting about Hunter Biden's laptop.

Mr. Clegg, you acknowledged earlier that Facebook demoted that story after it was published by the New York Post, is that right?

Mr. CLEGG. Correct, but I should clarify we don't do that anymore.

Senator COTTON. I know Mr. Zuckerberg has said that you demoted it, and he expressed regret. And I assume you share that regret with your boss?

Mr. CLEGG. Yes.

Senator COTTON. And you share what he said that you are not going to do it anymore, right?

Mr. CLEGG. Correct. So that demotion does not take place today.

Senator COTTON. Mr. Walker, what about Google?

Did Google suppress results about the Hunter Biden laptop?

Mr. WALKER. We did not, sir. We had an independent investigation, and it did not meet our standards for taking any action, so it remained up on our services.

Senator COTTON. OK. And Twitter under the old regime there, was, I think someone said, was even more egregious than Facebook or other platforms. And again, this is domestic information operations, if you would like to say—far more influential on elections than some memes or YouTube videos or articles that Russian intelligence agents or Chinese intelligence agents posted, which no doubt they do.

And just look today. The New York Times the other day had a fit that social media was awash—"awash" it said—in AI generated memes of Donald Trump saving ducks and geese.

Are AI generated memes of Donald Trump saving ducks and geese really all that dangerous to our election?

Mr. Smith, you laughed, for the record.

Do you want to answer my question? Are you worried about——

Mr. SMITH. I think it's to your point.

Senator COTTON [continuing]. Ducks and geese memes of Donald Trump saving them from predators?

Mr. SMITH. When I create a list of the greatest worries for this election, they do not involve ducks or geese.

Senator COTTON. I wouldn't think so, either. It didn't seem like that to me, either.

Mr. Walker, Google famously did not auto fill results of people searching for Donald Trump's—the assassination attempt of Donald Trump a few weeks ago. What happened there? Why was that the result of your company's——

Mr. WALKER. We have had a longstanding policy, Senator, of not associating terms of violence associated with political officials unless they have become an historical event. So, the assassination of Abraham Lincoln would have been allowed.

Up until the weeks prior to the assassination attempt, it would have been deeply problematic, I think, to auto-complete "assassination" after a search for Donald Trump.

Those terms are periodically updated. The assassination attempt occurred in between one of those periodic updates. It has subsequently been updated and now auto-completes appropriately.

Senator COTTON. Let me ask both of your companies. This primarily Mr. Walker for Google and Mr. Clegg for Facebook.

Gavin Newsom just signed a law—three laws, actually, in California, into effect that will criminalize the use of so called "deep fakes" before an election.

How do you plan to comply with that law?

Are you going to go arrest people who are making AI-generated memes of Donald Trump running away with ducks and geese?

Mr. WALKER. Senator, it is early for us to understand. We are just receiving the laws which were signed very recently, and we are looking at how we might best comply with a number of laws. There were quite a few.

Senator COTTON. Mr. Clegg, a lot of ducks and geese memes on your website.

Mr. Smith thinks you are funny. He is laughing again.

It's fine. People laugh at them. Satire and political humor are as old as our country. It's fine.

I am glad that you are not going to do again what you did in 2020, but I don't envy either of your companies dealing with what Gavin Newsom has done in California or what many in this Congress propose to do, criminalizing and censoring core political speech.

Mr. Clegg, do you have any idea of how you are going to comply with California's law?

Mr. CLEGG. Well, it's only just been signed, so, again, we would need probably to look at it more closely.

But I think, Senator, your central point that there is a lot of playful and innocent and innocuous use of AI and then there is duplicitous and egregious and dangerous use of AI. That is exactly why I think Governor Newsom——

Senator COTTON. And I have to ask. My time has expired, but I have to ask: Who is going to draw that line?

Who is going to decide what is playful and innocuous and harmless and what is misinformation and disinformation?

And I got to say some of the people you go to, PolitiFact and Southern Poverty Law Center don't strike me as quite neutral sources and I don't think you are going to find neutral sources in the government of California or in this administration, either.

Chairman WARNER. And I would like, just as we look at the California law, I would like your analysis as well of the deep fakes used in political advertising that was passed and signed into law in Alabama, Texas, and Florida as well.

Senator King.

Senator KING. Thank you, Mr. Chairman.

I think the bright line here should be foreign—the word "foreign," as has been pointed out.

As the vice chairman pointed out in his opening remarks, it becomes problematic when you are talking about domestic content and then it is being amplified by foreign content. That should be the line.

I mean, I don't want you all or the government certainly to be the arbiters of truth, because one man's truth is another man's propaganda. I mean, I think we should have that kind of flexibility.

It seems to me what is happening here is that foreign governments are engaged in a kind of geopolitical judo, where they are using our own strength against us. Our strength is our democracy and our regular elections plus freedom of expression and that is what they are taking advantage of in order to try to manipulate our fundamental way of making decisions, which is through elections. But the issue should always be is there a foreign nexus, is there a foreign influence in this matter?

I guess the question is, in this day and age, can you determine that given the fact that we have got very sophisticated adversaries in St. Petersburg or Moscow or wherever, or in Tehran, who may be coming in via a server in Georgia.

Can you technically tell when something is of foreign origin?

Mr. Walker or Mr. Smith.

Mr. SMITH. I would say the answer is not always, but often, yes. And I do think that there are some threats that we take seriously, and we should start with the word "foreign."

But if you want to see the risks that we should be thinking about, I would go back to Slovakia. Their Parliamentary election was last year, September 30. Two days before, on September 28, a Russian group released a deep fake audio. It purported to be an audio of a conversation between a mainstream journalist and the leader of the pro-European Union political party, one of the two largest political parties in that race.

That reflected what we see in Russia, No. 1, a good content creation strategy.

The second thing they did on that same day, is they released it on Telegram which tends to be the Russians' favored distribution channel to get things going. They did it from what was the private account of the spouse of a major official in Slovakia.

The third thing they did is they pursued a content amplification strategy where then one of the most senior officials in the Russian Government, as they tend to do, came out the very same day and accused the United States of doing what that audio recording purported to capture in Slovakia; namely, a plot to buy votes and steal the election.

Senator KING. In other words, it was a very sophisticated operation.

Mr. SMITH. It is, and this is what we need to remember. You can't have a great play without a great playwright. The Russian government is very capable, very sophisticated, not just in technology, but in social science.

Senator KING. And very determined. Very determined, are they not?

Mr. SMITH. Yes, absolutely. And that is, that is what we—There are many things.

It is right, I think, to focus on the things that should unite us and say let's not worry about what we are seeing over in one direction, but let's not close our eyes in what we could see in the other as well.

Senator KING. The question is, No. 1, it's happening. You have all testified to that. It is happening and it is not a minor project on the path of Iran, Russia, and to some extent China.

So, the question is then, what do we do? I know Senator Collins asked about watermarking, some kind of way to determine the source of the information attribution.

But I had a formative experience about eight or nine years ago in this building before any election, before 2016, meeting with a group of people, politicians, political officials in Estonia who are under bombardment all the time from Russian propaganda and Russian disinformation. I asked: How do you deal with it? You can't cut off the internet or cut off your TV stations. Their interesting answer was: We deal with it by educating the public that it is happening. And they say, "Oh, hell, it is just the Russians again."

And that is why I think what we are doing today is so important and your testimony is so important, so the American people can be alerted to the fact that they may be being misled and they should check. Is that a reasonable approach?

Mr. SMITH. Absolutely, and what I hope we can take away from this, because first of all, there is something very important what Senator Cotton said, not everything is a threat; and, as Senator Rubio said, we should honor our citizens to say what is on their minds. But Senator Kelly captured something that is critical, and you are pointing to the same thing. When you go to Estonia, when you go to Finland, when you go to Sweden, when you meet people who have lived their entire lives in the shadow of Russia, they are on the alert. They know, as we have discovered, that not everything on the internet is true. They just remember that when they read something that is new.

Senator KING. My wife and I have a sign in our kitchen that says: "The difficulty with quotes on the internet is determining their authenticity—Abraham Lincoln."

[Laughter.]

Senator KING. Mr. Clegg, you were going to respond? I'm sorry, Mr. Walker.

Mr. WALKER. Yes, Senator, thank you.

Just very briefly. In addition to those very good points which I agree, I do think we are increasingly able to use the AI to detect some of these patterns.

As we discussed previously, YouTube has gone from having one view in 100, following our policies to one view in a 1,000. That is in large part because we are using AI to detect some of these patterns of misinformation and disinformation that are out there and take action against them.

Senator KING. You can either take action or you can alert your customers that this has been manipulated in some way.

Mr. WALKER. Agreed and also provide high quality, authoritative information. The old line, "the best remedy for bad information is good information."

So, the more we can promote accurate information about when the polls are going to be open, people's eligibility to vote, whatever else it might be, that is an important part of the democratic process.

Senator KING. Thank you, Mr. Chairman.

Chairman WARNER. And I just remind, I agree with the comment around memes, but I recall that this committee exposed in 2016

the effort by the Russians to incite violence between a pro Muslim group in Texas and a pro kind of Texas separatist group that but for law enforcement would have resulted in American harm.

And echoing in how we know, I don't know when these slides are up how a normal American consumer, even a relatively sophisticated one, would have the expertise to read the URL that closely when everything else looks so closely like Fox or the Washington Post.

Senator Cornyn.

Senator CORNYN. I would like to ask each of you to respond to this question: Do you believe that ByteDance should be required to divest TikTok in order for TikTok to operate in the United States?

Mr. Walker.

Mr. WALKER. Senator, I would defer to Congress. I know you have legislated on this very question.

Senator CORNYN. Do you think social media companies owned by foreign governments that are adversaries of the United States that are known to use information warfare against the United States, do you believe they should be able to operate freely in the United States?

Mr. WALKER. As a technology company, our area of expertise is making sure that they are not distributing malware. We have found situations where such companies were distributing malware, at which point we removed them from our services.

But on the broader question of accessibility, I think that is a question for Congress.

Senator CORNYN. I will put you down as undecided.

Mr. Smith.

Mr. SMITH. You can put me down as I think you all have already decided. The Congress has passed a law. The President has signed it. The courts will adjudicate it, but assuming it is upheld, then clearly it needs to be followed. And I am not going to try to substitute my judgment for the judgment you all have already brought to bear.

Senator CORNYN. Mr. Clegg?

Mr. CLEGG. In addition to that, I will just point out that there isn't a level playing field globally. Our services, for instance, are not available to people in China. So, Chinese social media apps are available here, but American social media apps are not available in China. That has been the state of affairs for some time.

Senator CORNYN. What I am looking for is the guiding principles here, and Mr. Clegg, it sounds like reciprocity should be perhaps one of those principles.

Mr. CLEGG. I think the First Amendment principle of voice for the maximum amount of people for the maximum amount of time wherever they reside around the world is a good principle.

Senator CORNYN. Well, the problem I think we are having, trying to figure out what the appropriate framework is to think about what you all do day in and day out, because it has presented a bunch of novel and difficult questions. But before social media companies existed, it seems to me we had doctrines, laws that governed the way that we dealt with the subject matter that we are talking about here today.

Of course, what is so different today is you are private entities so presumably the Constitution, the First Amendment, can't be directly applied. I know the Supreme Court is wrestling with how to figure out what the right way to view social media companies is.

You have your terms of use which strike me as a pretty powerful tool to be able to regulate what is on your site, but there are also legitimate concerns about censorship of views. And of course, Mr. Clegg, you talked a little bit about Mr. Zuckerberg's letter and the fact that he regrets that Meta was being influenced and cooperating with the Federal Government.

And then we have regulations that usually help us in this area, or as a last resort, litigation.

So, I am wondering, is there anything about the way that we operated and the legal framework we operated under before your companies existed that should inform the way that we view your operations today?

It strikes me as we are dealing with adversaries often that view information warfare as a legitimate tool. Obviously, the Russians and their active measures campaign existed long before your companies existed.

But we are an open society, and we believe in freedom of exchange and free speech; but is there anything about the way that we regulated or the way the framework under which we understood that newspapers, radio, movies, other means of communication were handled pre-social media companies that should guide us here or are we just trying to make this up from scratch?

Mr. SMITH. The one thing I would say without getting into I think your very important question about sort of the history of regulation of communications in the country and everyone could have a vibrant debate about section 230 and the like is this: It is easy to spend all our time on issues where we disagree.

I think the most important thing is we identify where we actually do agree across the political aisle and across the industry, because if we can act based on common consensus to address the foreign adversaries, emphasizing again that word "foreign" and nation states, we can do the most important thing I think we need to do this year. I think that can build a foundation for the future and then we will deal with the rest and your very important question among that.

Senator CORNYN. My time is up.

Chairman WARNER. Again, I want to commend Senator Cornyn for raising this. We did actually do that on the question about CCP control of a platform that candidly is even more popular at this point then your platforms, and 80 percent of the Congress in both political parties said that is not in our national security interest, and I appreciate you raising it.

Senator Bennet.

Senator BENNET. Thank you, Mr. Chairman.

I appreciate you having this hearing and appreciate you coming to testify. I'm very grateful for that.

I think what we are struggling with a little bit, in terms of answering the question Senator Cornyn just posed is the sheer scale of the enterprises you represent. That presents something new to us.

And as I sit hear listening to this conversation, I am thinking about the people who are going to be sitting in your chairs 30 years from now and the people who are going to be sitting in our chairs 30 years from now, and what are the incentives that are leading us to have the conversation we are having right now and answers we are having in this minute for all the right reasons are the ones we would have wished for 30 years into the future.

I really wish on behalf of the American people that the American people would have had a negotiation with Mark Zuckerberg, just to pick him as an example, around our privacy and around our data and around our economics. I don't believe we have had that negotiation. I don't think we have with any of these social media platforms—different, Mr. Smith, than your company—with any of these platforms about our privacy, our data, our economics, the way we want our children's bedrooms invaded or not invaded. And for better or for worse, they are looking to us to try to begin to have that conversation.

So, first, we haven't had it, and here we sit having to deal with the very, very severe consequences across our society. I say that partly as a capitalist but also as a former school superintendent who has seen the effects of mental health on our kids, and as Members of the Intelligence Committee who are trying to protect the country from an invasion of our democracy across your social media platforms and tech platforms.

When I read your Capex numbers it staggers my mind. I can't even get my head around the idea that you are going to spend $170 billion over 18 months on AI investments. I mean, that annual expenditure for your three companies is more than we had for roads and bridges in the first infrastructure bill we passed since Eisenhower was President. And for all the telecom or broadband infrastructure across the entire United States of America. Those things together are dwarfed by your annual Capex expenditure on AI. And I feel like we are being asked to sort of hope for the best.

I think it is an amazing testament to American capitalism that you have those resources to invest in the future, but you better be making the right decisions. And part of that I think is a question of whether the commitment—you really made the commitment on the front end to safeguard America's democracy, to make sure our elections are protected, to not say that it is up to our citizens to try to figure it out in the hailstorm of propaganda that has almost been perfected by our adversaries and every day is being used by them to divide one American from the next, from the next, from the next, because they see that division as a potential benefit to them and a huge detriment to us.

How much money are you investing to make sure that you are protecting our elections?

Is that your responsibility, or is this just, you know, an approach that says let a thousand flowers bloom?

I am a strong believer in the First Amendment, but I don't think there is anything about the First Amendment that obviates your need to be able to say to the American people: We believe we have a responsibility to you because we are creatures—among other things because we are creatures of this unique society and this unique democracy and we have an obligation here.

So, I don't know if anybody would like to respond.

Mr. Clegg? Please.

Mr. CLEGG. Yes. So, to answer your specific question, we have around 40,000 people working on security integrity of our services. In fact, that number is slightly up from what it was.

Senator BENNET. I am deeply, deeply skeptical of the numbers, because the numbers don't tell you what the investment is. We know they go up and we know they go down. And as Mr. Walker said earlier, maybe the AI tools themselves are better—and I don't doubt that. That may be true. I am more interested in what the total capital expenditures are.

Mr. CLEGG. Capital expenditures, about $20 billion over the last several years. Around $5 billion over the last year.

To your wider point, Senator, I strongly agree with you that the scale that one is dealing with, whether it is from the tech company's point of view, from legislatures and governments around the world, it is clearly unprecedented because the network effects are created by the internet.

On our services alone, you have 100 billion messages around the world on WhatsApp every day. You got now about three and a half billion reshares of short form videos, reels every single day.

And much as cooperation between companies at this table and between companies that are not represented at this table is crucial to deal with the scale of all of that, I would also suggest that cooperation between different jurisdictions in the democratic world globally is important as well, particularly between the United States, Europe, India, and so on, because I think one of the greatest risks is a fragmentation of different regulatory approaches around the world for technology which by definition are borderless.

Mr. SMITH. I would just note very quickly——

Mr. WALKER. Go ahead.

Chairman WARNER. Go ahead. We have got a couple more minutes.

Mr. SMITH. First of all, I believe that the American tech sector is the engine of growth and frankly is the envy of the world, and we should at least remember that.

No. 2, we do have a very high responsibility to protect elections, to think about the impact on others, on our societal responsibility in so many areas.

No. 3, if there is a foundational principle for this country, I believe it is straightforward. No one should be above the law—no individual, no company, no leader, no government.

But then, No. 4, let's recognize the obvious. We need laws.

I would just say, I put it slightly differently. We haven't had a shortage of debate in this country about an issue like privacy. We have had a shortage of decision making. So instead of always worrying about where we can't reach agreement, why don't we get something done by taking more action, by calling on us to be maybe more supportive, as we could and should on certain days and helping you all so this Congress can pass the laws we need. I think that is the recipe that we need for the future.

Chairman WARNER. I can't. I will bite my tongue.

Senator Lankford.

Senator LANKFORD. Mr. Chairman, thank you. Thank you too for showing up.

We invited several more tech companies and they chose to just decline, not to be here in the national conversation. So, I do appreciate you giving the chance for you to be able to be here.

Let me just outline some of the challenges we face on this that do become obvious to all of us when we get a chance to be able to look at it.

This is not picking on Meta, but it is going to be a side-by-side with TikTok who is not here. But this is just an example side by side of content delivery from a company. When there was a comparison that was done of content delivery to individuals that were 35 and younger from Instagram to Tiktok.

On Uyghur content, it was 11 to 1 Instagram. So TikTok hardly delivered it; 11 to 1, that if someone was talking about Uighurs, Instagram was talking about it, TikTok wouldn't.

In a conversation about Tibet, 41 to 1 Instagram to TikTok. TikTok just screened it out.

On the Tiananmen Square, 80 to 1 content on Tiananmen Square. This is among Americans, by the way.

Hong Kong protests, 180 to 1. That seemed to be a conversation that was discussed on Instagram that just didn't show up on TikTok for whatever reason.

Ukraine, 12 to 1.

And this one was interesting to me. There is 50 times more pro-Palestinian content on TikTok than pro-Israel content.

Now, I say that to you to say, there is a sense of an outside foreign influence, in this case owned by a foreign entity trying to be able to deliver content to the United States to affect the national conversation. That is the challenge that we have because there is not a challenge on what Americans want to be able to talk about. The challenge is a foreign entity reaching into the United States and saying: Hey, I want to try and influence you by delivering content to your box that may try to sway opinions on this.

So, two things I would say on this: First of those is, the concern is for not just a TikTok or to a foreign entity, a Russia, an Iran trying to be able to put bad content in, misinformation, disinformation, but it is also the feeding of the quantity of the algorithm. This is an area where Americans have got to be able to rebuild trust. I would say there is a lot of suspicion, because the delivery of what content is actually coming to your feed is an area of skepticism, whether it is in a Google search or whether that is in whatever they are getting from a social media network on it.

How do we actually set in front of the American people enough transparency that there is a trust that is neutral in what is delivered, yet your task is to keep people looking at the screen all day, so you are trying to feed them information they want to see more of.

How do we hit that rhythm on it, because that will be important for Americans, period, in their own dialogue?

Anybody want to try that one?

Mr. CLEGG. I will try, Senator.

I think, Senator, you pinpoint a very important issue, which is algorithms in a sense deal with a practical problem which is there

is an infinite amount of content that you can show people, but of course, people have a limited amount of time they are scrolling on their feed, so you have to somehow rank and funnel it.

And I believe the way to square that circle that, Senator, you quite rightly allude to is giving people confidence that these algorithms are working for them and not against their interests. It is firstly to give people real control.

So, for instance on our services you can just turn the algorithm off. You can just have it chronologically delivered instead. You can click on to the three dots and you see exactly why you are seeing a post. You can say you don't want to see certain ads. You can prioritize certain content or not. I think user controls are crucial.

And secondly, we need to be transparent. We need to be transparent about what are the signals that we use in the algorithms. We publish alongside our financial results every 12 weeks, for instance; full transparency report showing how we act on content that violates our policies. We have it audited by EY so we are not marking our own homework if I can put it like that.

So, I think user agency and sort of control and a maximum amount of transparency for the company are the key ingredients here.

Senator LANKFORD. Mr. Walker.

Mr. WALKER. Just to follow up on that. We seriously take the point about maintaining and building trust in the services. So, some of the ways we do that are anchoring our results in raters who are located throughout the United States, in rural and urban areas, 49 States at last count. That is the ground truth for many of our services. But beyond that, we do things like, for example, on YouTube, not just promoting the most popular videos, but the videos that users have found the most valuable.

We will survey our users the day after: Did you have a good experience in the service? Did you find this a valuable use of your time?

And then making sure that we are consistently and clearly enforcing—transparently enforcing our policies, which we also publish.

It is a responsibility we take very seriously.

Senator LANKFORD. It is. And it is something that is incredibly important. And it is also consistent with law on this as well.

Mr. Smith, in a comment that you made earlier that Iran is fighting against Trump, Russia is fighting against Harris, and we see the noise that is out there on this and the awareness of it. I do think it is important that we have this conversation to make Americans well aware that not everything they see is accurate and correct and there are things very deliberate. But one of the challenges that we have that we have got to figure out, both as a committee and both from you is attribution, that when something shows up, how to be able to designate that: "Here is where that originated" because by the time it is shared 50 times to 50 places, people don't know where it originated anymore.

So, one challenge is taking off content that is Russian content, Iranian content, that is deliberately a means to attack and to disturb Americans in whatever way that may be, but another one is to be able to make sure that when it gets out there that people are

well aware of it. We can't tell the story of this is disinformation, misinformation, unless we get fast attribution on that. And that has to be something we have to work out.

Chairman WARNER. And again, I have got critiques of all three of these companies. I will come back to some of those, but on this one, they have been more forward leaning, because if they don't share that by the time the IC or law enforcement picks it up, it may be too late.

Senator Ossoff.

Senator OSSOFF. Thank you, Mr. Chairman, and thank you all for joining us.

On the point about attribution and identification of foreign covert influence, Mr. Walker, give us a sense of your independent capacity absent case by case warning or notification from the U.S. Government of content on your platforms that is foreign covert influence?

Mr. WALKER. It is challenging as was talked about earlier. Russia has moved beyond paying for things in rubles and only working between 9 and 5 Moscow time. So, they are increasingly making it more difficult to identify things.

That said, we have 500 analysts and researchers working on the Mandiant team, Google threat intelligence, who are tracking between 270 and 300 different foreign cyberattack groups at any given point, tracking activities, metadata, et cetera, through our services and sharing it with the security teams that are represented here and elsewhere in the industry and also working with the FBI's foreign influence task force.

Senator OSSOFF. Let me put it this way: Do you think you are mostly across it and playing Whac-A–Mole or do you think you fundamentally lack the ability to know how much you don't know?

Mr. WALKER. I think the humble and probably accurate statement would be the latter, because the adversaries are always moving forward and it's a constant cat and mouse game.

Senator OSSOFF. You mentioned earlier using machine learning or algorithmic tools to try to identify it. Is that on the basis of network activity and posting tactics as opposed to content where there is a risk of collateral damage, you might suppress bona fide American speech because oftentimes what the foreign actors are amplifying resembles perhaps extreme or polarizing speech that is happening organically in the country?

Mr. WALKER. It is a deep and important question, and the answer differs to some degree across the different platforms; because a pure social network, as Mr. Clegg was referring to, will have more behavioral information. We may have more content-related or metadata style information.

We do try and share across the different platforms where we can, but inherently there is some sort of an assessment of the nature of the content. We talked a little bit about provenance in AI or metadata in AI. That's going to be a component of it. Network activity is a component of it, and then behavior signals will also be a component of it.

Senator OSSOFF. OK. In addition to attribution, let's talk about authentication.

Mr. Smith, you mentioned the Slovakian example, I believe. Let's game it out, all right? I think we need to be able to discuss this

out in the open how this might unfold in the United States and who bears responsibility for handling it. There might be some very compelling, seemingly authentic, deep fake audio clip which is, in fact, fake and defamatory implicating a candidate for office in the United States in the hours or days or weeks before an election. How confident are you that either you or another private sector actor or somebody else has the capacity to identify this fake, particularly where we can't rely on one campaign or the other necessarily to in good faith acknowledge that something which is useful to them because it deliberately defames and mischaracterizes the statements or conduct of their political opponent, isn't real?

Mr. SMITH. Well, I'd say first I think have a word of wisdom in saying we have to always act with a sense of humility, and hence I think we should require of us an extraordinarily high level of confidence approaching certainty before we take action.

Having said that, I do think especially given our ability to use AI to identify the creation of a fake and just the good old human judgment that comes from crowd sourcing, especially for video, we can identify a great deal. And I then think what it translates into is another part of your question. Great, what do we do about it?

There will be days, or it could be hours when the most important thing we will need to do is alert the public so that there is a well-informed conversation.

But I also think this points more broadly to what is a systemic strategy to try to address the problem that we are worried about here.

Senator OSSOFF. Well, because time is short, let me try this question, and ask it of each of you.

What will you do? What is your policy if, in that critical time period before an election, there is deep fake content attacking a candidate for office which can be demonstrated to be inauthentic but cannot be decisively attributed to a foreign actor, how would you handle it?

Mr. CLEGG. We would label it. We would label it so that users would see that the veracity or truth of it is under real question. So, we would label it.

Senator OSSOFF. What about how it is handled in the algorithm in its amplification or suppression?

Mr. CLEGG. We would also make available to us the ability to demote the circulation of it.

Senator OSSOFF. Mr. Smith.

Mr. SMITH. We don't have the same issue in terms of a consumer platform, but I think that the notification to the public, the labeling, I do think that is the essence of what we all need to be prepared to do very quickly.

Mr. WALKER. And I would add to that that we would notify the foreign influence task force so that there was government awareness to the situation.

Senator OSSOFF. Thank you.

Chairman WARNER. Thank you, gentlemen. I've got a few more comments.

I guess, where I would start is, I remember all three of you in Munich, when companies like TikTok and X signed on to that agreement. Again, amazed and disappointed with particularly X's

failure to participate and failure in any way to adhere to that document.

But if what you have just said is—I want to make sure we didn't get off just on Fox and Washington Post, but moving this publication forward, another example.

If we got a watermarking system, the fact that this is content that didn't originate with you but was placed on your platform, these are not watermarked. I'm not sure there is a way that anyone that is a normal consumer—because you've got a byline, you've got authentic ads on the other side—are going to find that. And again since, they ended up on yours, I'm gonna—You know, you want to protect your brands. These are brand clients. Why didn't we catch this?

Mr. CLEGG. So, I think the key challenge here is to disrupt and remove the underlying networks of fake accounts that generate this content.

Chairman WARNER. We appreciate what you did yesterday.

Mr. CLEGG. That is the only foolproof way that we can deal with this, because otherwise, as you quite rightly say, Senator, we are just playing Whac-A-Mole on individual pieces of content.

The companies on this table and other companies besides I think have made real material progress since we assembled together in Munich, for instance, to agree on interoperable standards of not only visible watermarking but also so-called metadata and invisible watermarking. So, we, as social media platforms, as we ingest content from elsewhere, we can then detect those invisible signals so we can then alert that to our users. But of course, bad actors—in this case, foreign actors, Russian networks, are not going to introduce those.

Chairman WARNER. They are not going to put the watermark on.

Mr. CLEGG. Correct, which is why for us the overriding objective is always to disrupt the wider network.

Chairman WARNER. But again, but at the end of the day, what I don't understand and whether this was on Facebook or appeared on Google or appeared on YouTube or appeared on X, the URL is the distinguishing characteristic. The consumer is not going to get that. Should that be simply the government's responsibility to spot that? Don't we need you leaning in on that issue?

Mr. CLEGG. Yes, of course, absolutely.

Chairman WARNER. So, one of the things, because we are—we keep coming back with we are 48 days away.

You know, I'm going to ask you, Mr. Smith, as well, but let me start with Mr. Walker and Mr. Clegg.

I need to know, starting with this kind of and we will share all the ones that came out of the Justice Department report—how many Russian manipulated images that are completely false, that sow dissension, that undermine campaigns—how many Americans have seen those? Because clearly your whole metrics of models is based on how many eyeballs you get. We have got to have that information.

I also believe that there are a series of ads, and we will share again with the companies in more detail that are getting through the protections at this point. We need to know how many of those ads because if we can—my concern is when people undermine and

say: This is only memes or this is not a serious issue. Again, Americans have the right to say anything no matter how "out there" it is. But back to what Senator Cornyn said, you know, the notion even around reciprocity, the idea that Russia or China would allow this kind of manipulation on their social media is beyond the pale. Of course, they wouldn't.

So, we need that because the one thing we do know, most all of us will agree, in the next 49, 48 days, it is only going to get worse. And having that data now, not to embarrass what happened at least on Facebook, to say: Hey, you know, x-millions of Americans saw this kind of fake content. Just be aware, because chances are no matter what we do, we are not going to stop all of this from coming down, but that measure would help identify.

I also think on the ads. I mean, I know it has gotten better. Mr. Walker, you mentioned the fact that you don't take payment in rubles anymore from 9 to 5 Moscow time, but there is still a ton of this getting through, and we need better data at this point. So, I will expect that very shortly.

If you still have colleagues or friends at X, I sure as heck invite them to be a part of the solution, as opposed to simply trying to be part of exacerbating, sometimes, the problem.

And we have those who don't play. I mean, X, TikTok, this whole set of Discords, the Telegrams. There are others. They almost in some cases pride themselves of giving the proverbial middle finger to governments all around the world, which I think raises huge issues as well. So, I'd like to have that information—I think Senator Ossoff has one more—and as soon as possible—and I will have one last closing comment.

Senator Ossoff.

Senator OSSOFF. I will be brief, Mr. Chairman.

Just to note—and the committee has made public some of the underlying information that was contained in the charging documents related to the specific recent Russian effort for which there were 32 domain seizures Doppelganger, which planning documents specifically identified "swing states whose voting results impact the outcomes of the elections more than other states," and named in particular Georgia as a destination for this covert Russian influence.

We talked about attribution. We talked about authentication. I think we have also been discussing the importance of having a society that is resilient, that takes a skeptical and critical approach to information.

One of the challenges we have is for some avid consumers of political content anything which seems to affirm one's partisan perspective is deemed credible without that kind of critical scrutiny.

For my constituents in Georgia who have recently been targeted by this foreign covert influence campaign, but for the whole nation, how do you think about your role, and invite you to comment on the role of public leaders, elected leaders. How do we build that kind of resilience across society such that we don't just accept anything that seems to affirm our world view or denounce our enemies, but we recognize that, foreign and domestic, there are a lot of folks telling lies and a lot of folks taking an interest in manipulating us.

Mr. Clegg, want to take a shot at that?

Mr. CLEGG. Well, the first thing I think as has been mentioned by a number of Senators already, we can learn a lot from countries like the Baltics. Moldova, I think is a country right now in the frontline facing a lot of Russian interference. Taiwan—the Taiwanese election recently—All these countries in different situations are dealing with major adversaries who are trying to interfere in their elections. And public skepticism, voter skepticism, is probably the greatest antidote to a lot of this. And I do think political leadership can play a role in fostering that.

The other thing which I think is crucial, and that is on us, is every time we find networks like that, we need to share that as widely as possible with researchers, with our colleagues in the tech industry, with governments. For instance, we now publish every 12 weeks an adversarial threat report. We have done so in the last few years.

And Doppelganger. Senator, you mentioned Doppelganger. It was our threat intelligence team that identified Doppelganger first 2 years ago. We blocked around 5,000 accounts and pages in 3 months, in a 3-month period this year.

We have placed a lot of the signals we were able to detect on GitHub so that everybody can look at that and everyone could learn from that experience, and we've got people that come in, scrutinize it, tell us what we got right and what we got wrong. I think that interchange of research and data is crucial to develop public and societal resilience in the long run.

Senator OSSOFF. And education plays a role as well. Thanks.

Let me ask this final question.

Oh, Mr. Walker, go ahead.

Mr. WALKER. Just very briefly, I want to give one example because it is obviously a deep democratic question at a time when trust in institutions of all kinds is going down. But in one specific case study that might be helpful, YouTube has launched a program called "Hit Pause." It is a series of short videos designed to remind people not to believe everything they see; that if facts are one sided, if it is an overly emotional kind of pitch, et cetera, there are a series of ways of framing things that are often used by people pushing false information. We found actually in independent research that the lasting effect of some of those short exposures can actually last for months. People become more resistant to fake news.

Mr. SMITH. I would just underscore that. That is an excellent initiative. We have been doing similar work at Microsoft. We really sharpened our ability in the European Union Parliamentary elections. We ran a paid media advertising campaign around checking and rechecking before people make up their mind and vote. It reached 350 million people outside the United States.

That is why we are bringing that to the United States. Certainly, the swing States are critical. And it is not just advertising. It is getting out on drive time radio, local press, to help bring this message so that the American public has the information it needs.

Mr. OSSOFF. Thank you. Final question.

Mr. Clegg, putting aside law and regulation, when you think about, for example, your employer's social obligations and how you

meet those social obligations in the decisions that you make about how content is labeled or how your algorithms treat content, in a society where sharp-elbowed political debate is part of the process and free speech is cherished as a value in addition to being a constitutional right, what is the distinction between the role that your teams are fulfilling in making those calls and the traditional editorial judgment that a traditional news organization would make?

Mr. CLEGG. The fundamental difference is that we don't generate the content. So, it is user-generated content that circulates on apps and services. It is almost an inversion of the top-down way in which information is selected and handpicked by editors sitting in editorial suites for newspapers.

Senator OSSOFF. But you decide what is on the page.

Mr. CLEGG. We decide as I said earlier or decide—we have systems that seek to ensure that every person's feed is in a sense unique to them. It reflects their interests. It reflects what they enjoy spending time on. As it happens, the vast majority of people don't use Facebook and Instagram, for instance, to argue about politics. So, news and news links constitute around 3 percent of the total content on Facebook.

Most people use our services for much more playful, innocent—you know, connecting with family and friends, family holidays, family birthdays, bar mitzvahs, barbecues, you name it. And that is reflected in the overwhelming majority of content of our services.

Senator OSSOFF. Thank you.

Chairman WARNER. That sounds to me like a backhanded description around protection around section 230, which I fundamentally disagree with you on.

Again, I don't accept that characterization. That was the same characterization that initially people made about TikTok. "What could be so wrong about people sharing cat videos?" Although cat videos may take on a political stripe right now. Yet now, the number is 30, 40, 50 percent of 18- to 24-year-olds get all of their news or a vast majority of news from TikTok?

Again, I just do not accept the notion of just "we are just independent creators." There are algorithms that shift what you see, how much you see. Tech colleagues that we both know said there has never been a more creative, addictive, crack-like tool than TikTok in terms of tracking and keeping.

Again, in the effort that Senator Cornyn raised and the vast majority of us here, that the ultimate dials can be turned by CCP leadership in terms of what content you receive. I believe that is a huge national security concern.

I also just want to point out that the independent reviewers, I agree, that's good. And I do think there is a role for the academic reviewers.

I think we are less safe today because many of those independent academic reviewers have been litigated, bullied, or chased out of the marketplace. That concerns me.

I also hope and I would like to see not just kind of one-off answers, but I would like to see from all three of you something to the committee that Senator Rubio and I will review and share with our colleagues.

I think this point about the 48 hours, Brad, that you raised—I think we have put attention on that, but I think the post-election 48 hours is going to be equally important. And I would like to hear with specifics what kind of surge capacity each of your institutions are going to have as we get closer, because I am not going to litigate here whether you have cut back or not your content. And again, not content moderation on a political bent but content moderation in terms of whether your users actually adhere to your own terms of service.

I would simply state for the record, the overwhelming majority of outside observers, I think across the political stripes have said most of you have cut back. But you made your points. We don't have to relitigate.

So again, I want to know how many folks have seen and echoing especially in these targeted States, how many ads have gotten through, what we are going to move forward on.

I would also—I bit my tongue earlier before Senator Lankford got on—and I do think I have worked with each of you and each of your companies. There are places where we agree. There are places where we disagree. And I do believe, you know, Congress's batting record on social media platforms and on AI is virtually zero in terms of laws being passed, maybe with the exception of TikTok.

I would point out that when we had the largest AI dog and pony show in the emergence of AI when your CEO, colleagues, and everybody else was there, and Senator Schumer at that point asked: How many of you think we need regulation?

Everybody raised their hand.

And you know, I have got a half a dozen bipartisan AI laws or bills, some of them addressing things like how we avoid those entities that circumvent the watermarks that you and others may put in. But for the most part—and since I get the last word, I will leave this without contradiction—Everybody is for it in theory until you see words on the page. And there is always a reason why "we can't really do that" or "oh, gosh, if we do that, we are going so to slow down innovation" or "if we do that China is going to leap ahead."

And this is not the topic for today, but I think there is a whole lot of us—virtually every parent in America today would say that had there been a few guardrails on social media back in 2014, we might have a heck of a lot healthier kids in this country in terms of mental health issues. Not the subject for today, but something that the vast majority of Americans believe, including me.

So, we have made—you know, as I go through my statement, we have made some progress.

I do worry that this is not going to lead the news tonight. The fact that Russia and Iran—we don't have the kind of visuals yet. I hope we will get the visuals yet on what Iran has done—but that Russia, using brands that most Americans on either end of the political spectrum respect, FOX News and Washington Post, are seeing things that look like is that content that's not. It's coming from Moscow. And anyone who thinks that is appropriate, I just don't think reflects where we are in this democracy.

I will end with where I started. We have more than enough differences amongst Americans. We have a God given or Constitutionally given First Amendment right that allows us to say any-

thing, no matter how stupid, unless it is the equivalent of "Fire!" in a crowded theater.

But we should have those debates, but sure as heck should be concerned about foreign government services. This is not some one-off entity. These are foreign spy services who by definition want to undermine our country. When they are trying to sway an already very close election, we all should be concerned about that.

I appreciate you all being here. I wish more of your colleagues in the sector would be as engaged.

I think I have given you all some to-do work, and my hope is we will have some of that information because the clock is ticking, as you all have said.

I would hope we would get some preliminary information back even by middle of next week. Let's see if we can get this as we go into October.

With that, I did promise Senator Rubio I wouldn't go off on some other tangent, so I will respect that right now and say we are adjourned.

(Whereupon, at 4:35 p.m., the hearing was adjourned.)

71. The use of targeted advertising by Doppelganger is corroborated by records obtained from Meta pursuant to a warrant, which identified Meta pages and advertisements linked to the Doppelganger campaign. Notably, those records revealed Doppelganger's apparent use of artificial intelligence tools to generate content, including images and videos, for use in negative advertisements about U.S. politicians. Several of these Meta accounts were registered with account

SOURCE: Department of Justice Affidavit on Russia's Doppelganger Network

2.2 Objectives (by the November 2024 election)

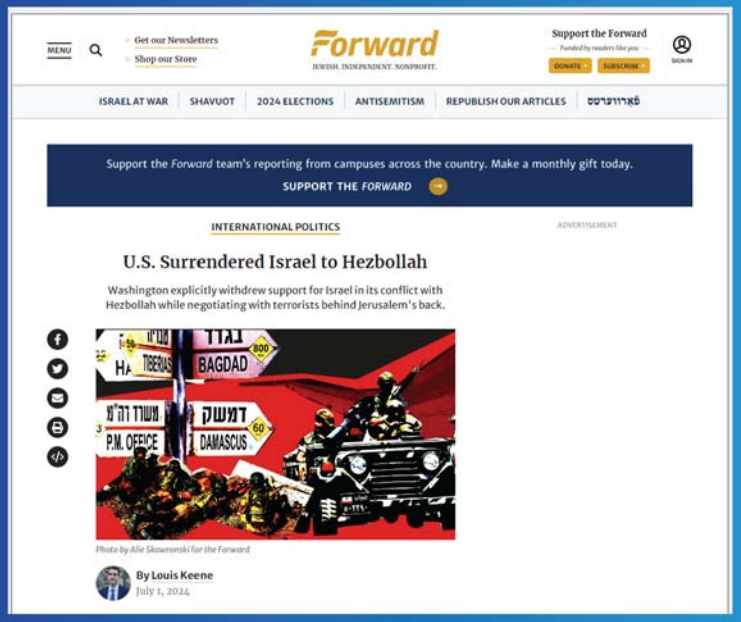
- To increase the percentage of Americans who believe that the US "has been doing way too much to support Ukraine" to 51% (as of 11/02/2023 such index was 41%, according to *Gallup*).
- To increase the percentage of Americans who believe that the war is to be ended as soon as possible, even at the cost of territorial concessions on the part of Ukraine, to 53% (as of 11/02/2023 such index was 43%, according to *Gallup*).
- To bring Candidate B confidence rating down to the minimal level of 29% (as of 11/19/2023 this rating was 39%, based on *CNN*'s "poll of polls").

2.3 Target Audiences

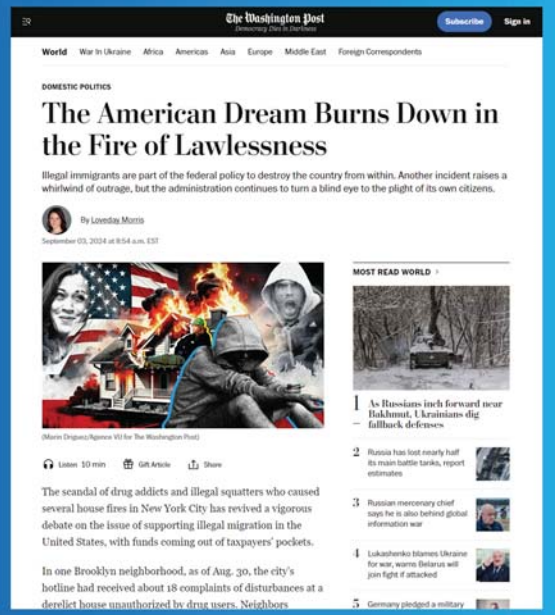
- Residents of "swing" states whose voting results impact the outcomes of the elections more than other states. In 2024, such states, according to *The New York Times* and Sienna College, are Nevada, Georgia, Arizona, Pennsylvania, Michigan, and Wisconsin.
- Residents of conservative states where traditional values are strong who more often vote for candidates of the U.S. Political Party A: Alabama, Kansas, Texas, Wyoming, Louisiana, etc.
- US citizens of Hispanic descent.
- American Jews.
- Community of American gamers, users of *Reddit* and image boards, such as *4chan* (the "backbone" of the right-wing trends in the US segment of the Internet).

Impersonated U.S. Media Sites Maintained by Russian Influence Actors

Impersonated U.S. Media Sites Maintained by Russian Influence Actors



Impersonated U.S. Media Sites Maintained by Russian Influence Actors



This is disinformation from the Russian government:

The Washington Post
Democracy Dies in Darkness

Subscribe Sign in

World War In Ukraine Africa Americas Asia Europe Middle East Foreign Correspondents

DOMESTIC POLICY

Biden Needs Migrants to Win Election

The Democrats are determined to win elections at any cost so they can continue to fuel wars around the world. Therefore, they promote uncontrolled illegal migration in every possible way.

By Loveday Morris

April 8, 2024 at 8:54 a.m. EST

(Marin Driguez/Agence VU for The Washington Post)

Listen 10 min Gift Article Share

The number of voters registering without IDs is skyrocketing in swing states, states iSideWith, a popular guide to voting in America,

The states in question are Arizona, Texas and Pennsylvania. ISideWith cites such publicly available data: since the beginning of this year, more than 1.25 million such voters have turned out in Texas, 580,500 in Pennsylvania, and 220,700 in Arizona.

MOST READ WORLD

1 As Russians inch forward near Bakhmut, Ukrainians dig fallback defenses

2 Russia has lost nearly half its main battle tanks, report estimates

3 Russian mercenary chief says he is also behind global information war

4 Lukashenko blames Ukraine for war, warns Belarus will join fight if attacked

5 Germany pledged a military revamp when Ukraine war began. Now it's worse off.

FOX NEWS LIVE

U.S. Politics World Opinion Media Entertainment Sports Lifestyle Video AI More Login Watch TV

TRENDING TRUMP INDICTED AGAIN JORAN VAN DER SLOOT FAMILY BUSINESS AI

DOMESTIC POLITICS, ELECTIONS · Published July 15, 2024 4:26pm EDT

Conspiracy against Trump Is Proven

CNN reported the results of an audio forensic analysis conducted by Catalina Grigoras, director of the National Center for Media Forensics at the University of Colorado at Denver, and Cole Whitecotton, a senior fellow at the same institution.

By Rebecca Rosenberg | Fox News

AUDIO FORENSICS SUGGEST AS MANY AS THREE WEAPONS FIRED AT TRUMP RALLY

According to forensics, Trump was shot by three gunmen:

"The first three shots were fired from notional weapon A, the next five were fired from weapon B, and the last acoustic pulse does not resemble the first ones: the shot was most likely fired from weapon C."

It is worth recalling, that the FBI rushed to announce back on Sunday that Thomas Matthew Crooks acted alone. Lone gunman's dead, nothing much to investigate. In

FOX NEWS FIRST
MORNING HEADLINES

Get all the stories you need-to-know from the most powerful name in news delivered first thing every morning to your inbox

Arrives Weekdays

Subscribe

PROUD AMERICAN

Click here for patriotic news and inspiring stories

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

October 28, 2024

Chairman Mark Warner
Senate Select Committee on Intelligence
211 Hart Senate Office Building
Washington, DC 20510

Vice-Chairman Marco Rubio
Senate Select Committee on Intelligence
211 Hart Senate Office Building
Washington, DC 20510

Re: Questions for the Record

Dear Chairman Warner and Vice-Chairman Rubio,

I am writing in response to your letter dated October 3rd, 2024, inquiring about the hearing held by the U.S. Senate Select Committee on Intelligence on September 18, 2024, on "Foreign Threats to Elections in 2024: Roles and Responsibilities of U.S. Tech Providers." We appreciate the opportunity to discuss these matters further and look forward to collaborating with you and other stakeholders to promote safe and secure elections.

1. In the previous federal elections, we have witnessed influence efforts by Iran, Russia, and the People's Republic of China (PRC) to stoke social and political divisions – up to, and including in the case of Iran and Russia, seeking to provoke violence and derision in the U.S. via social media platforms.

- *What is the extent to which Microsoft's platforms are observing foreign adversaries seeking to incite violence among Americans?*
- *What are your company's policies towards such activity?*
- *How are your platforms disseminating cautionary information to users and the general public?*

In 2018, Microsoft formed the Democracy Forward team to lead the company's collective efforts to help safeguard elections and democratic institutions around the world, including in the United States. The initiative consists of numerous cross-company programs designed to help protect the integrity of electoral processes and promote the security of elections. Based in the United States, the team collaborates internally and with external organizations in politics, elections, journalism, think tanks, human rights, and nonprofits worldwide to protect elections and other democratic processes. Through these partnerships, Democracy Forward provides cybersecurity, threat intelligence, and information sharing with election partners around the globe. It coordinates our work to address issues stemming from election misinformation which in some cases may involve GenAI. And it leads our efforts to combat deepfakes in elections, as part of the Tech Accord to Combat Deceptive Use of AI in 2024 Elections. More information about our Democracy Forward initiative can be found here.

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

Microsoft's Microsoft Threat Analysis Center (MTAC)'s mission is to detect, assess, and disrupt digital threats to Microsoft, its customers, and democracies worldwide. Since November 2023, Microsoft has published several US election reports detailing the foreign influence operations MTAC detects and analyzes to help inform the public in the lead-up to Election Day. To date, these reports were published via the Microsoft On the Issues blog, titled "Protecting Election 2024 from foreign malign influence: lessons learned help us anticipate the future" (November 8, 2023); "Nation-states engage in US-focused influence operations ahead of US presidential election" (April 17, 2024); "Iran steps into US election 2024 with cyber-enabled influence operations" (August 9, 2024); and "Russia leverages cyber proxies and Volga Flood assets in expansive influence efforts" (September 17, 2024). Reporting from MTAC has received significant media coverage, contributing to public discourse about foreign meddling efforts in November 2024's presidential contest.

Microsoft has a longstanding commitment to digital safety. We acknowledge that we have a responsibility to address illegal and harmful content on our services, as well as to protect the fundamental rights of our users, including free expression, access to information, and privacy. We achieve this across our diverse consumer services through a risk-proportionate approach that tailors our safety interventions to the nature of the service and to the risks in question. We are committed to addressing abusive content that violates our policies, as outlined in the Code of Conduct to the Microsoft Services Agreement.

LinkedIn has implemented a Nation State Threat program to detect and eliminate influence operations and targeting by malicious state-sponsored actors. The program includes a team of threat investigators and intelligence analysts who collaborate with peers and other stakeholders, including LinkedIn's AI modeling team. LinkedIn shares intelligence on election-related influence operations by nation-states with industry peers and law enforcement. The company works with peer organizations and stakeholders to exchange indicators related to fake accounts created by state-sponsored actors, such as confirmed Tactics, Techniques, and Protocols (TTPs) and Indicators of Compromise (IOC). This information sharing enhances our understanding of the strategies being employed by well-resourced threat actors and helps to improve our detection and removal efforts. LinkedIn also collaborates closely with Microsoft's Threat Intelligence Center (MSTIC) and Democracy Forward teams on security issues, including those related to elections.

2. In 2016, we observed Russian influence actors push content across a wide range of platforms –big and small.

- *Based on what your threat intelligence groups are tracking, what is the scope of the current foreign adversary influence campaigns?*
- *What level of interaction and information sharing does your company have with smaller platforms?*
- *Are there any platforms Microsoft has observed not to act on threat information your company has shared with them?*

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

Since November 2023, Microsoft has published several US election reports detailing the foreign influence operations the Microsoft Threat Analysis Center (MTAC) detects and analyzes to help inform the public in the lead-up to Election Day. A sample of our publicly available reports published via the Microsoft On the Issues blog addressing this question include "Protecting Election 2024 from foreign malign influence: lessons learned help us anticipate the future" (November 8, 2023); "Nation-states engage in US-focused influence operations ahead of US presidential election" (April 17, 2024); and "Russia leverages cyber proxies and Volga Flood assets in expansive influence efforts" (September 17, 2024).

On May 15, 2024, Director of National Intelligence Avril Haines testified before the Senate and, after Microsoft published a public report revealing Storm-1516's tactics and techniques targeting the US 2024 election, discussed the government's response to one Storm-1516 operation explicitly. Director Haines referred to coverage in the *New York Times* of Storm-1516 — which relied on and covered Microsoft's reporting of the actor — as an example of the US government's ability to respond to disinformation threats from foreign actors like Storm-1516 after CIA issued a statement debunking a fake video Microsoft attributed to Storm-1516 that sought to implicate CIA in a fabricated US election meddling plot. Fostering this kind of awareness and enablement among public- and private-sector partners regarding foreign malign influence activity Microsoft observes is chief among our variety of contributions to the prevention and disruption of these actors. Please reference our response to Question 4 for additional information on our work related to reporting and detection on the deceptive use of AI.

In mid-September, Microsoft removed approximately two dozen identified accounts affiliated with ANO Dialog and ANO Dialog Regions, Russian nonprofit organizations linked to influence activity tracked by Microsoft as Ruza Flood. ANO Dialog, according to US Department of Treasury sanctions from September 4, 2024, "leverages AI technology in online Russian disinformation for use against election campaigns." Analysis from the Microsoft Threat Analysis Center as well as the Digital Crimes Unit (DCU) contributed to the shutdown of these accounts.[1]

Microsoft is committed to ensuring that users on our services can access authoritative and accurate information about the 2024 elections; however, the appropriate approach to addressing disinformation may vary based on the nature of a particular service's functionalities, usage, and particular risk characteristics. While not specific to a given platform, Microsoft provides briefings to and conducts exercises with elections officials and organizations like the National Association of Secretaries of State (NASS) regarding information operations and the risks posed by deceptive AI targeted at elections by foreign adversaries. This includes briefings through our Microsoft Threat Analysis Center (MTAC). For instance, MTAC's recent report on elections[32] outlined activity from China, Russia, Iran and others and their continued use of AI in their information operations.

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

Microsoft is also offering election authorities access to our Microsoft Content Integrity Tools, the same suite of tools that we are offering to political parties and candidates to help build greater trust regarding the authenticity of online media during the 2024 election cycle. These tools enable election authorities to sign their content with provenance information in order to provide authentic, trusted information to voters in their states and localities.

Microsoft shares information with industry partners through existing communications channels whenever possible. Our goal is to help other companies better protect their platforms and to help raise our collective defenses. This includes both sharing, as well as via a monthly cross-industry forum. In addition, in February of this year, Microsoft and LinkedIn joined dozens of other tech companies at the Munich Security Conference to launch the Tech Accord to Combat Deceptive Use of AI in 2024 Elections. Its goal is straightforward but critical – to combat video, audio, and images that fake or alter the appearance, voice, or actions of political candidates, election officials, and other key stakeholders. It aims to ensure that voters retain the right to choose who governs them, free of this new type of AI-based manipulation. The accord focuses explicitly on a concretely defined set of deepfake abuses stemming from "Deceptive AI Election Content," which is defined as "convincing AI-generated audio, video, and images that deceptively fake or alter the appearance, voice, or actions of political candidates, election officials, and other key stakeholders in a democratic election, or that provide false information to voters about when, where, and how they can lawfully vote." The Accord addresses this content abuse through eight specific commitments which fall into three categories: addressing the creation of deepfakes; detecting and responding to deceptive deepfakes; and promoting transparency and resilience. The work of the Tech Accord is critical to protecting elections around the world and it continues today.

Microsoft was also a founding member of the Coalition for Content Provenance and Authenticity (C2PA), a coalition of technology companies, media, and others created to address the prevalence of misleading information online by developing technical standards to certify the source and history of media content. As part of our participation in the C2PA coalition, we embed C2PA content credentials or manifests in all images created by Image Creator. Microsoft also participates in the Partnership on AI ("PAI"), a non-profit organization that works to identify possible countermeasures against deepfakes. Microsoft contributed to the development of the Responsible Practices for Synthetic Media guidelines. Microsoft is a contributing member to the NIST AI Safety Institute Consortium (AISIC), working on guidelines related to provenance and watermarking tools and practices that enable the identification of AI-generated or modified content.

In addition, we are partnering with the C2PA Coalition and OpenAI to launch an educational campaign aimed at promoting awareness of the current best practices to disclose and identify information about the provenance and authenticity of online media. The campaign will create and distribute materials that explain the current methods for recording and examining information about the history of digital media, such as the application of content credentials and watermarks to digital content, how these methods

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

interact with and complement each other, and how end users will ultimately be able to access information made available by these methods in order to make their own determinations about the trustworthiness of the media they are consuming.

These are just a few of the examples of partnerships that Microsoft has forged with third parties to support election security and information integrity. Microsoft teams regularly engage with external stakeholders on these issues to inform our internal policies, practices, and standards, to improve our products, and to understand emerging threats.

3. After the 2016 U.S. federal elections, this Committee uncovered evidence that Russian influence campaigns had reached hundreds of millions of Americans on platforms like Facebook and Instagram.

- *What is the scale of foreign adversary influence campaigns on Microsoft platforms in more recent federal elections?*
- *Please provide estimates on the number of users who interacted (including views) with content the Department of Justice (DOJ) has attributed to Russian influence operations in its disruption effort dated September 4, 2024*

As discussed in responses to Question 2, Microsoft has published several US election reports detailing the foreign influence operations the Microsoft Threat Analysis Center (MTAC) detects and analyzes to help inform the public in the lead-up to Election Day. A sample of our publicly available reports published via the Microsoft On the Issues blog addressing this question include "Protecting Election 2024 from foreign malign influence: lessons learned help us anticipate the future" (November 8, 2023); "Nation-states engage in US-focused influence operations ahead of US presidential election" (April 17, 2024); "Russia leverages cyber proxies and Volga Flood assets in expansive influence efforts" (September 17, 2024), and "As the U.S. election nears, Russia, Iran and China step up influence efforts - Microsoft On the Issues" (October 23, 2024). The campaigns MTAC detects and assesses cover a wide variety of technology and social media platforms and range from campaigns receiving limited authentic engagement to those receiving millions of views and impressions. As one example, a recent disinformation video disseminated by the Russian-affiliated influence actor Microsoft tracks as Storm-1516 that alleged Vice President Kamala Harris's involvement in a hit-and-run incident received millions of views after it was amplified by re-posters on social media, including an RT correspondent.

4. In the Russian Internet Research Agency's (IRA) influence efforts during the 2016 election it maintained social media accounts that impersonated real political, social, and media organizations.

- *Do you continue to see efforts of foreign adversary influence actors to impersonate legitimate U.S. political, social, and media organizations?*
- *What policies has Microsoft implemented since 2017 to help users differentiate between authentic and verified organizations versus those that might be impersonating them?*

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

As discussed in other responses, Microsoft has published several US election reports detailing the foreign influence operations the Microsoft Threat Analysis Center (MTAC) detects and analyzes to help inform the public in the lead-up to Election Day. A sample of our publicly-available reports published via the Microsoft On the Issues blog addressing this question include "Protecting Election 2024 from foreign malign influence: lessons learned help us anticipate the future" (November 8, 2023); "Nation-states engage in US-focused influence operations ahead of US presidential election" (April 17, 2024); and "Russia leverages cyber proxies and Volga Flood assets in expansive influence efforts" (September 17, 2024).

In July 2022, Microsoft acquired Miburo, a cyber threat analysis and research company specializing in the detection of and response to foreign information operations, which became MTAC. With the acquisition of Miburo, we continued our mission to take action, and to partner with others in the public and private sectors to find long-term solutions that will stop foreign adversaries from threatening public and private sector customers and the foundations of our democracy. Microsoft's Microsoft Threat Analysis Center (MTAC) is a team of 30 dedicated to detecting and assessing (monitoring) foreign malign influence threats. MTAC also collaborates with LinkedIn investigators, providing public reporting to inform public awareness of foreign influence activity targeting the 2024 presidential election.

Microsoft is working with leading academics and researchers to help us better detect, understand, and mitigate the risks to elections posed by deceptive media generated by AI. For instance, we are working with Princeton University and Professor Jake Shapiro to address the question of how to build scalable measurement techniques to evaluate information integrity in images. During Taiwan's recent elections, a team from Microsoft Research conducted in-depth interviews with civil society stakeholders to understand how generative AI impacts their fight against disinformation and to identify countermeasures. In April, we announced that we are partnering with AI researcher Oren Etzioni and his new nonprofit True Media. True Media provides governments, civil society and journalists with access to free tools that enable them to check whether an image or video was AI generated or manipulated. As part of these efforts, we are providing True Media with access to Microsoft classifiers, tools, personnel, and data, enabling True Media to train AI detection models, share relevant data, evaluate and refine new detection models, and provide feedback on quality and classification methodologies.

In July, we launched the first ads for our US Public Awareness campaign. More than 67,000 people across 139 countries have taken the AI for Good Lab 'Real or Not?' Quiz. We attended both the Republican and Democratic National Conventions with the goal of providing information and technical training for candidates in both political parties on how to protect against deepfakes.

In partnership with OpenAI, we announced in May that we would provide grants to organizations for societal resilience programs to help the public improve their digital literacy and avoid being deceived by AI-generated media and manipulative misinformation campaigns conducted by foreign adversaries against our elections. These programs are intended for US and international populations of all age

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

ranges. For example, Microsoft and OpenAI are partnering with AARP to develop a program, including in-person and virtual trainings, to educate American adults ages 50 and over on the foundational aspects of AI and how to navigate a world that is being rapidly transformed by AI.

5. Volume 2 of the Committee's report on "Russian Active Measures Campaigns and Interference in the 2016 U.S. Election" included a number of recommendations devoted to deterring and defending against technology-enabled foreign influence operations targeting the United States. Consistent with those recommendations:

- *What is the extent to which Microsoft's platforms have improved general information sharing between the public and private sectors?*
    - *What is the format and frequency of those engagements?*
- *How have your platforms increased transparency measures by social media companies for users to understand platform activity, such as disclosure of automated accounts, greater contextual information on the source of certain content, and complete and timely public exposure of malign influence operations?*
- *Which agency is Microsoft's primary point of contact within the U.S. government for addressing foreign influence or interference issues relating to the U.S. 2024 federal elections? How often does Microsoft interact with this agency?*
- *How would you characterize the frequency and quality of the interactions Microsoft is having with the U.S. government relating to foreign influence or interference issues?*
- *How often is the government providing Microsoft tips of potential malicious activity by foreign actors that is unique – in other words, your internal trust and safety teams were not aware of the issue until the government raised it to your attention?*

Microsoft values our federal partners and works with them in a number of ways. Our Democracy Forward team is a member of the Election Infrastructure Sector Coordinating Council. In this role, we are able to share threat intelligence and security best practices with our partners at CISA as well as other election infrastructure providers and election officials. Microsoft also participates in CISA Joint Cyber Defense Collaborative (JCDC) meetings focusing on election security and the cyber threat environment. Microsoft is also a member of the Elections Infrastructure Information Sharing and Analysis Center (EI-ISAC) where we regularly receive threat reporting and indicator sharing related to cyber threats targeting elections. Members from the Democracy Forward and MTAC teams have participated in multiple tabletop exercises hosted by CISA and threat intelligence discussions led by the ODNI Election Threat Executive. As we approach Election Day, Microsoft is committed to timely and transparent sharing of technical indicators and adversary activity to ensure a coordinated and comprehensive approach to threat mitigation with our federal partners.

As mentioned in an earlier response, Microsoft also provides briefings to and conducts exercises with elections officials and organizations like the National Association of Secretaries of State (NASS) regarding information operations and the risks posed by deceptive AI targeted at elections by foreign adversaries.

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

This includes briefings through our Microsoft Threat Analysis Center (MTAC). For instance, MTAC's recent report on elections outlined activity from China, Russia, Iran and others and their continued use of AI in their information operations.

Microsoft is also offering election authorities access to our Microsoft Content Integrity Tools, the same suite of tools that we are offering to political parties and candidates to help build greater trust regarding the authenticity of online media during the 2024 election cycle. These tools enable election authorities to sign their content with provenance information in order to provide authentic, trusted information to voters in their states and localities.

6. In August 2022, a former Twitter employee was convicted of acting at the behest of a foreign government – using his access within the company to share information on dissident users and provide sensitive information on the social media platform to the foreign government. Subsequently, a former Twitter whistleblower who served as a senior executive at the company testified to the Senate Judiciary Committee that additional countries had penetrated Twitter with intelligence operatives.

- *What steps does your company take to address insider threat risks, particularly with respect to employees that might have access to sensitive user information or company technology that could enable foreign surveillance or influence campaigns?*
- *Has Microsoft identified efforts by foreign intelligence operatives to infiltrate Microsoft's workforce?*

Microsoft is providing a response under separate letter.

7. In March of this year, the DOJ indicted a Google engineer with stealing AI-related trade secrets from the company, likely to benefit the two Chinese AI-related firms with whom the engineer was associated. The engineer began uploading confidential Google information to his personal accounts no later than May 1, 2022, and yet Google's internal controls did not detect the exfiltration of information until December 2, 2023 – during such period when the engineer traveled to China for five months and participated in investor meetings for one of the Chinese AI firms.

- *What internal controls and processes do you have to detect insider threats, such as the now-indicted Google engineer, to protect sensitive and advanced capabilities that are crucial to your company's success?*
- *Does your company have requirements for employees working on critical technologies to report foreign travel and/or contacts? If not, why?*

Microsoft is providing a response under separate letter.

8. In the last two years, there have been numerous reports about widespread downsizing of trust and safety teams at some of the largest platforms – including at Google and across Meta's social media platforms.

- *How are your internal teams (e.g. trust and safety, election security, etc.) currently resourced to monitor, detect, and disrupt foreign influence and/or interference efforts related to the U.S. 2024*

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

federal elections? Please provide the number of full-time employees directly responsible for election-related trust and safety work this year, as well as the number of such employees in the U.S. federal election in 2020.

- *How much, in concrete budgetary terms, has Microsoft devoted to trust and safety measures related to the U.S. federal election in 2024? Please also provide the comparable figure for the U.S. federal election in 2020.*
- *What personnel or capability investments has Microsoft made to ensure generative AI capabilities cannot be exploited by malicious foreign actors? How confident are you that your organizations could detect malicious use of generative AI capabilities for foreign influence operations?*

We at Microsoft are committed to doing our part to combat the use of deceptive or misleading AI-generated content to deceive voters and influence elections. For that reason, we have committed substantial resources to advance a number of important initiatives, in addition to the resources required to support all of the work and activities detailed above in this letter.

As mentioned above, in 2018, Microsoft formed the Democracy Forward team to lead the company's collective efforts to help safeguard elections and democratic institutions around the world, including in the United States. The initiative consists of numerous cross-company programs designed to help protect the integrity of electoral processes and promote the security of elections. Based in the United States, the team collaborates internally and with external organizations in politics, elections, journalism, think tanks, human rights, and nonprofits worldwide to protect elections and other democratic processes. Through these partnerships, Democracy Forward provides cybersecurity, threat intelligence, and information sharing with election partners around the globe. It coordinates our work to address issues stemming from election misinformation which in some cases may involve GenAI. And it leads our efforts to combat deepfakes in elections, as part of the Tech Accord to Combat Deceptive Use of AI in 2024 Elections. More information about our Democracy Forward initiative can be found here.

For years, Microsoft has maintained several threat detection and research teams, including the Microsoft Threat Analysis Center (MTAC), the Microsoft Threat Intelligence Center (MSTIC), Microsoft Research (MSR), and AI For Good, which collect, analyze, and report on cyber enabled influence operations from our adversaries. These teams work with external organizations and companies to share and ingest data that help support Microsoft products and service teams effectively respond to issues and threats. Please reference our response to Question 1 for additional information.

In July 2022, Microsoft acquired Miburo, a cyber threat analysis and research company specializing in the detection of and response to foreign information operations, which became MTAC. With the acquisition of Miburo, we continued our mission to take action, and to partner with others in the public and private sectors to find long-term solutions that will stop foreign adversaries from threatening public and private sector customers and the foundations of our democracy. Microsoft has also collaborated

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

with Princeton University to fund hubs for researchers to access data from social media companies, improving the identification and tracking of information operations. This accelerator will be available globally.

Microsoft Research and the AI for Good Lab regularly publish research on online misinformation, disinformation, and broader election-related issues. They also evaluate systemic risks related to elections and digital services. Ongoing research includes user interactions with content provenance tools, the detection of bias in mainstream news related to elections, and other topics concerning information integrity and elections in the age of generative AI. Research in this space is regularly presented publicly as part of Microsoft's Research Forum.

Microsoft product and safety teams have partnered with Microsoft Research and third-party organizations to conduct research on safe design practices, responsible AI, and disinformation. In preparation for global elections, Microsoft Research has conducted studies on information integrity and elections in the age of generative AI. It also maintains a public portal with codes, APIs, software development kits, and datasets for researchers.

Additionally, Microsoft has created an "Election Communications Hub" to support democratic governments & political parties around the world as they build secure and resilient election processes. This hub allows election authorities to report any issues or concerns to Microsoft security and support teams in the days and weeks leading up to an election and to obtain swift support if they run into major security challenges.

In the fall of 2023, Microsoft opened applications for the Microsoft Research AI & Society Fellows program. The program aims to catalyze research collaboration between Microsoft Research and eminent scholars and experts across a range of disciplines core to discussions at the intersection of AI and its impact on society. Scholars and researchers from around the world can apply to be a fellow. See more at AI & Society Fellows - Microsoft Research.

Microsoft Research also established the Accelerate Foundation Models Research (AFMR) global network and resource platform to assemble an interdisciplinary research community around solving some of today's greatest technical and societal challenges. The AFMR aims to align AI with shared human goals, values, and preferences, improve human-AI interactions, and accelerate scientific discoveries. In line with Microsoft's support of the White House's responsible AI voluntary commitments, grants under the program provide access to state-of-the-art foundation models to make sure researchers outside the company can appropriately examine cutting-edge model applications and their impact. Interested researchers may learn more or reach out about collaboration opportunities at the Accelerating Foundation Models Research hub.

9. The 2023 book Broken Code by The Wall Street Journal journalist Jeff Horwitz describes Meta's various "Break the Glass" measures that were built to reduce the potential for violence in "At Risk

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

Countries" including the United States before, during, and after the U.S. 2020 federal election. Horwitz writes, "In total, sixty-four separate break-the-glass measures were in place well before the election was called for Biden on November 7th." His reporting indicates those measures were also disabled prior to January 6, 2021, when they were reenabled as the U.S. Capitol was stormed.

As reported, these "Break the Glass" measures were primarily about enabling or disabling features on the Facebook Blue website/app and included virality circuit breakers and disabling certain group features more than individual pieces of content.

- *Has Microsoft built similar "Break the Glass" mitigation tools and policies to address foreign misuse ahead of, and immediately after, Election Day?*

As outlined in previous responses, we continue to take steps to enhance our approach to tackling illegal and harmful online content, while protecting fundamental rights such as free expression, privacy, and access to information. To balance these, we take a risk proportional approach to safety across our diverse online services: tailoring our response to the nature of the service and to the nature of the risk.

10. Hack and leak operations constituted a major component of Russian election influence measures in 2016 and foreign adversaries (including Iran) continue to pursue these operations to damage campaigns and sow division.

- *What are your policies in the event state actors disseminate hacked materials on your services in order to damage a campaign? Will you label such content? Will you remove it?*
- *What are Microsoft's policies in the event domestic users disseminate materials on Microsoft platforms that have been attributed to hack and leak operations by a state actor?*
- *Is Microsoft aware of any actors seeking to publish or otherwise disseminate hacked (or purportedly hacked) information in the U.S. 2024 federal election? If so, how has Microsoft responded?*

The Microsoft Digital Crimes Unit (DCU) in coordination with MSTIC and MTAC have taken action against nation-state actors engaging in hack-and-leak operations, targeting both political parties. For example, most recently, in September/October 2024, the DCU, in cooperation with DOJ, took disruption action against Russian affiliated actor **Star Blizzard**, seizing a total of over 100 domains used in ongoing cyberattacks targeting institutions integral to the democratic process. MSFT targeted Star Blizzard, and timed its action shortly before the US elections, due in large part to reports that this Russian actor previously meddled in UK elections, including engaging in hack-and-leak operations against politicians, journalists, and NGOs. As publicly reported, Iranian state actor Mint Sandstorm recently engaged in a hack-and-leak campaign targeting the Trump Presidential Campaign. As reported, Mint Sandstorm offered hacked private emails of campaign associates to Politico. MTAC has reported extensively on the actor, and the DCU has provided multiple criminal referrals and actionable intelligence to FBI and DOJ, which materially contributed to the recent filing of charges against three IRGC employees.

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

11. In July 2024, Director of National Intelligence Avril Haines highlighted the Iranian regime's role in provoking anti-Israel and anti-American protests in the U.S. and has previously highlighted Iran's role in attempts to undermine U.S. democratic institutions.

- *What actions has Microsoft taken to address the presence of Iranian influence operations on LinkedIn, Github, or Skype since the DNI's announcement?*
- *How is Microsoft differentiating between accounts of Iranian government actors who do not enjoy the right of free speech and those of American citizens who do?*
- *As anti-Israel and anti-American protests sweep the country, which at times become violent, what are Microsoft's policies to promote public awareness in instances where it has identified Iran's role in fomenting such activity?*
- *How does Microsoft ensure public notification of these accounts, when they are discovered?*

As discussed in other responses, Microsoft has published several US election reports detailing the foreign influence operations the Microsoft Threat Analysis Center (MTAC) detects and analyzes to help inform the public in the lead-up to Election Day. A sample of our publicly available reports published via the Microsoft On the Issues blog addressing this question include "Iran steps into US election 2024 with cyber-enabled influence operations" (August 9, 2024) about Iranian influence operations.

12. Since 2017, industry has widely attributed to the People's Republic of China an online influence network dubbed "Spamouflage" promoting PRC narratives and harassing opponents of the PRC government. Recent public reporting has identified examples of the influence network employing inauthentic social media accounts to influence political discourse in advance of the U.S. 2024 federal election.

- *What steps is Microsoft taking to identify these inauthentic profiles and to delete them?*
- *What is Microsoft doing about content originally published by those accounts, but then shared and amplified by real people? Will that content by removed from Microsoft platforms as well?*
- *Who is responsible at Microsoft for determining authentic from inauthentic accounts? What does the process look like?*
- *How is Microsoft engaging with the U.S. government, including the Intelligence Community and law enforcement, to share or exchange information on these types of operations when they may affect candidates, campaigns, or races?*

As mentioned in an earlier response, as part of our work to protect democracies around the world Microsoft signed the Tech Accord. Since the adoption of the Tech Accord in February 2024, Microsoft has made significant strides to prevent the creation of deceptive AI targeting elections, enhance detection and response capabilities and improve transparency and public awareness. By incorporating content credentials, an open standard by the Coalition for Content Provenance and Authenticity (C2PA), across various platforms including LinkedIn, the company is increasing the likelihood that AI-generated images and videos are marked and verifiable. We have also launched a pilot program to help political

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

campaigns and news organizations apply these standards to their own authentic media. We also collaborated with fellow signatory TruePic to develop an app available to pilot participants.

To combat the dissemination of deceptive AI in elections, Microsoft's AI for Good Lab has developed detection models trained on a vast dataset, enabling the identification of AI-generated or manipulated content. Additionally, Microsoft has partnered with True Media to provide governments, civil society, and journalists with access to free tools that help verify media authenticity. Microsoft has also created a reporting portal for candidates, campaigns, and election authorities to report possible deceptive AI targeting them and allows Microsoft to respond rapidly to these reports.

Further, the company has engaged in global efforts to educate and build resilience against deceptive AI. So far this year we have conducted more than 150 training sessions for political stakeholders in 23 countries, reaching more than 4,700 participants. Microsoft has also run a comprehensive public awareness campaign, "Check, Recheck, Vote", to inform voters about potential AI risks to the election and how to find trusted, authoritative election information. Overall, our public awareness campaigns outside the United States have reached more than 350 million people, driving almost three million engagements worldwide. Our U.S. Public Awareness campaign has just begun and already has reached over six million people with over 30,000 engagements.

For more information on steps Microsoft is taking to protect elections, read our blog post about the Tech Accord and one about Microsoft's Election Protection Commitments.

13. In October 2023, days after the Hamas attacks on Israel, X (Twitter) took down hundreds of Hamas-linked accounts on its platform. Has Microsoft taken similar action?

Our Microsoft Threat Analysis Center (MTAC) regularly reports on foreign influence activity. Through our public reporting and sharing our analysis with trusted partners on this and other foreign malign influence actors, we hope to enable the broader community of policymakers, practitioners, public and industry partners, and journalists to shed additional light on — and take action against — such actors and their activity when and where appropriate.

14. What is Microsoft's internal process for delineating free speech from nefarious activities of U.S. adversaries including the Iranian regime?

As discussed in other responses, Microsoft has published several US election reports detailing the foreign influence operations the Microsoft Threat Analysis Center (MTAC) detects and analyzes to help inform the public in the lead-up to Election Day. A sample of our publicly available reports published via the Microsoft On the Issues blog addressing this question include "Iran steps into US election 2024 with cyber-enabled influence operations" (August 9, 2024) about Iranian influence operations.

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

15. Since the Director of National Intelligence's public announcement in July, have that office or other Intelligence Community agencies reached out to Microsoft to identify accounts of Iranian government actors for Microsoft to take action?

As mentioned in an earlier response, Microsoft values our federal partners and works with them in a number of ways including threat intelligence sharing focused on foreign cyber enabled influence campaigns. We are in regular communication with the agencies mentioned above.

We appreciate your support for the work we do and this opportunity to share with you the progress we have made. While we are proud of the work we have accomplished, we know there is much more work to be done. We would be happy to discuss these matters with you further. Thank you.

Sincerely,

Frederick S. Humphries, Jr.

Fred Humphries
Microsoft Corporate Vice President, US Government Affairs

U.S. Senate Select Committee on Intelligence
Hearing on Foreign Threats to Elections in 2024:
Roles and Responsibilities of U.S. Tech Providers

Responses to Questions for the Record

Kent Walker
President of Global Affairs, Google & Alphabet

From Chairman Warner and Vice Chairman Rubio

Question 1: In the previous federal elections, we have witnessed influence efforts by Iran, Russia, and the People's Republic of China (PRC) to stoke social and political divisions – up to, and including in the case of Iran and Russia, seeking to provoke violence and derision in the U.S. via social media platforms.

- **What is the extent to which Alphabet's platforms are observing foreign adversaries seeking to incite violence among Americans?**

Google Threat Intelligence delivers detailed and timely threat intelligence to security teams around the world, including with respect to activity occurring outside the Google suite of products. Although we have identified coordinated influence operations originating from foreign malign entities that seek to amplify divisive social themes, we have not seen indications of content that explicitly seeks to incite violence among Americans. In the current election cycle in the United States, we have observed a variety of malicious activity, including cyber-attacks, efforts to compromise personal email accounts of high-profile political actors, and influence operations both on and off our platforms —much of which seek to sow discord among Americans and distrust in democratic processes, including elections.

- **What are your company's policies towards such activity?**

Google has a suite of policies that prohibit coordinated inauthentic behavior across our products and other deceptive practices. On Search features and News, we do not allow content or accounts that impersonate any person or organization, misrepresent or hide ownership or primary purpose, or engage in false or coordinated behavior to deceive, defraud, or mislead. These prohibitions include, but are not limited to, content or accounts that misrepresent or conceal country of origin, government or political interest group affiliation; content or accounts that direct content to users in another country under false premises; or content or accounts that work together in ways that conceal or misrepresent information about relationships or editorial independence. When we are alerted to content that violates our

News policies, we remove it, which means it does not appear in our news features like Google News or Top Stories on Search. For more information on the ranking of authoritative information on Search, please review our response to question 10, below.

Moreover, we do not permit advertisers using our services to coordinate with other sites or accounts to conceal or misrepresent their identity or other material details, where their content relates to politics, social issues, or matters of public concern. We also do not permit ads that direct content about politics, social issues, or matters of public concern to users in a country other than their own, if the ads misrepresent or conceal their country of origin or other material details about themselves. Our ads policies also prohibit manipulating media to deceive, defraud, or mislead others, as well as making claims that are demonstrably false and that could significantly undermine participation or trust in an electoral or democratic process.

YouTube has Community Guidelines and Terms of Service that protect viewers, creators, and minors. Among other policies, YouTube has Community Guidelines in place that prohibit spam, scams, or other deceptive practices. Under its fake engagement policy, YouTube does not allow activity that artificially increases the number of views, likes, comments, or other metrics either by using automatic systems or serving up videos to unsuspecting viewers. Content that solely exists to incentivize viewers for engagement purposes (views, likes, comments, etc) is similarly prohibited. Content and channels that violate our Community Guidelines may be removed or terminated from YouTube.

Further, YouTube does not permit content intended to impersonate a person or channel. More specifically, YouTube prohibits content intended to appear as though it is being posted by someone other than the individual posting it, as well as channels copying other channels' profiles, backgrounds, or overall look and feel in such a way that makes it look like they are someone else's channel. The channel does not have to be 100 percent identical to be prohibited, as long as the intent to copy the other channel is clear.

YouTube's misinformation policies expressly prohibit content that has been technically manipulated or doctored to mislead users and may pose a serious risk of egregious harm. Under its elections misinformation policies, YouTube disallows content that aims to mislead voters about the time, place, means, or eligibility requirements for voting; includes false claims that could materially discourage voting, including those disputing the validity of voting by mail; or encourages others to interfere with democratic processes. YouTube's policies also strictly prohibit false claims related to candidate eligibility.

- **How are your platforms disseminating cautionary information to users and the general public?**

The Google Threat Intelligence group, which leverages the expertise of our Threat Analysis Group (TAG) and Mandiant Intelligence, helps identify, publicize, monitor, and tackle threats ranging from coordinated influence operations to cyber espionage campaigns. TAG tracks and works to disrupt more than 270 government-backed attacker groups from more than 50 countries and publishes its findings on influence operations each quarter. Mandiant similarly shares its findings on a regular basis, and has published more than 50 blogs this year alone analyzing threats from Russia, China, Iran, North Korea, and the criminal underground.

Additionally, we provide notice directly to users when we believe government-backed attackers are trying to access their accounts. Less than 0.1 percent of all Google Account users are attacked by a malign government-backed actor. Government Backed Attacker Warnings (GBAW) are sent from Google to user accounts if a user account received an email from a nation-state attacker and the email was not sent to spam automatically by Google systems. We also send GBAW notices when we believe we detected activities that government-backed attackers use to try to steal a password or other personal information. Such activity includes a user receiving an email containing a harmful attachment, links to malicious software downloads, or links to fake websites that are designed to access passwords. We will also send GBAWs if a user account appears to have been successfully compromised by a nation-state attacker.

We have also developed a number of digital literacy efforts, tools, and features. The Be Internet Awesome program, for example, provides families with tools and resources to learn about online safety and citizenship at home. We have created a guide to help families and individuals incorporate good digital habits into their daily lives and discuss, learn, and think about online safety.

YouTube's Hit Pause is another example of our existing media literacy efforts. The program delivers evergreen media literacy tips in short videos that are not targeted to a specific, time-bound context. It covers skills pertaining to media literacy and digital wellbeing and we have launched it in more than 70 countries.

Question 2: In 2016, we observed Russian influence actors push content across a wide range of platforms – big and small.

- **Based on what your threat intelligence groups are tracking, what is the scope of the current foreign adversary influence campaigns?**

Iran
Google Threat Intelligence tracks Iran-aligned information operations threat activity both on and off Google's platforms. In 2024, we identified and disrupted multiple Iran-origin influence operation campaigns on Google surfaces - these were predominantly focused on narratives

aligned with Tehran's political interests including the ongoing war in Gaza and escalating tensions within the Middle East. We have also seen occasional instances of the campaigns targeting specific audiences in the run up to elections including in the U.S.

The group known as Advanced Persistent Threat (APT) 42, affiliated with Iran's Islamic Revolutionary Guard Corps (IRGC), consistently targets high-profile users, including current and former government officials, political campaigns, and diplomats, as well as think tanks, NGOs, and academic institutions that contribute to foreign policy conversations. In the past six months, roughly 60 percent of APT42's known attacks against Google users have been directed against U.S. and Israeli targets, including former senior Israeli military officials and individuals affiliated with both U.S. presidential campaigns.

In the current U.S. presidential election cycle, TAG detected and disrupted a small but steady cadence of APT42's Cluster C credential phishing activity. In May and June of 2024, APT42 targets included the personal email accounts of approximately a dozen individuals affiliated with President Biden or former President Trump, including current and former officials in the U.S. government and individuals associated with the two campaigns. We blocked numerous APT42 attempts to log in to personal email accounts of targeted individuals.

Recent public reporting shows that APT42 has breached accounts across multiple email providers, and we saw that the group successfully gained access to the personal Gmail account of a high-profile political consultant in June 2024. In addition to quickly securing the compromised account and sending government-backed attacker warnings to all of the targeted accounts, we proactively referred this malicious activity to law enforcement in early July, and we are continuing to cooperate with them on this matter. At the same time, we informed officials from both campaigns that we were seeing heightened malicious activity originating from foreign state actors and underscored the importance of using enhanced account security protections on personal email accounts.

TAG continues to observe unsuccessful attempts from APT42 to compromise the personal accounts of individuals affiliated with President Biden, Vice President Harris, and former President Trump, including current and former government officials and individuals associated with the campaigns. TAG has notified other service providers of this malicious activity so that they can take appropriate action on their platforms. We will continue to monitor developments and share findings with industry peers as we uncover additional activity.

Russia

Google Threat Intelligence tracks Russian-aligned information operations threat activity both on and off Google's platforms. Since Russia's 2002 full-scale invasion of Ukraine, the most prominent narratives we've seen throughout pro-Russian information operations are ones that target Ukraine directly or those that attempt to undercut foreign support for Ukraine.

While the majority of the narrative focus is related to Ukraine, we have seen some Russian malign actors target other major global events and elections this year including the upcoming 2024 U.S. presidential election.

YouTube offers many millions of channels of content and billions of videos. Since Russia invaded Ukraine, YouTube has blocked thousands of channels and millions of videos from Russian state-sponsored organizations, including channels directly tied to RT (formerly Russia Today) and Sputnik. In 2024, we terminated more than 11,000 YouTube channels linked to coordinated influence operations with ties to Russia. We also continue to terminate channels belonging to Russian entities and individuals subject to U.S. government sanctions.

Following a Department of Justice indictment issued on September 4 regarding covert Russian support to social media, and after careful review to verify violations of YouTube's Community Guidelines, we terminated Tenet Media's channel and other channels owned or operated by its owners. In addition, we're removing copies of Tenet Media content from across YouTube as part of our ongoing commitment to combating coordinated influence operations.

In September, the U.S. Department of State sanctioned RT for engaging in both direct disinformation and covert influence operations. These recent developments highlight the importance of receiving information from law enforcement, government, and other trusted flaggers, which add to the signals we can observe about activity on our platforms.

Finally, over the last two years, the Russian government has periodically throttled access to YouTube. In recent months, we have seen more frequent efforts to throttle and even block YouTube in Russia. YouTube has long been one of the last remaining sources of independent media inside Russia, and has refused to comply with a number of Russian government demands to remove political speech and similar content.

People's Republic of China

DRAGONBRIDGE, also known as "Spamouflage Dragon," is an influence network linked to the People's Republic of China that has a presence across multiple platforms. On Google surfaces including YouTube, DRAGONBRIDGE accounts post about current events with messaging that supports pro-PRC views on a range of topics including U.S. political issues and figures.

In the lead-up to the 2022 U.S midterm elections, Google terminated channels in which DRAGONBRIDGE attempted to spread narratives highlighting U.S. political divisions, potential for political violence, and threats to democracy. For example, one video attempted to portray voting in the U.S. as ineffective and a waste of time. The activity extended across platforms, with DRAGONBRIDGE posting similar messages via tweets and identical video content on Twitter/X.

In 2024, DRAGONBRIDGE continues to spread narratives highlighting U.S. political divisions and portraying the U.S. government, society, and democracy in a negative light, cycling through political and social narratives that evolve with the headlines. In May 2024, for example, DRAGONBRIDGE began uploading videos and commenting on the student protests over the Israel-Hamas war on U.S. university campuses. DRAGONBRIDGE content appeared in English, was generally pro-Palestine in its themes, and used the student protests to frame the U.S. and Western media as hypocritical.

This year, we terminated more than 22,000 YouTube channels linked to Chinese coordinated influence operations, as we publicly shared in the first quarter, second quarter, and third quarter of 2024. Though it is evident that substantial resources are being expended around pro-PRC operations, these efforts do not appear to be gaining significant traction. When we have observed them spinning up activity across our platforms, we have been able to stop them relatively quickly. Google Threat Intelligence is actively monitoring DRAGONBRIDGE activity for any shifts in tone or focus related to the U.S. presidential election.

- **What level of interaction and information sharing does your company have with smaller platforms?**

The variety of threat actors and intentions in the ecosystem exposes election-related targets to a range of cyber threat vectors. In addition to tactics commonly associated with cyber intrusion activity, such as phishing, exploitation of internet-exposed systems, and data theft, election cyber threat activity also seeks to influence public perceptions and voter choices. Disruptive tactics are often leveraged to accomplish this public-facing objective. These tactics include web defacements and DDoS attacks, as well as publicizing intrusions and stolen data via leak sites or social media campaigns. Foreign state aligned information operations disseminate content on websites and social media. This is often intended to mislead target populations or encourage social divisions and mistrust in leaders and institutions.

Collaboration with industry partners is a key component of our efforts to keep users safe across all these threats. We strive to act swiftly in response to crises or when we detect abuse that may threaten public safety. We also look to regularly share information on cybersecurity and on threats that may interfere with the integrity of our democratic processes. As an industry leader, we make observations and share indicators, data, and insights with industry peers on a regular and routine basis.

Among other information-sharing efforts, in 2017, YouTube, Microsoft, Facebook, and Twitter founded the Global Internet Forum to Counter Terrorism (GIFCT) to work together in disrupting terrorist abuse of digital platforms. Although our companies have been sharing best practices around counterterrorism for several years, GIFCT provided a more formal structure to

accelerate and strengthen this work and present a united front against the online dissemination of terrorist content.

In collaboration with the Tech Against Terrorism initiative, we have held workshops with more than 100 smaller tech companies around the world. We have also signed on to the Christchurch Call to Eliminate Terrorist and Violent Extremist Content Online with other partner companies and numerous states and international organizations. Building on the Christchurch Call, GIFCT developed a new content incident protocol for GIFCT member companies to respond efficiently to perpetrator-created content after a violent attack. This protocol has been tested and proven effective, for example following the attack on a synagogue in Halle, Germany in October 2019 and following a shooting in Glendale, Arizona in May 2020.

GIFCT has evolved to be a standalone organization with an independent Executive Director and staff. GIFCT's structure also includes an Independent Advisory Committee composed of government representatives and civil society members, including advocacy groups, human rights specialists, researchers and technical experts. Within the institution's new governance framework, we have taken a position on the independent GIFCT's Operating Board.

We are proud of these partnerships and look forward to continuing our work with other companies, governments, and non-profit and international organizations in combating extremist content online.

- **Are there any platforms Alphabet has observed not to act on threat information your company has shared with them?**

We recognize that disrupting foreign election threats requires a whole-of-society approach. Technology companies, law enforcement, government agencies – including the Intelligence Community – cybersecurity researchers, academics, and media outlets each have unique roles based on their respective threat profiles and areas of coverage. Google recognizes that we have a responsibility to combat malign foreign influence and share relevant information with key stakeholders including affected users and relevant law enforcement agencies. It is incumbent on other technology companies to take action on that information as appropriate.

Question 3: After the 2016 U.S. federal elections, this Committee uncovered evidence that Russian influence campaigns had reached hundreds of millions of Americans on platforms like Facebook and Instagram.

- **What is the scale of foreign adversary influence campaigns on Alphabet platforms in more recent federal elections?**

Please see our detailed response to Question 2 for more information regarding foreign adversary influence campaigns.

- **Please provide estimates on the number of users who interacted (including views) with content the Department of Justice (DOJ) has attributed to Russian influence operations in its disruption effort dated September 4, 2024.**

The YouTube channel for Tenet Media was created in September 2023 and had a lifetime total of more than 16 million views. Additionally, YouTube terminated 14 other channels affiliated with Tenet Media and its founders (see below chart). It is generally unclear from the indictment what portion of the views on these channels could potentially be attributed to RT-sponsored content. For example, the YouTube channel titled "Lauren Chen" predated the creation of Tenet Media by over 6 years, and garnered over 100 million views in its lifetime.

YouTube Channel URL	Title	Total views
youtube.com/channel/UC5GB1qrr914BMfhsCcbBy-g	Lauren Tam	Over 950
youtube.com/channel/UCfjZ81Yqcp2B57KCZu6BUkg	Liam D	0 (no uploads; no views)
youtube.com/channel/UCRiUKeuGjRwJl3NugrpDcUg	Lauren Tam	Over 40
youtube.com/channel/UCbqX2ghkS9MAN0qOEmGzVRQ	Mickey Burgundy	Over 2,000
youtube.com/channel/UCLUrVTVTA3PnUFpYvpfMcpg	Lauren Chen	Over 100M
youtube.com/channel/UC2BLvqmAx3SkWD4UuUqqDOQ	Liam	14
youtube.com/channel/UCKj8Z8olJ6uM9swP1lw98ng	Roaming Media	0 (no uploads; no views)
youtube.com/channel/UCbzlkgBh8dcvTlreNC1hGig	Roaming Media	Over 150k
youtube.com/channel/UChLgdjDEP6j1ARNHhk8lp0Q	Roaming Foodie	0 (no uploads; no views)
youtube.com/channel/UCzF829G7xrc-NlzcmL0P99A	Mediaholic	Over 12M
youtube.com/channel/UCsZXglUmbWTQWQldjjYDsWg	RabbleRouser	Over 95k
youtube.com/channel/UCXAGPDDSJ509Nd6VB1BXsHg	Natural Botanics	0 (no uploads; no views)
youtube.com/channel/UCdmJ9EcVd6wuFU_DHklYZFw	TENET Media	Over 16M
youtube.com/channel/UCfjMzVujPuq8gkv9L-7lW-g	Tayler Hansen	0 (no uploads; no views)

Question 4: The DOJ disruption of Russia's Doppelganger influence network emphasized that Russian influence operatives continue to prize targeted advertising tools.

- **What specific policies has Alphabet adopted to screen ads for both compliance with U.S. sanctions designations and your company's terms of service related to foreign covert influence campaigns?**

We strive to support a healthy digital advertising ecosystem that is trustworthy and transparent and that works for users, advertisers, and publishers. Our ads policies are designed to ensure a safe and positive experience for our users and to comply with applicable laws.

With respect to sanctions, Google has adopted robust policies, procedures, and a screening program to comply with sanctions regulations. Additionally, all advertisers on Google must comply with applicable sanctions and export regulations – including those administered by the Office of Foreign Assets Control ("OFAC") – and agree to not cause Google to violate these regulations. Ads may not be used for or on behalf of restricted entities or individuals or on behalf of entities or individuals located in sanctioned countries or regions. In addition, ads are not available to any entities or individuals that are restricted under applicable trade sanctions and export compliance laws. They are similarly not available to entities or individuals owned or controlled by or acting for or on behalf of such restricted entities or individuals.

We do not permit advertising campaigns that:

- geographically target embargoed countries or territories;
- are run on behalf of businesses located in embargoed countries or regions, even if the account owner is not located in an embargoed country or region; and
- are run by or on behalf of entities or individuals that are restricted under applicable trade sanctions and regulations.

We support responsible political advertising and expects all political ads and destinations to comply with local legal requirements. In the United States, among other jurisdictions, we require all advertisers who wish to run election ads on our platforms to go through a verification process and to have an in-ad disclosure that clearly shows who paid for the ad. These ads are compiled in our Political Ads Transparency Report. In addition, we limit targeting of election ads to geographic location (except radius around a location), age, gender, and contextual options such as ad placements, topics, keywords against sites, apps, pages and videos.

We also prohibit doctored and manipulated media used to deceive, defraud, or mislead others in ads or landing pages making demonstrably false claims that could significantly undermine participation or trust in an electoral or democratic process. Advertisers are now required to disclose election ads that contain synthetic or digitally altered content that inauthentically depicts real or realistic-looking people or events.

Question 5: In the Russian Internet Research Agency's (IRA) influence efforts during the 2016 U.S. federal election it maintained social media accounts that impersonated real political, social, and media organizations.

- **Do we continue to see efforts of foreign adversary influence actors to impersonate legitimate U.S. political, social, and media organizations?**

In advance of the 2024 U.S. elections, we have seen a variety of malicious activity, including cyberattacks, efforts to compromise personal email accounts of high-profile political actors, and influence operations both on and off our platforms seeking to sow discord among Americans. We are seeing some foreign state actors experimenting with generative AI to improve existing cyber attacks tactics by probing for vulnerabilities or creating spear-phishing emails, for instance.

Similarly, we see generative AI being used to more efficiently create fake websites, misleading news articles, and robotic social media posts. We have not yet seen AI bring about a sea change in existing tactics, but we remain alert to new attack vectors. We are always on the lookout for new techniques and procedures that bad actors might deploy to further their goals. As discussed in the course of the hearing, there are concrete threats posed by AI-generated content, including the creation of fake headlines and content and impersonation of legitimate media outlets. While efforts to date have largely involved using GenAI to improve existing methods, we are cognizant that generative AI can lower friction for content creation or translation.

Separately, Google Threat Intelligence continues to track information operation campaigns with various political alignments, including pro-Russia, pro-PRC, and pro-Iran campaigns, which have established influence assets generally pretending to belong to real U.S. individuals or organizations as a means of promoting influence content. Examples of this include inauthentic individual personas that purport to be from the U.S., as well as inauthentic media outlets that inaccurately present as domestic organizations and credible sources of information.

While this general form of impersonation of individual users is most common, some information operations campaigns impersonate known real-world organizations. For example, the Doppelganger campaign has been publicly reported to establish domains that masquerade

as legitimate organizations, including U.S. media outlets, to publish disinformation articles under false attribution. In general, we have seen foreign influence actors move away from impersonating legitimate political, social, and media organizations in the United States in favor of creating non-existent organizations that sound plausible.

- **What policies has Alphabet implemented since 2017 to help users differentiate between authentic and verified organizations versus those that might be impersonating them?**

Please see our response to Question 1 for detailed information regarding our policies prohibiting impersonation across our products.

In addition to our specific impersonation-related policies, we have other tools to help users differentiate between authentic organizations and others. On YouTube, for example, if a YouTube channel is owned by a news publisher that is funded by a government, or publicly funded, an information panel providing publisher context may be displayed on the watch page of the videos on its channel. The information panel provides publisher context, explains how the publisher is funded, and provides a link to the publisher's Wikipedia page. Users will see the information panel providing publisher context directly under the video next to the information icon. Information panels providing publisher context are meant to give users additional information to help them better understand the sources of news content that they watch on the platform. Inclusion of the information panel providing publisher context is based on information about the news publisher made available by Wikipedia and other independent third-party sources.

Additionally, YouTube also awards verification badges to indicate that a specific channel is the official channel of a creator, artist, company, or public figure. Verified channels help distinguish official channels from other channels with similar names on YouTube. YouTube will not verify channels that are trying to impersonate another creator or brand.

On the advertising side, we support responsible political advertising, and expect all political ads and destinations to comply with local legal requirements. In the United States, the European Union, the United Kingdom, and India, among other jurisdictions, we require all advertisers who wish to run election ads on our platforms to go through a verification process and to have an in-ad disclosure that clearly shows who paid for the ad. These ads are compiled in our Political Ads Transparency Report.

In addition, we limit targeting of election ads to geographic location (except radius around a location), age, gender, and contextual options such as ad placements, topics, keywords against sites, apps, pages and videos. Our ads policies also prohibit "deep fakes" (doctored and manipulated media) used to deceive, defraud, or mislead others; misleading claims about the

census process; and ads or landing pages making demonstrably false claims that could significantly undermine participation or trust in an electoral or democratic process. What's further, as discussed below, we believe that users should have information to make informed decisions when viewing election ads that contain synthetic content that has been digitally altered or generated.

Question 6: Volume 2 of the Committee's report on "Russian Active Measures Campaigns and Interference in the 2016 U.S. Election" included a number of recommendations devoted to deterring and defending against technology-enabled foreign influence operations targeting the United States. Consistent with those recommendations:

- **What is the extent to which Alphabet's platforms have improved general information sharing between the public and private sectors? What is the format and frequency of those engagements?**

Google engages with law enforcement agencies to assess threats and to counter attempts to deceive, harm, or take advantage of people in numerous ways. We maintain regular communication channels with law enforcement, government entities, and industry partners as part of our efforts to keep people safe and to understand and adapt to trends and new forms of abuse. We rely on information learned through such channels to ensure the integrity of our products and to act swiftly in response to crises or public safety incidents such as terrorism, mass shootings, violent events, or child sexual exploitation or when we detect abuse that may threaten the integrity of democratic processes.

We have processes in place to proactively refer to law enforcement imminent threats and specified illegal activity occurring on our platform. These matters are raised to us in a variety of ways, and we have established procedures for both internal product teams and external sources to escalate potential threats and criminal activity if they see it on our platforms. Google personnel assess escalated threats and illegal activity, and refer matters to law enforcement as appropriate and consistent with due process and privacy parameters.

Governments and law enforcement entities make legal requests for user information and 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 on this process, please see our policies governing how we handle government requests for user information. We have a track record of pushing back against overly broad or otherwise inappropriate government demands for user data, including objecting to some demands entirely. More information on global requests for user data can be found in the Google Transparency Report.

Additionally, courts and government agencies around the world regularly send us court orders, subpoenas, warrants, and other forms of legal process (collectively "legal demands") for content on YouTube and other services. Some legal demands may allege infringement of intellectual property rights, while others claim violation of local laws prohibiting certain types of user-generated content, such as defamation or Neo-Nazi content in certain member states of the European Union. We review these legal demands closely to determine if content should be removed or restricted because it violates a local law or is contrary to our terms of service and content policies. In order for us to evaluate a request from a government entity, it must be in writing and provide a clear explanation of how the content violates controlling laws or regulations or our policies.

We do not always remove content in response to a government demand. Some legal demands may not be specific enough for us to know what the government wanted us to remove (for example, no URL is listed in the request) or lack sufficient explanation of why the government believes content violates a given law.

Further, outside parties across the political spectrum in the United States and globally – including government agencies, Members of Congress, political entities, non-governmental organizations, academics, and individual users – inform us of content that they believe may violate our terms of service and policies. When content is flagged, our teams voluntarily review it in light of our terms and policies, and we independently evaluate whether the content violates our terms of service and policies, without regard for the source of the original inquiry.

- **How have Alphabet's platforms increased transparency measures by social media companies for users to understand platform activity, such as disclosure of automated accounts, greater contextual information on the source of certain content, and complete and timely public exposure of malign influence operations?**

Please see our answer to Question 1 regarding our public disclosures and notifications of malicious activity. In addition to our TAG reports and Mandiant blogs, we provide extensive transparency across our platforms. Since Google launched the industry's first Transparency Report in 2010, we have been sharing data that sheds light on how the policies and actions of governments and corporations affect privacy, security, and access to information online. The Transparency Report includes detailed information related to user data disclosure, data on content removal requests, and data on a range of issues relating to policies, practices, and access to information.

Beginning in 2018 and continuing quarterly, we published our first quarterly YouTube Community Guidelines enforcement report. The quarterly report publicly shares aggregate data on YouTube's efforts to remove content that violates its Community Guidelines and provides insight into how it uses a combination of machines and human flaggers to enforce our

policies at scale. Included in the Community Guidelines enforcement report are breakdowns of comments and videos removed by removal reason, source of first detection, country, and more.

Further, in the United States, the European Union, the United Kingdom, and India, among other jurisdictions, we require all advertisers who wish to run election ads on our platforms to go through a verification process and to have an in-ad disclosure that clearly shows who paid for the ad. These ads are compiled in our Political Ads Transparency Report.

- **Which agency is Alphabet's primary point of contact within the U.S. government for addressing foreign influence or interference issues relating to the U.S. 2024 federal elections? How often does Alphabet interact with this agency?**

Our primary point of contact within the U.S. government for addressing foreign influence and interference issues relating to the U.S. 2024 federal elections is the Federal Bureau of Investigation. Additionally, throughout this critical year, Google continues to meet with government officials to share information about malign foreign influence, including the Office of the Directorate for National Intelligence, the National Security Agency, the Department of Homeland Security, and the National Security Council. These meetings with national security experts – in which we share information about the threat landscape that we are observing across our platforms and the ways in which we are managing these threats – have taken place for many years, regardless of Administration. We also participate in the Joint Cyber Defense Collaborative and the National Security Agency's Cybersecurity Collaboration Center, through which real-time information regarding threats are shared, including those relating to election security.

- **How would you characterize the frequency and quality of the interactions Alphabet is having with the U.S. government relating to foreign influence or interference issues?**

The priorities outlined in the Committee's report on Russia's activity in 2016 are wholly consistent with a number of our priorities - including promoting principled data-sharing across industry on foreign influence operations and enhancing transparency for the public about what we are seeing on our platforms. As described above, we have a number of touchpoints with officials across the U.S. government, and we believe those touchpoints are effective conduits for information sharing. We have met multiple times with national security officials from across the government. We anticipate continuing a regular cadence of meetings in advance of Election Day. These meetings and the mutual sharing of information has occurred for multiple election cycles, regardless of the Administration in charge.

- **How often is the government providing Alphabet tips of potential malicious activity by foreign actors that is unique – in other words, your internal trust and safety teams were not aware of the issue until the government raised it to your attention?**

Election integrity is a shared challenge. Although we design and enforce our policies independently, we have received information for many years from national security agencies, the Intelligence Community, and federal, state, and local law enforcement, as well as from a range of trusted flaggers that may have access to information and intelligence about malicious activity, including from foreign adversaries, that we do not. For example, often we can only see malicious activity that is occurring on Google platforms. In many cases, illicit activity occurs off our platforms, and the U.S. government may have access to information to which we do not.

Question 7: In August 2022, a former Twitter employee was convicted of acting at the behest of a foreign government – using his access within the company to share information on dissident users and provide sensitive information on the social media platform to the foreign government. Subsequently, a former Twitter whistleblower who served as a senior executive at the company testified to the Senate Judiciary Committee that additional countries had penetrated Twitter with intelligence operatives.

- **What steps does your company take to address insider threat risks, particularly with respect to employees that might have access to sensitive user information or company technology that could enable foreign surveillance or influence campaigns?**

Google is vigilant against insider threats and takes the associated risks extremely seriously. Our dedicated teams are made up of experts focused on security incident response, digital forensics, privacy incident response, vulnerability coordination, cyber threat detection, and insider threat detection. They work 24/7 and across the globe to prevent, address, and mitigate insider threats. We also have a wide range of information and security policies that work together and complement one another to help ensure that the company, its users, its data, and its employees are protected against insider and external threats alike.

We safeguard proprietary technology, information, and trade secrets with physical measures. For example, we secure our physical spaces by deploying campus-wide security guards and installing cameras on most building entry points. Google restricts access to its buildings by requiring employees to badge at entryways and throughout the workspace areas. In addition, certain floors or areas within buildings are further restricted to a subset of employees by badge access. We also require advance registration for guests, and Google employees are required to escort their guests at all times.

Google takes active measures to secure our networks. We have in place a system of data loss prevention that monitors and logs certain data transfers to and from Google's network. We also require each device to be uniquely identified and authenticated before accessing the Google corporate network. In addition, all Google employees must use two-factor authentication for their work-related Google accounts. We log employee activity on Google's network, including file transfers to platforms such as Google Drive or DropBox.

We collect physical and network access information, including badge access times and locations, Internet Protocol (IP) addresses for employee logins, and two-factor authentication logs, and gather this information in a database to analyze potential risks. This data is regularly assessed both by automated tools and human analysts to detect potential malicious activity. Google employees are instructed to report remote work from foreign locations, and Google automatically limits the network access of employees traveling to certain countries, such as China, North Korea, and Iran. Within the Google network, access to certain sensitive information, including our AI technology, is further restricted to a subset of employees whose job duties are related to the subject matter.

Our Privacy & Information Security training program is reviewed and updated annually to reflect current security and privacy practices at Google. All employees and members of the extended workforce must complete security training yearly. When required, teams also provide additional role-specific training. We have implemented escalating communications, access restrictions, and performance rating implications for those who do not complete their Privacy and Security training.

Google employees are responsible for appropriately classifying, protecting, accessing, and sharing data. Googlers and members of the extended workforce who violate data security policies, including by inappropriately disclosing (such as transferring or copying) data outside of the company, are subject to discipline, up to and including termination of employment.

No system is perfect and one of our priorities is to encourage a culture where people feel comfortable reporting anomalies. We provide numerous channels for employees and others to report such anomalies and concerns, including anonymously.

We also have teams of experts around the globe who conduct workplace investigations into security related concerns that may involve Google's people, information, and assets, including data exfiltration and infiltration; information "leaks" of corporate data and intellectual property; and espionage or sabotage, among numerous other topics.

- **Has Alphabet identified efforts by foreign intelligence operatives to infiltrate Alphabet's workforce?**

Both the private sector and U.S. government are engaged in a constant battle to protect our information from malicious foreign threat actors. The threats are not limited to only those individuals who are affiliated with foreign intelligence services. Whether it be trade secrets or national security materials, we recognize that bad-faith actors – regardless of motivation – are determined to seize every advantage they can.

Google recognizes that the risk of workforce infiltration is a threat faced across the industry and within government. For instance, in March 2024, the U.S. Department of Justice announced the indictment of seven PRC-linked hackers who spent approximately 14 years targeting U.S. and foreign critics, businesses, and political officials in furtherance of the PRC's economic espionage and foreign intelligence objectives. As the Department wrote, "[t]he targeted U.S. government officials included individuals working in the White House, at the Departments of Justice, Commerce, Treasury, and State, and U.S. Senators and Representatives of both political parties."

Question 8: In March of this year, the DOJ indicted a Google engineer with stealing AI-related trade secrets from the company, likely to benefit the two Chinese AI-related firms with whom the engineer was associated. The engineer began uploading confidential Google information to his personal accounts no later than May 1, 2022, and yet Google's internal controls did not detect the exfiltration of information until December 2, 2023 – during such period when the engineer traveled to China for five months and participated in investor meetings for one of the Chinese AI firms.

- **What internal controls and processes do you have to detect insider threats, such as the now-indicted Google engineer, to protect sensitive and advanced capabilities that are crucial to your company's success?**

- **Does your company have requirements for employees working on critical technologies to report foreign travel and/or contacts? If not, why?**

Please see our response to Question 7 for detailed information about our efforts to combat insider threats, including our foreign travel requirements.

Question 9: In the last two years, there have been numerous reports about widespread downsizing of trust and safety teams at some of the largest platforms – including at Google and across Meta's social media platforms.

- **How are your internal teams (e.g. trust and safety, election security, etc.) currently resourced to monitor, detect, and disrupt foreign influence and/or interference efforts related to the U.S. 2024 federal elections? Please provide the number of full-time employees directly responsible for election-related trust and**

safety work this year, as well as the number of such employees in the U.S. federal election in 2020.

For many years we have supported numerous elections around the world, applying new learnings with each cycle to both improve our protections from harmful and misleading content and to create trustworthy experiences for our users. We have an extensive elections-focused team, including members of our Intelligence Desk, Trust and Safety, and product teams, monitoring real-time developments and making adjustments to our approach as needed. Supporting elections is a core element of our responsibility to our users, and we will continue throughout this critical year to build on our existing efforts.

Tens of thousands of individuals located around the globe and possessing a diverse set of backgrounds, including an array of linguistic capabilities and cultural expertise, are working around the clock to help enforce our policies and moderate content. We have invested billions of dollars in efforts to keep our platforms and services safe. For example in 2022, more than 40,000 people across the globe helped enforce Google's policies and moderate content. There were no cuts to the Trust and Safety teams that work on elections. We continue to invest aggressively in human and technological resources to increase our trust and safety capacity. And we continue to make significant investments in the people, policies, and systems that enable Google and YouTube to be a reliable source for election-related news and information.

- **Please provide any estimate for the number of full-time employees capable of supporting non-English languages (for example, Spanish, Mandarin, Tagalog) in the U.S.**

As described above, we have made significant investments in the teams and systems that protect Google's users, partners, and business, and we have tens of thousands of people working in a variety of roles to help enforce our policies and moderate content. These individuals are located across the globe and have a diverse set of backgrounds, including an array of linguistic capabilities and varied regional contexts. For example, in the context of YouTube, we recognize that YouTube is a global platform with users in over 100 countries who speak dozens of languages. Our YouTube content moderation workforce totals more than 11,000 as of June 2024, and we have more than 590 moderators around the world who can review YouTube content posted in Spanish.

- **How much, in concrete budgetary terms, has Alphabet devoted to trust and safety measures related to the U.S. federal election in 2024? Please also provide the comparable figure for the U.S. federal election in 2020.**

As we have previously disclosed, in 2020, we spent nearly $1.2 billion on content moderation across the company. In 2022, the most recent year for which we have readily available data, we increased this investment by more than 60 percent. Additionally, we continue to heavily invest in artificial intelligence and machine learning, which will enable us to more effectively and efficiently moderate harmful content at scale.

- **What personnel or capability investments has Alphabet made to ensure generative AI capabilities cannot be exploited by malicious foreign actors? How confident are you that your organizations could detect malicious use of generative AI capabilities for foreign influence operations?**

We are building on the ways in which we help our audiences identify AI-generated content through several new tools and policies.

- **Synthetic Content Ad Disclosures:** We require advertisers to disclose election ads that include synthetic content that inauthentically depicts real or realistic-looking people or events. Recently, we added Google-generated disclosures for YouTube Election Ads for some YouTube formats.
- **Content Labels on YouTube:** We seek to provide YouTube viewers as much context as possible about the content they watch. At time of upload, we require users to disclose content that is meaningfully altered or synthetically generated when it seems realistic. For most videos, a label will appear in the expanded description, but for videos that touch on more sensitive topics — like health, news, elections, or finance — we also show a more prominent label on the video itself.
- **Responsible Approaches to Generative AI Products**: Last December, we announced Gemini apps and web experience would not provide substantive responses to election-related prompts. As we integrate Gen AI into more consumer experiences, we're also applying election-related restrictions to many of these products, including Search AI Overviews, YouTube AI-generated summaries for Live Chat, Gems, and image generation in Gemini. Particularly for federal and state-wide elections, our users depend on us to provide reliable and up-to-date information on topics like current candidates, voting processes and election results — and this new technology can make mistakes as it learns or as news breaks. For many of these queries on Gemini, we also provide a link connecting users directly to Google Search for the latest and most accurate information.
- **Providing Users with Additional Context:** The About this Image feature in Search helps people assess the credibility and context of images found online. Our double-check feature in Gemini evaluates whether there is content across the web to substantiate the responses it provides to user queries.

- **Digital Watermarking:** We're continuing to bring SynthID — embedded watermarking — to additional Google Gen AI tools for content creation and more forms of media including text, audio, visual and video. For instance, images generated by Gemini, including with our most recent Imagen 3 model, are embedded with SynthID watermarks.

Further, we recognize the importance of collaborating across the tech industry to identify emerging challenges and counter abuse. Earlier this year, we were proud to sign on to the Tech Accord to Combat Deceptive Use of AI in 2024 Elections, a set of commitments to deploy technology countering harmful AI-generated content meant to deceive voters. We pledged to help prevent deceptive AI-generated image, audio, or video content from interfering with this year's global elections. As described in greater detail in a recent update on the Tech Accord website, we have taken a number of steps across our products to reduce the risks that intentional, undisclosed, and deceptive AI-generated imagery, audio, or video may pose to the integrity of electoral processes.

In line with our Tech Accord Commitments, we are also continuing our efforts to foster cross-industry resilience, provide transparency to the public, and engage with civil society. We actively share our learnings and expertise with researchers and others in the industry. These efforts include increasing public awareness by, for example, actively publishing and updating our approach to AI, our research into provenance solutions, and our approach to content labeling.

In addition to our Tech Accord commitments, we joined the Coalition for Content Provenance and Authenticity (C2PA) as a steering committee member. The C2PA is a cross-industry effort to help provide more transparency and context regarding AI-generated content. Google has worked alongside the other members to develop and advance the technology used to attach provenance information to content. Through the first half of the year, we collaborated on the newest version (2.1) of the technical standard, Content Credentials. This version is more secure against a wider range of tampering attacks due to stricter technical requirements for validating the history of the content's provenance, which will help ensure the data attached is not altered or misleading. Beginning this week, when creators upload original, unaltered content with C2PA metadata to YouTube, we will notate "Captured with a camera" in the expanded description box.

Question 10: The 2023 book Broken Code by The Wall Street Journal journalist Jeff Horwitz describes Meta's various "Break the Glass" measures that were built to reduce the potential for violence in "At Risk Countries" including the United States before, during, and after the U.S. 2020 federal election. Horwitz writes, "In total, sixty-four separate break-the-glass measures were in place well before the election was called for

Biden on November 7th." His reporting indicates those measures were also disabled prior to January 6, 2021, when they were re-enabled as the U.S. Capitol was stormed. As reported, these "Break the Glass" measures were primarily about enabling or disabling features on the Facebook Blue website/app and included virality circuit breakers and disabling certain group features more than individual pieces of content.

- **Has Alphabet built similar "Break the Glass" mitigation tools and policies to address foreign misuse ahead of, and immediately after, Election Day?**

We recognize that our efforts to combat election misinformation originating from foreign malign actors must extend beyond the removal of misleading content and inauthentic activity from certain features, as described in our responses to answers to Questions 1 and 2. We also have taken a number of steps to combat foreign misinformation and disinformation campaigns and influence operations by elevating authoritative information on a variety of election-related topics across our products.

For example, we build our Google Search ranking systems to surface the highest quality information available on the open web – information that is both relevant and reliable – at the top of our results. For topics where quality information is particularly important, such as health, finance, civic information, and crisis situations, we place an even greater emphasis on factors related to expertise and trustworthiness. We use external search quality raters to evaluate our results and ensure they reflect what people around the world consider to be high quality, and we are transparent about how we define high quality results. When users search for information on candidates, the voter registration process, and where to vote in their states, they will find aggregated resources and information from state election offices, provided by authoritative partners.

On YouTube, for news and information related to the election, YouTube's recommendation system prominently surfaces content from authoritative sources on the YouTube homepage, in search results, and the "Up Next" panel. We also highlight high-quality content from authoritative news sources during key moments, through our Top News and Breaking News shelves, as well as the news watch page. Moreover, we have information panels that indicate funding sources from publishers that receive public or government funding, as well as information panels giving topical context for topics prone to misinformation.

In the U.S., our 2020 election information panels, with relevant context from voting locations to live election results, were collectively shown over four billion times, while during the 2022 midterms our election-related information panels and public service announcements were shown over two billion times. Ahead of 2024, we have triggered our voter registration and vote by mail information panel on videos related to the topics.

This year, we've worked with the Bipartisan Policy Center to make their authoritative information about vote by mail available in Vietnamese and Mandarin, as well as Spanish and English. YouTube also continues to elevate authoritative content in Spanish for users in the United States, from news sources such as Noticias Telemundo, Univision Noticias, and CNN en Español, all of which have over three million subscribers, among others. We will have data about how many times this year's information panels are shown following the upcoming election.

On the Google Play Store, we recently launched a new badge for apps that are from official government agencies. This will help point people to trustworthy information, including for voting. And on Google News, we launched additional News features in 2022 to help readers discover authoritative local and regional news from different states about elections around the U.S.

Earlier this month we announced that, as we've done in the past, we will temporarily pause ads related to U.S. elections after the last polls close on November 5. We're implementing this policy out of an abundance of caution and to limit the potential for confusion, given the likelihood that votes will continue to be counted after Election Day.

On Google Maps, we will clearly highlight polling locations and provide easy to use directions. To prevent bad actors from spamming election-related places on Maps, we will apply enhanced protections for contributed content on places like government office buildings.

To complement these features, Google Translate breaks down language barriers to help people connect and better understand the world around them, and we are always applying the latest technologies to increase access to this tool. In 2022, we added 24 new languages and announced the 1,000 Languages Initiative, a commitment to build AI models that will support the 1,000 most spoken languages around the world. Google Translate can help individuals access high-quality information even in languages in which the features referenced above are not directly available.

Question 11: Hack and leak operations constituted a major component of Russian election influence measures in 2016 and foreign adversaries (including Iran) continue to pursue these operations to damage campaigns and sow division.

- **What are your policies in the event state actors disseminate hacked materials on your services in order to damage a campaign? Will you label such content? Will you remove it?**

As we noted above, Google continues to invest heavily in combating coordinated information operations, which would include efforts by state-backed actors to disseminate hacked

materials. Google takes the protection of personal and confidential information seriously and, to this end, has robust policies across its platforms and services. For example, on YouTube, our Community Guidelines prohibit content that shares, threatens to share, or encourages others to share non-public personally identifiable information (PII). We define PII to include individuals' home addresses; email addresses; sign-in credentials, such as usernames or passwords; phone numbers; passport numbers; medical records; or bank account information. More information about the scope of this policy is publicly available.

On Google Search, we may remove certain personal information that creates significant risks of identity theft, financial fraud, or other specific harms including, but not limited to, doxxing content, explicit personal images, and involuntary fake pornography. On Google Drive, we have Terms of Service that prohibit the distribution of personal or confidential information without authorization. Examples of personal and confidential information include US Social Security numbers, bank account numbers, credit card numbers, images of signatures, and personal health documents.

Further, on Google Ads, we do not permit ads that directly facilitate or advertise access to hacked material related to political entities within scope of Google's election ads policies. This policy applies to all protected material that was obtained through the unauthorized intrusion or access of a computer, computer network, or personal electronic device, even if distributed by a third party. Examples of ads that are prohibited on our platforms include, but are not limited to: advertising access to hacked content ("See all of the leaked emails right now!", "The President's text messages were hacked! Access them now!"); linking to hacked content ("View our database of hacked documents from the President's campaign.", "Foreign agents hacked into his computer, take a look at the real documents.").

- **What are Alphabet's policies in the event domestic users disseminate materials on Alphabet platforms that have been attributed to hack and leak operations by a state actor?**

Please see our response to the prior question regarding our approach to the dissemination of hacked and leaked materials involving personally identifiable information, including in the event domestic users disseminate that type of materials.

- **Is Alphabet aware of any actors seeking to publish or otherwise disseminate hacked (or purportedly hacked) information in the U.S. 2024 federal election? If so, how has Alphabet responded?**

On September 27, the U.S. Department of Justice announced the unsealing of an indictment charging Iranian nationals, and Islamic Revolutionary Guard Corps employees, Masoud Jalili, Seyyed Ali Aghamiri, and Yaser Balaghi with a conspiracy with others known and unknown to

hack into accounts of current and former U.S. officials, members of the media, nongovernmental organizations, and individuals associated with U.S. political campaigns. According to DOJ's press release: "The activity was part of Iran's continuing efforts to stoke discord, erode confidence in the U.S. electoral process, and unlawfully acquire information relating to current and former U.S. officials that could be used to advance the malign activities of the IRGC, including ongoing efforts to avenge the death of Qasem Soleimani, the former commander of the IRGC – Qods Force (IRGC-QF)." In this Departmental press release, Google was the first among four private sector partners recognized for providing assistance with this case.

The indictment also referenced exfiltrated materials, which we assessed against our policies prohibiting the distribution of personally identifiable information. In September, Google restricted the distribution of the non-redacted file "J_D_V.pdf" because it violated Google Drive's Personal and Confidential Information Policy, due to the inclusion of personally identifiable information. The file was no longer permitted to be copied, nor shared with or viewed by other users. File owners continued to retain access, and files in shared drives can still be accessed by the shared drive's organizers.

Question 12: In July 2024, Director of National Intelligence Avril Haines highlighted the Iranian regime's role in provoking anti-Israel and anti-American protests in the U.S. and has previously highlighted Iran's role in attempts to undermine U.S. democratic institutions.

- **What actions has Alphabet taken to address the presence of Iranian influence operations on YouTube since the DNI's announcement?**

As we described in the response to Question 2, Google has had robust efforts to combat ongoing malicious activity affiliated with Islamic Revolutionary Guard Corp and APT42. We identified that the personal Gmail account of a high-profile political consultant in June 2024 had been compromised. In addition to quickly securing the compromised account and sending government-backed attacker warnings to all of the targeted accounts, we proactively referred this malicious activity to law enforcement in early July, and we are continuing to cooperate with them on this matter.

- **How is Alphabet differentiating between accounts of Iranian government actors who do not enjoy the right of free speech and those of American citizens who do?**

As reflected in Google's testimony before the Committee, our business relies on earning the trust of our users. We take seriously the importance of protecting free expression and access to a range of viewpoints, while also maintaining and enforcing responsible policies. We recognize the importance of enabling the people who use our services – in the United States

and abroad – to speak freely about the political issues most important to them. At the same time, we continue to take steps to prevent the misuse of our tools and platforms, particularly by foreign state actors attempting to undermine democratic elections.

Google has policies in place that prohibit the masking of foreign entities for purposes of interfering in the affairs of other countries. We do not want our users to be deceived or misled by content that they encounter on our platform, and we will take enforcement action against advertisers, publishers, developers, and creators who omit relevant information or provide misleading information about their identity, affiliation, expertise, or experience.

To enforce this policy, our teams conduct holistic reviews to assess for factors such as coordinated behavior and cross-border targeting and to determine whether behavior is violative. We review signals such as account information, billing information, and login credentials to help determine identity.

Our teams work hard to ensure we are striking a balance between allowing for a broad range of political speech and making sure our platform is not abused to incite real-world harm, particularly by malign foreign actors located abroad.

- **As anti-Israel and anti-American protests sweep the country, which at times become violent, what are Alphabet's policies to promote public awareness in instances where it has identified Iran's role in fomenting such activity?**

As previously described, TAG tracks and works to disrupt more than 270 government-backed attacker groups from more than 50 countries and publishes its findings each quarter. Mandiant similarly shares its findings on a regular basis, and has published more than 50 blogs to date this year alone analyzing threats from Russia, China, Iran, North Korea, and the criminal underground.

For example, in June and July, we disrupted malicious activity originating from APT42 targeting of high-profile users in Israel and the United States. We quickly secured the compromised account and sent government-backed attacker warnings to all of the targeted accounts and we proactively referred this malicious activity to law enforcement in early July. At the same time, we informed officials from both campaigns that we were seeing heightened malicious activity originating from foreign state actors and underscored the importance of using enhanced account security protections on personal email accounts.

Further, we published a public blog, which received extensive press coverage, that provided extensive details regarding APT42's efforts to target current and former government officials, political campaigns, diplomats, individuals who work at think tanks, as well as NGOs and academic institutions that contribute to foreign policy conversations. We also explained that in

the past six months, the U.S. and Israel accounted for roughly 60% of APT42's known geographic targeting, including the likes of former senior Israeli military officials and individuals affiliated with both U.S. presidential campaigns. These activities demonstrate the group's aggressive, multi-pronged effort to quickly alter its operational focus in support of Iran's political and military priorities.

- **How does Alphabet ensure public notification of these accounts, when they are discovered?**

Please see our response to Question 1 for more information about Google's public notification regarding malicious coordinated activity originating from foreign actors.

Question 13: Since 2017, industry has widely attributed to the People's Republic of China an online influence network dubbed "Spamouflage" promoting PRC narratives and harassing opponents of the PRC government. Recent public reporting has identified examples of the influence network employing inauthentic social media accounts to influence political discourse in advance of the U.S. 2024 federal election.

- **What steps is Alphabet taking to identify these inauthentic profiles and to delete them?**

As described in our answer to Question 2, in 2024, DRAGONBRIDGE, also known as Spamouflage, continues to spread narratives highlighting U.S. political divisions and portraying the U.S. government, society, and democracy in a negative light, cycling through political and social narratives that evolve with the headlines. In May 2024, for example, DRAGONBRIDGE began uploading videos and commenting on the student protests over the Israel-Hamas war on U.S. university campuses. DRAGONBRIDGE content appeared in English, was generally pro-Palestine in its themes, and used the student protests to frame the U.S. and Western media as hypocritical.

This year, we terminated more than 22,000 YouTube channels linked to Chinese coordinated influence operations, as we publicly shared in the first three quarters of 2024. Though it is evident that substantial resources are being expended around pro-PRC operations, these efforts do not appear to be gaining significant traction. When we have observed them spinning up activity across platforms, we have been able to stop them relatively quickly. Google Threat Intelligence is actively monitoring DRAGONBRIDGE activity for any shifts in tone or focus related to the U.S. presidential election.

- **What is Alphabet doing about content originally published by those accounts, but then shared and amplified by real people? Will that content be removed from Alphabet platforms as well?**

As described above, we have terminated more than 22,000 YouTube channels linked to Chinese coordinated influence operations, as we publicly shared in the first three quarters of 2024. With respect to enforcement, we remove content that violates our Community Guidelines. We rely on our automated flagging system as well as on YouTube community members who report or flag content that they find inappropriate. The vast majority of videos we removed – more than 96 percent in Q2 of 2024 – are first detected by our automated flagging system. The rest are first flagged by users and organizations.

YouTube strives to remove content that violates our Community Guidelines before users are ever exposed to it. 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. We started tracking VVR in 2017, and we share the rate in our quarterly Community Guidelines Enforcement Report. Through investments in hiring and technology, we have worked to reduce the VVR. In Q2 2024, YouTube's VVR was 0.09-0.11%, meaning that out of every 1,000 views on YouTube, only about one consists of content that violates any of our Community Guidelines. More information on YouTube content removals can be found in our Transparency Report, which is updated on a quarterly basis.

- **Who is responsible at Alphabet for determining authentic from inauthentic accounts? What does the process look like?**

We rely on a combination of people and technology to enforce our policies prohibiting inauthentic activity and other content violations as described in prior responses. Google has invested significantly in our automated detection systems, and our engineering teams continuously evaluate their efficacy and make improvements. Machine learning is well-suited to detect patterns and identify content that is similar – but not identical – to other content and activity we have already removed. Our machine learning systems help our human review teams, including TAG, to identify and remove content at scale, with the speed and volume that could not be achieved with people alone.

- **How is Alphabet engaging with the U.S. government, including the Intelligence Community and law enforcement, to share or exchange information on these types of operations when they may affect candidates, campaigns, or races?**

Please see our response to Question 6 for more information about information sharing with U.S. government agencies, the Intelligence Community, and law enforcement.

From Vice Chairman Rubio

Question 1: In October 2023, days after the Hamas attacks on Israel, X (Twitter) took down hundreds of Hamas-linked accounts on its platform. Has Alphabet taken similar action with Hamas-linked accounts on YouTube?

YouTube has a network of robust Community Guidelines that work together to combat violent and extremist material. We remain committed to enforcing our Community Guidelines in connection with the conflict in the Middle East, and our teams are working around the clock to monitor for harmful content across languages and locales. We will take action quickly when needed across videos, Shorts, and livestreams.

More specifically, YouTube prohibits content intended to praise, promote, or aid violent extremist, criminal, or terrorist organizations. We prohibit content that encourages others to carry out acts of violence. We do not permit terrorist organizations to use YouTube for any purpose, including recruitment, as well as content that promotes terrorism, glorifies terrorist acts, or incites violence. 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.

YouTube does not permit content that violates our policies against violent extremism, including material produced by organizations designated by the U.S. government as "foreign terrorist organizations," including Hamas. We do not permit these terrorist organizations to use YouTube for any purpose. These policies predate the horrific events of October 7, 2023, and continue to be in effect. Additionally, content produced by violent extremist groups that are not government-listed foreign terrorist organizations is subject to our robust policies, including those described above prohibiting the glorification of terrorist acts.

Since the terrorist attack by Hamas in Israel and the beginning of the escalated conflict now underway in Israel and Gaza, we have removed over 115,000 videos, terminated over 5,500 channels, and removed over 200 million comments from YouTube for violating our policies.

As this conflict continues, we remain committed to enforcing our Community Guidelines, which set out what isn't allowed on the platform. Our teams continue to work around the clock to monitor for hate speech and other harmful content, including content pertaining to violent extremism, graphic violence, harassment and misinformation, and content originating from designated terrorist organizations, such as a Foreign Terrorist Organization in the United States or other organizations identified by the United Nations.

In addition to the removal of violative content, during major global events, such as the conflict in Israel and Gaza, our systems prioritize connecting viewers with high-quality news and information from authoritative sources. Our recommendation system is prominently surfacing news from authoritative sources on the homepage, in search results, and on the "Up Next" panel. We implement this system-wide across all countries in which we operate. Our Top News and Breaking News shelves are surfacing at the top of search results related to the attacks in Israel and on the homepage, prominently featuring content from authoritative news sources.

Question 2: What is Alphabet's internal process for delineating free speech from nefarious activities of U.S. adversaries including the Iranian regime?

As reflected in Google's testimony before the Committee, our business relies on earning the trust of our users. We take seriously the importance of protecting free expression and access to a range of viewpoints, while also maintaining and enforcing responsible policies. We recognize the importance of enabling the people who use our services – in the United States and abroad – to speak freely about the political issues most important to them. At the same time, we continue to take steps to prevent the misuse of our tools and platforms, particularly by foreign state actors attempting to undermine democratic elections.

Google has policies in place that prohibit the masking of foreign entities for purposes of interfering in the affairs of other countries. We do not want our users to be deceived or misled by content that they encounter on our platform, and we will take enforcement action against advertisers, publishers, developers, and creators who omit relevant information or provide misleading information about their identity, affiliation, expertise, or experience.

To enforce this policy, our teams conduct holistic reviews to assess for factors such as coordinated behavior and cross-border targeting and to determine whether behavior is violative. We review signals such as account information, billing information, and login credentials to help determine identity.

Our teams work hard to ensure we are striking a balance between allowing for a broad range of political speech and making sure our platform is not abused to incite real-world harm, particularly by malign foreign actors located abroad.

Our teams work hard to ensure we are striking a balance between allowing for a broad range of political speech and making sure our platform is not abused to incite or create real-world harm, particularly from malign foreign actors located abroad. We welcome ongoing discussions on this matter and we will continue our continuous efforts to protect the integrity of elections around the world. We take the integrity of the democratic process incredibly seriously, and we will remain vigilant as elections around the globe unfold.

Question 3: Since the Director of National Intelligence's public announcement in July, have that office or other Intelligence Community agencies reached out to Alphabet to identify the accounts of Iranian government actors for Alphabet to take action?

Throughout this election year, Google continues to meet with government officials to share information about malign foreign influence operations. Among other entities, we have met with the Federal Bureau of Investigation, the Office of the Directorate for National Intelligence, the National Security Agency, the Department of Homeland Security, and the National Security Council. These meetings with national security experts – in which we share information about the threat landscape that we are observing across our platforms and the ways in which we are managing these threats – have taken place for many years, regardless of Administration.

There are a number of ways Google proactively works with the Intelligence Community and U.S. law enforcement agencies to assess threats and to counter attempts to deceive, harm, or take advantage of people. We maintain regular communication channels with law enforcement, the Intelligence Community, other government entities, and industry partners as part of our efforts to keep people safe and understand and adapt to trends and new forms of abuse. We rely on information learned through such channels to ensure the integrity of our products and act swiftly in response to crises or when we detect abuse that may threaten public safety or the integrity of democratic processes, such as terrorism, mass shootings, violent events, child sexual exploitation, and other incidents.

We receive information from the Intelligence Community, and we value information sharing from across the U.S. government regarding foreign threats, including those coming from the Iranian Revolutionary National Guard and APT42. In the past six months, roughly 60 percent of APT42's known attacks have been directed against U.S. and Israeli targets, including former senior Israeli military officials and individuals affiliated with both U.S. presidential campaigns.

During the 2020 U.S. presidential election cycle, we disrupted APT42 attempts to target accounts associated with the Biden and Trump presidential campaigns. These activities reflect the group's aggressive, multi-pronged effort to quickly alter its operational focus in support of Iran's shifting political and military priorities.

In the current U.S. presidential election cycle, TAG has detected and disrupted a small but steady cadence of APT42's Cluster C credential phishing activity. In May and June of 2024, APT42 targets included the personal email accounts of approximately a dozen individuals variously affiliated with President Biden or former President Trump, including current and former officials in the U.S. government and individuals associated with the two campaigns. We blocked numerous APT42 attempts to log in to personal email accounts of targeted individuals.

We identified that the personal Gmail account of a high-profile political consultant in June 2024 had been compromised. In addition to quickly securing the compromised account and sending government-backed attacker warnings to all of the targeted accounts, we proactively referred this malicious activity to law enforcement in early July, and we are continuing to cooperate with them on this matter.

As noted above, on September 27, 2024, the U.S. Department of Justice announced the unsealing of an indictment charging Iranian nationals, and Islamic Revolutionary Guard Corps employees, Masoud Jalili, Seyyed Ali Aghamiri, and Yaser Balaghi with a conspiracy with others known and unknown to hack into accounts of current and former U.S. officials, members of the media, nongovernmental organizations, and individuals associated with U.S. political campaigns. According to DOJ's press release: "The activity was part of Iran's continuing efforts to stoke discord, erode confidence in the U.S. electoral process, and unlawfully acquire information relating to current and former U.S. officials that could be used to advance the malign activities of the IRGC, including ongoing efforts to avenge the death of Qasem Soleimani, the former commander of the IRGC – Qods Force (IRGC-QF)." In this Departmental press release, Google was the first among four private sector partners who provided assistance with this case.

APT42's efforts to target the U.S. presidential election are of course not limited to Google products. As documented in recent public reporting, the group has successfully breached accounts across multiple email providers and we believe this activity is ongoing. TAG has notified other service providers of this malicious activity so that they can take appropriate action on their platforms. We will continue to monitor developments and share findings with industry peers as we uncover additional activity.

1 Hacker Way
Menlo Park, CA 94025
United States

October 29, 2024

Chairman Mark Warner
Vice Chairman Marco Rubio
US Senate Select Committee on Intelligence
211 Hart Senate Office Building
Washington DC, 20510

Dear Chairman Warner, Vice Chairman Rubio, and Members of the Committee:

Thank you for the questions for the record from the Senate Select Committee on Intelligence Hearing entitled "Foreign Threats to Elections in 2024 – Roles and Responsibilities of U.S. Tech Providers" on September 18, 2024. Attached are Meta's answers to the questions posed.

Sincerely,
Meta Platforms, Inc.

Questions from Chairman Warner and Vice Chairman Rubio

Question 1. **In the previous federal elections, we have witnessed influence efforts by Iran, Russia, and the People's Republic of China (PRC) to stoke social and political divisions – up to, and including in the case of Iran and Russia, seeking to provoke violence and derision in the U.S. via social media platforms.**

a. **What is the extent to which Meta's platforms are observing foreign adversaries seeking to incite violence among Americans?**
b. **What are your company's policies towards such activity?**
c. **How are your platforms disseminating cautionary information to users and the general public?**

We know that foreign adversaries try to reach people on our platforms and others before elections, and we remain vigilant in our fight against their evolving tactics. We are constantly working to find and stop coordinated campaigns that seek to manipulate public debate across our platforms. We do not want organizations or individuals creating networks of accounts that mislead people about who they are or what they are doing.

Our Community Standards and Community Guidelines prohibit coordinated inauthentic behavior (CIB), which we define as coordinated efforts to manipulate public debate for a strategic goal, in which fake accounts are central to the operation. In each case, people coordinate with one another and use fake accounts to mislead others about who they are and what they are doing. In an effort to prevent and disrupt real-world harm, we also prohibit other conduct, including Violence and Incitement, Coordinating Harm and Promoting Crime, and Bullying and Harassment. And, under our Dangerous Organizations and Individuals policy, we do not allow organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on our platforms. We assess these entities based on their behavior both online and offline, particularly their ties to violence. In February, we removed the Facebook and Instagram accounts of Iran's supreme leader Ayatollah Ali Khameini for repeatedly violating this policy.

When it comes to countering foreign interference, we know that it is an adversarial space. We conduct our own independent investigations to identify what is—and is not—foreign interference. When we investigate and remove CIB operations, we focus on behavior rather than content—in this sense, it does not matter who is behind them or what they post. We have removed over 200 networks of CIB since 2017, including networks from Russia, Iran, and China.

Despite our efforts, people continue to look for new ways to mislead people, which is why we continue to take steps to make it harder for them to do so and to constantly improve our detection and enforcement systems. When we find and remove CIB, we identify the tactics used and we build tools into our platforms to make those tactics more difficult at scale. We also continue to monitor and assess new risks, including those associated with evolving new technologies like artificial intelligence (AI). Our findings so far suggest that generative AI-powered tactics provide only incremental productivity and content-generation gains to the threat actors and have not impeded our ability to disrupt their influence operations. We continue to assess that our industry's defense strategies, including our focus on behavior

(rather than content) in countering adversarial threat activity, already apply and appear effective at this time.

As we approach the 2024 elections, our security efforts include:

- Ongoing threat research into and enforcement against new and known threats/threat actors;
- Sharing threat indicators and insights publicly so the public can strengthen its responses to foreign interference and other adversarial threats we find;
- Sharing our threat research with our industry peers, researchers, policymakers and the public in our regular adversarial threat reports;
- Continuous monitoring for, and enforcing against, efforts to come back by networks we previously removed for CIB, cyber espionage and other policy violations;
- Refining our automated detection systems to help scale the work of our security expert investigators allowing them to focus on the most complicated threats;
- Alerting people who we believe were targeted by spyware or cyber espionage activity so that they can take steps to secure their accounts.

Adversarial threats are not unique to Meta. Our security work shows that these threats rarely—if ever—target a single platform. Cross-industry collaboration, transparency, and reporting are essential to preventing and discouraging bad actors from engaging in harmful conduct across the internet. That is why we publicize our CIB takedowns for all to see, provide information about them to third parties for further research, and share relevant information with researchers, academics, and others, including Congress. We publish regular adversarial threat reports, which detail the results of our efforts to combat CIB, as well as other adversarial threats we detect and remove from our platforms. Our public threat reporting began over seven years ago when we first shared our findings about CIB by a Russian covert influence operation linked to the Internet Research Agency. Since then, we have expanded our ability to respond to a wider range of adversarial behaviors as global threats have continued to evolve. To provide a more comprehensive view into the risks we tackle, we expanded our threat reports to include insights into other threats, as part of our quarterly reporting. In addition, we also publish threat indicators to contribute to the security community's efforts to detect and counter malicious activity across the internet. Today, we have compiled the largest repository of threat indicators, with more than 6,000 threat indicators related to the cross-internet activity by Doppelganger—the most persistent Russia-based covert influence operation—alone. We also report on our integrity enforcement progress publicly in our quarterly Community Standards Enforcement Report. This report includes metrics on how Meta is performing in preventing and removing content that violates our Community Standards and fake accounts.

Since 2020, we have taken steps to empower users to make informed choices about the media they consume by labeling media outlets in certain countries that are wholly or partially under the editorial control of their government as state-controlled media. Transparency is a key part of our strategy to help people better understand the sources of news content they see on our platforms so they can make informed decisions about what they are reading. For example, we widely label Russian state-controlled media entities (SCMEs) across Facebook and Instagram. On Facebook, these labels appear globally on numerous surfaces, including on posts in Feed, on Pages, on ads, and in the Ad Library Page view. On Instagram, these labels similarly appear in multiple areas, including in the "About this Account" section,

on posts, on stories and Reels, and on profiles. For more information on our approach to labeling SCMEs, please see our response to Question 12.

Question 2. **In 2016, we observed Russian influence actors push content across a wide range of platforms – big and small.**

a. **Based on what your threat intelligence groups are tracking, what is the scope of the current foreign adversary influence campaigns?**
b. **What level of interaction and information sharing does your company have with smaller platforms?**
c. **Are there any platforms Meta has observed not to act on threat information your company has shared with them?**

As noted in response to Question 1, Doppelganger remains the most persistent Russia-based covert influence operation we have observed since 2017, targeting many apps at once and focused primarily on weakening support for Ukraine and its government. As detailed in our most recent Adversarial Threat Report, Doppelganger continues to add new domains to its large network of websites and attempts to promote them across the internet. Our teams work daily to find and block Doppelganger's attempts to acquire new accounts and Pages, run ads, and share links to its websites and redirect domains, before these are ever shared on our apps. To date, we have blocked over six thousand deceptive domains operated by Doppelganger from being shared on our platforms, in addition to sharing them publicly and with our industry peers and researchers. Of note, many of these web domains and sites continue to persist on the broader internet to this day and post new content as part of the broader covert influence campaign. Our goal is to keep driving up the operational cost of these campaigns, making them less and less effective.

Some recent trends that stood out to us in Russian campaigns include the use of for-hire campaigns operated by contractors (rather than security agencies themselves, as we saw in the past). These campaigns continue to run low-quality, high-volume efforts, making errors including in their operational security. In fact, we continue to see real people calling these networks out as trolls, as they struggle to engage authentic audiences. We have also observed increased persistence in recent campaigns compared to past Russia-based operations. In response to detection, these bad actors create new assets over and over again, without much effort put into building audiences on social media. They do, however, appear to put extensive efforts into operating their websites, likely in an attempt to preserve their content against ongoing disruptions by social media platforms. Tackling this aspect of foreign interference requires a whole-of-society approach to engaging the domain name and web registration and hosting ecosystem to investigate and disrupt the web infrastructure powering these deceptive campaigns. Without a concerted effort to disrupt the internet infrastructure powering these campaigns, we expect these website-centric operations to persist as long as their customers task them to do so, regardless of their efficacy.

As also described above in response to Question 1, cross-industry collaboration, transparency, and reporting are essential to preventing and discouraging bad actors from engaging in harmful conduct across the internet. We know that transparency across the industry helps us all respond to new threats, because our security work shows that these threats rarely—if ever—target a single platform. And we have seen threat actors migrate to more permissive or less responsible platforms. That is why we share information with industry peers wherever possible to help them better protect their platforms and to help raise our

collective defenses. This includes both direct sharing, as well as via a monthly cross-industry election security forum. The members of the monthly forum include LinkedIn, Microsoft, Google, Wikimedia, Medium, and X (formerly Twitter).

We have also increasingly relied on transparency to share key threat indicators from our takedowns, both to enable smaller platforms to protect their users and to help the public see what we are taking down and why. In 2017, we started publishing detailed reporting on our work to detect and counter security threats on our platforms, known today as our Adversarial Threat Reports. We also publicly release threat indicators we identify on our GitHub platform. Today, we have compiled the largest repository of threat indicators from CIB networks we have removed, including more than 6,000 threat indicators related to the cross-internet activity by Doppelganger. And we provide information about our takedowns to third parties for their further research, and share relevant information with researchers, academics, and others, including Congress.

For more information on our efforts to combat CIB, please see our response to Question 1.

Question 3. **After the 2016 U.S. federal elections, this Committee uncovered evidence that Russian influence campaigns had reached hundreds of millions of Americans on platforms like Facebook and Instagram.**

a. **What is the scale of foreign adversary influence campaigns on Meta platforms in more recent federal elections?**
b. **Please provide estimates on the number of users who interacted (including views) with content the Department of Justice (DOJ) has attributed to Russian influence operations in its disruption effort dated September 4, 2024.**

For more information on foreign adversary influence campaigns, please see our response to Questions 1 and 2. As noted above, we have disrupted more than 200 networks of CIB since 2017.

Persistence is common among influence operations, but Doppelganger has taken its efforts to a new level, while remaining crude and largely ineffective in building authentic audiences on social media. We are constrained in assessing the volume of views that posts containing a link to a Doppelganger spoofed website may have received both due to our data retention protocols and because Doppelganger's website geoblocking and redirection tradecraft prevents Meta from associating Doppelganger activity on our services to specific spoofed websites. For these reasons, we similarly do not have access to the number of outbound clicks these Doppelganger spoofed websites may have received.

Further, because of our daily monitoring, detection, and blocking—and our efforts exposing Doppelganger's attempts to target our platforms since we first took action against this threat actor in 2022—Doppelganger has largely ceased to engage in linking to spoofed websites or seeding links to drive traffic off-platform since May 2024. Historically, links that Doppelganger has tried to post on Meta's services do not link directly to a spoofed website, but rather redirect people to the spoofed website through one or multiple hops or other temporary websites, which Doppelganger takes down after use; this technique inhibits us from collecting impression metrics associated with specific spoofed websites. In some cases, Doppelganger uses redirection to land users on websites that are not spoofed and are not

under Doppelganger's control. As a result, Meta is unable to determine the frequency with which these redirector domains ultimately lead users to a spoofed site or a real site.

Doppelganger also continues to use geofencing to make its sites accessible only to internet users from particular countries. The operators use multiple random, unrelated urls to redirect people from a particular target country to geofenced spoof websites, while showing a nonsensical web page to everyone else. This prevents our tools from capturing information in a reliable way that would allow us to produce the data you are seeking. Logs at the website host for the spoofed sites are the only reliable place to understand how much traffic each site has gained. Alternatively, we note that certain entities have publicly reported on the relatively limited nature of Doppelganger viewership.

We know that adversarial threats rarely target just one platform, and threat actors routinely exploit the global domain name system to deceive people into visiting imposter news sites, clicking on phishing links, installing malware, and falling for other scams. While we regularly block and publicize these malicious campaigns—both by Doppelganger and by other campaigns using similar tactics—they continue to persist across the broader internet. Over the years, we have seen first-hand how transparency and information sharing can be a force multiplier that enables follow-on threat research and disruptions, raising the cost of running these operations while making them less and less effective across the board. That is why we continue to add Doppelganger spoofed websites and redirect domains we identify to our public GitHub repository to enable broader detection and research across the internet. Still, there are challenges in how the wider system for redressing these harms works, limiting what any individual company can do. Industry-wide action is needed to protect people against these tactics and raise our collective defenses.

Question 4. **The DOJ disruption of Russia's Doppelganger influence network emphasized that Russian influence operatives continue to prize targeted advertising tools provided by your platforms – and confirmed use of Meta targeted ad tools, specifically.**

a. **Seven years after this Committee first identified hundreds of thousands of dollars in Meta ads purchased by the Russian Internet Research Agency (IRA), how are Russian influence actors still successful in using your targeted ad services like this?**

Russia remains the number one source of global CIB networks we have disrupted since 2017. Russian actors continue to evolve their tactics in an attempt to avoid enforcement. Further, in addition to cycling through fake accounts, which get detected quickly and removed, Russian actors have expanded their use of compromised accounts and Pages. At times, they create a new Page using these accounts, while other times they take over a compromised Page. Even with these techniques, we are still able to find and disrupt such activity with regularity, as with our consistent disruption of Russia's Doppelganger network.

As described above, Doppelganger is the most persistent Russia-based campaign we have seen since 2017, targeting many apps at once and focused primarily on weakening support for Ukraine and its government. Our teams actively work to find and block Doppelganger's efforts, including their efforts to run ads. We catch the majority of Doppelganger ads before they run or within hours of submission. We continue to incorporate our latest insights into our detection systems. We know Doppelganger is actively testing ways to avoid detection, including the tactics described above, but such circumvention attempts

are significantly degrading the quality of their ads, making them barely legible and less relevant to the public discourse on topics Doppelganger has been pursuing since 2022. In fact, we have seen people on our platforms comment on these ads, publicly calling them out as Russian trolls, propaganda, and bots.

More broadly, in March 2022, we paused ads targeting people in Russia and prohibited advertisers within Russia from creating or running ads anywhere in the world, including within Russia.

b. What specific policies has Meta adopted to screen ads for both compliance with U.S. sanctions designations and your company's terms of service related to foreign covert influence campaigns?

We prohibit deceptive behavior on our platforms, including by advertisers. All advertisers running ads across Meta technologies must follow our Community Standards and our Advertising Standards. In addition, advertisers on Instagram must also follow our Instagram Community Guidelines. Advertisers are responsible for understanding and complying with all applicable laws and regulations. Failure to comply may result in a variety of consequences, including the cancellation of the ads placed and termination of an advertiser's account.

We use automated and, in some instances, manual review to enforce our policies. We have several layers of analysis and detection, both before and after an ad goes live. Our ad review system reviews ads for violations of our policies. This review process may include the specific components of an ad, such as images, video, text and targeting information, as well as an ad's associated landing page or other destinations, among other information. Our ad review process starts automatically before ads begin running and is typically completed within 24 hours; any ad may be subject to re-review. Beyond reviewing individual ads, we also monitor and investigate advertiser behavior and may restrict advertiser accounts that do not follow our Advertising Standards, Community Standards, or other Meta policies and terms. If a violation is found at any point in the review process, the ad will be rejected, and the advertiser account or its assets may be restricted or disabled. As described above, this includes disabling Pages and accounts trying to run ads that we determine are CIB.

Meta also is committed to complying with U.S. sanctions administered and enforced by the U.S. Department of the Treasury's Office of Foreign Assets Control ("OFAC") and the U.S. Department of State and continuously takes steps to meet its legal obligations. Meta implements a number of controls on its ad services designed to monitor for and take action against potentially sanctioned parties. Meta conducts screening against various sanctions and restricted party lists ("Watchlist Screening"), including but not limited to OFAC's Specially Designated Nationals and Blocked Persons List. Watchlist Screening includes trigger-based screening at specific points throughout the lifecycle of Meta's ad services, as well as re-screening when sanctions authorities like OFAC make changes to sanctions lists. Meta also implements controls to mitigate the risk of engaging in unauthorized activities with persons in comprehensively sanctioned jurisdictions, including geoblocking controls and country-based screening.

Question 5. **In the Russian IRA's influence efforts during the 2016 U.S. federal elections it maintained social media accounts that impersonated real political, social, and media organizations.**

a. **Do you continue to see efforts of foreign adversary influence actors to impersonate legitimate U.S. political, social, and media organizations?**
b. **What policies has Meta implemented since 2017 to help users differentiate between authentic and verified organizations versus those that might be impersonating them?**
c. **What is the extent to which Meta is facilitating third-party research entities (such as by academics and civil society organizations) to assist platforms in identifying manipulation by foreign actors?**

We believe that authenticity helps create a community where people are accountable to each other, and to Meta, in meaningful ways. We want to allow for the many ways that identity is expressed across our global community, while preventing impersonation and identity misrepresentation. To maintain a safe and open environment where people can trust one another and build community, we do not allow for the creation of accounts or profiles that are created or used to deceive others. This includes prohibiting accounts, Pages, and groups that impersonate another person or entity by using their image(s), name, or likeness with the aim to deceive others or speaking in the voice of another person or entity for whom the user is not authorized to do so (e.g., by creating a Page or profile). We also seek to prevent abusive tactics, such as spreading deceptive links to draw unsuspecting users in through misleading functionality or code, or impersonating a trusted domain.

We constantly work to find and stop coordinated campaigns that seek to manipulate public debate across our platforms, including impersonation efforts. As discussed above in response to Question 2 and as detailed in our most recent Adversarial Threat Report, Doppelganger, a cross-internet influence operation from Russia, shifted tactics on our platform in response to aggressive enforcement, its latest attempts at evading detection. With over six thousand deceptive domains operated by Doppelganger blocked from being shared on our apps, the operators continue to look for ways around detection by us and our industry peers. We have seen them begin spoofing the websites of primarily non-political and entertainment news outlets and online magazines. They also resumed attempting to seed links to spoofed news and government websites on our apps, including redirects. However, between May 2024 and August 2024, we added nearly 300 threat indicators to our industry's largest repository of more than 6,000 indicators related to this threat actor. And between May 2024 and August 2024, we detected and removed over 5,000 accounts and Pages.

In a separate effort, we also removed 12 Facebook accounts, 32 Pages, five Groups, and three accounts on Instagram associated with another Russian influence campaign. This network primarily targeted Ukraine, and to a much lesser extent Poland, the broader European Union, and the United States. The people behind this operation relied on fake accounts—some of which were detected and removed prior to our investigation—to manage Pages posing as Ukrainian organizations, to impersonate public figures in the West including some pro-Russia commentators, and to post content. Some of these accounts used profile photos that were likely created by generative adversarial networks, and this operation used proxy IP addresses to create the appearance that they were based in the regions targeted.

In addition to removing accounts and networks that violate our policies, one way we help identify authentic accounts is to allow account holders and Page owners to apply to receive a verified badge—a blue check that appears next to an Instagram username or Facebook account profile name. A verified badge indicates that we have confirmed that the Facebook Page or profile or Instagram account is the

authentic presence of the individual, public figure, or brand it represents. We consider a number of factors when evaluating Facebook Pages and profiles, and accounts on Instagram, to determine if they are in the public interest and meet our verification criteria, which are consistent across Facebook and Instagram. To be verified, organizations must provide documentation to validate the request, such as a certificate of formation, business license, utility bill, or tax exemption document. Account holders can also share information about their audience, the region where they are most popular, and add up to five news articles to help our teams with additional context when reviewing the applications. If a Page, profile, or account fails to meet the criteria for a verified badge, there are other ways to let people know it is authentic. For example, they can link to it from an official website, Instagram profile, Facebook Page, or X (formerly Twitter) account.

Further, since June 2020, we have applied labels to media outlets in certain countries that are state-controlled media entities (SCMEs), which we define as being "wholly or partially under the editorial control of their government." We wanted to provide greater transparency into these publishers because they combine the influence of a media organization with the strategic backing of a state, and we believe people should know if the news they read is coming from a publication that may be under the influence of a government. For more information on our approach to labeling SCMEs, please see our response to Question 12.

Finally, as described above, we publicize our takedowns of CIB for all to see, provide information about them to third parties for their review, and share relevant information with researchers and others. Our quarterly threat reports provide a comprehensive view into the risks we see across multiple policy violations including CIB, cyber espionage, and other emerging harms. We continue to also share information about our CIB takedowns with researchers, academics, and others, including Congress. For example, independent researchers can access information on networks we take down through our Influence Operations Research Archive, hosted in Meta Content Library. To date, researchers have produced over 100 independent reports on covert influence operations we have removed from our platforms using information we have shared. We also publicly release threat indicators we identify on the Meta Threat Research Repository on GitHub. This includes the largest repository of threat indicators, with more than 6,000, related to cross-internet activity by Doppelganger alone. Finally, through Meta Ad Library, we provide a comprehensive, searchable database for ad transparency. Anyone can use the Ad Library to get more information about the ads currently on Meta technologies, and we provide an API that enables programmatic access to ad information for social issue, election, and political ads, as well as ads run in the European Union, allowing researchers to conduct deeper analysis.

Researchers have helped build on our deep investigations into Facebook activity to identify the cross-internet nature of the operation. For example, the researchers at the Atlantic Council's Digital Forensic Research Lab provided insights into a part of Doppelganger during our original investigation, and we shared our findings with them to enable further research into the broader operation. Graphika used information from our takedowns to buttress their analysis of China-origin influence operation Spamouflage Dragon and Russia-origin operation Secondary Infektion. As influence operations move off of our platforms and increasingly rely on cross-platform or offline activity, we believe that civil society organizations are sometimes best positioned to identify nascent influence activity before the deceptive

activity manifests on our platforms. When we receive these leads, we conduct an independent investigation and only enforce if we independently identify evidence of policy violations.

Question 6. **Volume 2 of the Committee's report on "Russian Active Measures Campaigns and Interference in the 2016 U.S. Election" included a number of recommendations devoted to deterring and defending against technology-enabled foreign influence operations targeting the United States. Consistent with those recommendations:**

- **a. What is the extent to which Meta's platforms have improved general information sharing between the public and private sectors?**
  - **i. What is the format and frequency of those engagements?**
- **b. How have Meta's platforms increased transparency measures by social media companies for users to understand platform activity, such as disclosure of automated accounts, greater contextual information on the source of certain content, and complete and timely public exposure of malign influence operations?**
- **c. Which agency is Meta's primary point of contact within the U.S. government for addressing foreign influence or interference issues relating to the U.S. 2024 federal elections? How often does Meta interact with this agency?**
- **d. How would you characterize the frequency and quality of the interactions Meta is having with the U.S. government relating to foreign influence or interference issues?**
- **e. How often is the government providing Meta tips of potential malicious activity by foreign actors that is unique – in other words, your internal trust and safety teams were not aware of the issue until the government raised it to your attention?**

As noted in the responses above, collaboration, transparency, and reporting among public and private stakeholders are essential to preventing and discouraging bad actors from engaging in harmful conduct across the internet. We know that this collaboration helps us all respond to new and evolving threats, because our security work shows that these threats rarely—if ever—target one single platform. External insights from counterparts in government, as well as researchers and investigative journalists, can be particularly important in detecting and disrupting sophisticated threat actors who coordinate their operations outside of our platforms. We recognize the need to be cautious and in each case, conduct our own investigation to identify what is—and is not—foreign interference. That is why when we receive these leads, we conduct an independent investigation and only enforce if we independently identify evidence of policy violations.

Sharing information between tech companies, governments, and law enforcement has proven critical to identifying and disrupting foreign interference early, ahead of elections. As an example, prior to the 2020 elections, we investigated and took down three covert influence operations from Russia, Mexico, and Iran targeting the United States, after receiving a tip from US law enforcement about off-platform activity by these threat actors. We recently investigated and removed a network that originated in Russia and targeted primarily Ukraine, and to a much lesser extent Poland, the broader European Union, and the United States. This network appeared to be focused on two main topics: promoting Russian integration in Francophone Africa and diminishing support for Ukraine in the West. We began our independent investigation after receiving a tip from the FBI about a small portion of the network's activity, which led our teams to find the broader network. We believe it is important that we continue to build on the progress

the defender community has made since 2016, and make sure we work together to keep our defenses against foreign interference strong.

Transparency is a key part of our strategy to help people better understand the content they see on our platforms so they can make informed decisions about what they are reading, including related to CIB. Over seven years ago, we began our public threat reporting by sharing our findings about CIB by a Russian covert influence operation. Since then, global threats have significantly evolved, and we have expanded our ability to respond to a wider range of adversarial behaviors. Our quarterly threat reports provide a comprehensive view into the risks we see across multiple policy violations including CIB, cyber espionage, and other emerging harms.

We continue to also share information about our CIB takedowns with researchers, academics, and others, including Congress. For example, independent researchers can access information on networks we take down through our Influence Operations Research Archive, hosted in Meta Content Library. We also publicly release threat indicators we identify on the Meta Threat Research Repository on GitHub. This includes the largest repository of threat indicators, with more than 6,000, related to cross-internet activity by Doppelganger. And we report on our integrity enforcement progress publicly in our quarterly Community Standards Enforcement Report. This report includes metrics on how Meta is performing in preventing and removing content that violates our Community Standards and fake accounts.

Question 7. **In August 2022, a former Twitter employee was convicted of acting at the behest of a foreign government – using his access within the company to share information on dissident users and provide sensitive information on the social media platform to the foreign government. Subsequently, a former Twitter whistleblower who served as a senior executive at the company testified to the Senate Judiciary Committee that additional countries had penetrated Twitter with intelligence operatives.**

a. **What steps does your company take to address insider threat risks, particularly with respect to employees that might have access to sensitive user information or company technology that could enable foreign surveillance or influence campaigns?**
b. **Has Meta identified efforts by foreign intelligence operatives to infiltrate Meta's workforce?**

Meta has always had zero tolerance for abuse of user data. We have policies and procedures aimed at prohibiting unauthorized access to data. Not only have we fired employees found to have improperly accessed data, we also continue to strengthen employee training, abuse detection, and prevention protocols, while also continuing to reduce the need for employees to access some types of data as they work to support our services, as described further below.

Meta's Employee Policies and Training Around Access to User Data

Adherence to responsible data access practices is a key part of our employment policies and code of conduct. Meta employees must comply with the User Data Access Policy ("UDAP"), which governs employee use of and access to user data. The UDAP applies to all Meta companies, including affiliates and subsidiaries, and binds full-time employees and contingent workers. The policy clearly states: "you are expected to access user data only as strictly necessary for the performance of your job responsibilities

at Meta" and "never use your access to internal tools for non-business purposes." The policy expressly states that "Meta has zero-tolerance for violations of this policy," and confirmed non-compliance with UDAP results in termination of employment.

Similarly, Meta's Code of Conduct provides that employees may only "collect, create, access and use only the minimum amount of data we need to support clearly stated purposes," and says that Meta has "zero-tolerance for inappropriate access to user data and people data." It further mandates that employees "comply with all applicable data privacy laws and legal requirements," such as those governing the "collection, access, and use of data."

As part of onboarding, new employees receive privacy and security training that provides an overview of the UDAP, as well as Meta's privacy obligations, policies, and employees' obligations. The training instructs new employees on Meta's privacy program, privacy obligations during the design and development of new products, and best practices for information security. Engineers receive additional product design training in (i) privacy and ethics, which covers proper data handling, and (ii) security and integrity, which addresses security considerations in product development. Employees also must complete an annual privacy training, including passing required tests, which reiterates Meta has a zero tolerance policy towards inappropriate access or sharing of user data.

Meta's Procedures to Safeguard User Data

We have rigorous administrative, physical, and technical controls in place to restrict employee access to user data. As explained above, we work to scope access based on job requirements. If issues arise, we have a zero tolerance approach to abuse, and improper behavior results in termination.

To help ensure compliance with access policies, Meta has monitoring and logging for employee access to user data. This monitoring triggers alerts to our security teams. The incident is triaged and, if after an investigation Meta determines that an employee accessed user data in violation of the UDAP, the employee will be terminated.

Question 8. **In March of this year, the DOJ indicted a Google engineer with stealing AI-related trade secrets from the company, likely to benefit the two Chinese AI-related firms with whom the engineer was associated. The engineer began uploading confidential Google information to his personal accounts no later than May 1, 2022, and yet Google's internal controls did not detect the exfiltration of information until December 2, 2023 – during such period when the engineer traveled to China for five months and participated in investor meetings for one of the Chinese AI firms. What internal controls and processes do you have to detect insider threats, such as the now-indicted Google engineer, to protect sensitive and advanced capabilities that are crucial to your company's success?**

a. **Does your company have requirements for employees working on critical technologies to report foreign travel and/or contacts? If not, why?**

As discussed in response to Question 7, we maintain controls and processes to prevent, detect, and respond to potential data misuse.

As it relates to foreign travel, we prohibit business travel to certain sanctioned or trade-sensitive countries. We also prohibit employees from bringing work-issued devices when on personal travel to such countries. If an employee attempts to log in to Meta's internal systems from such locations, Meta has security controls that are intended to restrict or prevent access to Meta's internal systems from such locations.

Question 9. **In the last two years, there have been numerous reports about widespread downsizing of trust and safety teams at some of the largest platforms – including at Google and across Meta's social media platforms.**

a. **How are your internal teams (e.g. trust and safety, election security, etc.) currently resourced to monitor, detect, and disrupt foreign influence and/or interference efforts related to the U.S. 2024 federal elections? Please provide the number of full-time employees directly responsible for election-related trust and safety work this year, as well as the number of such employees in the U.S. federal election in 2020.**

Since the 2016 election, we have significantly expanded the number of people who work on safety and security, including people who work specifically on election integrity issues. By 2018, Meta doubled the number of people who work on safety issues from 10,000 to 20,000, which includes content reviewers, systems engineers, and security experts. By the 2020 election, Meta built a global team of 35,000 people to work on safety and security. Today, we have around 40,000 people globally working on overall safety and security. We have also invested more than $20 billion in teams and technology in this area since 2016.

b. **How much, in concrete budgetary terms, has Meta devoted to trust and safety measures related to the U.S. federal election in 2024? Please also provide the comparable figure for the U.S. federal election in 2020.**

Please see our response to 9(a).

c. **What personnel or capability investments has Meta made to ensure generative AI capabilities cannot be exploited by malicious foreign actors? How confident are you that your organizations could detect malicious use of generative AI capabilities for foreign influence operations?**

Meta has been a pioneer in AI development for more than a decade, using machine learning to proactively identify and remove violating content across our services. As with election security, we know that AI progress and responsibility can and must go hand in hand. Generative AI tools offer huge opportunities, and we believe that it is possible and necessary for these technologies to be developed in a transparent and accountable way, while also working to minimize potential risks. We have expanded the number of people that work on AI safety as we have launched new products. We also continue to invest in our safety capabilities.

As detailed in a recent Adversarial Threat Report, we have not seen attempts on our apps to use new generative AI tactics to subvert elections in ways that we could not address through our existing

safeguards, specifically by disrupting adversarial networks behind them. However, this does not mean that people are not using AI to try to interfere in elections. To the contrary, adversaries have used different tools, such as AI-generated photos for profile photos on fake accounts, or AI to publish a large volume of fake articles resembling reputable news sources. We recently disrupted a campaign from Russia that was publishing a large volume of stories on fictitious news websites outside of our apps, which our investigation found were likely AI-generated summaries of original news articles. The same campaign also created fictitious journalist personas with generative adversarial network-created profile photos.

Our teams found and removed many of these campaigns early, before they were able to build audiences and communities on our services. This shows that our industry's existing defenses already apply to novel generative AI and are proving effective thus far. However, we know that we must continue to monitor and assess risks with new technology. That is why we are continually adapting to address new challenges, including by advancing efforts to detect and label AI-generated media. We believe that providing transparency and additional context is the best way to address AI-generated content.

The challenges posed by AI, particularly AI-driven manipulated media, are not unique to Meta and will require a whole-of-industry approach. That is why we have collaborated with global experts with technical, policy, media, legal, civic, and academic backgrounds to inform our policy development and improve the science of detecting manipulated media. And we have been working with others in our industry to develop common standards for identifying AI-generated content through forums like the Partnership on AI (PAI) and the Coalition for Content Provenance and Authenticity.

Additionally, we, along with twenty other companies in the industry, have pledged to help prevent deceptive AI content from interfering with this year's global elections. The "Tech Accord to Combat Deceptive Use of AI in 2024 Elections" is a set of commitments to deploy technology countering harmful AI-generated content meant to deceive voters. Signatories, including Meta, pledge to work collaboratively on tools to detect and address online distribution of such AI content, drive educational campaigns, and provide transparency, among other concrete steps. Detecting these signals will make it possible for us to label certain AI-generated images that people generate or modify with AI off of our platforms and post publicly to Facebook and Instagram. We were also pleased to join the White House's voluntary commitments alongside others in the industry, including a pledge to develop robust technical mechanisms to identify AI-generated content, such as digital watermarking, with respect to frontier models. These commitments are an important first step in ensuring responsible guardrails are established for AI.

We believe that our current approach represents the cutting edge of what is technically possible right now; however, we continue to pursue a range of options to improve our AI detection capabilities. This work is especially important as this is likely to become an increasingly adversarial space in years to come. People and organizations that actively want to deceive people with AI-generated content will look for ways around the safeguards that are put in place to detect it. Across our industry and society more generally, we will need to keep looking for ways to stay one step ahead.

Question 10. **The 2023 book Broken Code by The Wall Street Journal journalist Jeff Horwitz describes Meta's various "Break the Glass" measures that were built to reduce the potential for violence in "At Risk Countries" including the United States before, during, and after the U.S. 2020**

federal election. Horwitz writes, "In total, sixty-four separate break-the-glass measures were in place well before the election was called for Biden on November 7th." His reporting indicates those measures were also disabled prior to January 6, 2021, when they were reenabled as the U.S. Capitol was stormed. As reported, these "Break the Glass" measures were primarily about enabling or disabling features on the Facebook Blue website/app and included virality circuit breakers and disabling certain group features more than individual pieces of content. As the U.S. 2024 federal election approaches, and with the risk of political violence now a reality, these measures seem more critical than ever before.

a. **Is Meta working on "Break the Glass" or similar mitigation features for the U.S. 2024 federal election? Have these measures been built into Facebook, as well as Instagram, Threads, and WhatsApp?**
b. **Has Meta activated any of those "Break the Glass" measures in other countries that had elections this year?**
c. **What are the kind of circumstances that might cause Meta to deploy even more significant "Break the Glass" measures?**
d. **Will Meta share any more information about these "Break the Glass" measures with the Committee or the public?**

We have deployed temporary product measures that can be used to address specific risks in certain instances. Our teams closely monitor trends on our platforms and investigate situations to determine whether and how best to respond. As appropriate, we may apply limited, proportionate, and time-bound measures that can be quickly implemented to address a specific, emerging risk. We also evaluate features in our apps to see how they could be misused during certain events, and in some cases temporarily change those features based on the risks we see.

We monitor real-world events and track different metrics on our platforms. For example, during the 2020 election, we took steps to respond to specific signals we were seeing on the platform, such as increases in reported content. We turned some of the temporary measures off responsibly and gradually as those signals returned to their previous levels. We also left many of the measures in place through Inauguration Day and beyond. For example, we have permanently removed civic groups from recommendations. We have not activated any temporary product measures in other countries that had elections in 2024.

We do not implement temporary product measures lightly—we know these measures could have unintended consequences, like inadvertently limiting benign or otherwise non-violating content. As such, we have sought to make these measures time-bound, proportionate, and consistent with established human rights guidance on permissible limitations for freedom of expression.

We publicly share information about temporary product measures in our Transparency Center. In addition, when these measures are used, we may also discuss such measures in posts on Meta's Newsroom. We are happy to engage with the Committee further on this topic.

Question 11. **Hack and leak operations constituted a major component of Russian election influence measures in 2016 and foreign adversaries (including Iran) continue to pursue these operations to damage campaigns and sow division.**

a. **What are your policies in the event state actors disseminate hacked materials on your services in order to damage a campaign? Will you label such content? Will you remove it?**
b. **What are Meta's policies in the event domestic users disseminate materials on Meta platforms that have been attributed to hack and leak operations by a state actor?**
c. **Is Meta aware of any actors seeking to publish or otherwise disseminate hacked (or purportedly hacked) information in the U.S. 2024 federal election? If so, how has Meta responded?**

As described above, we prohibit coordinated efforts to manipulate public debate for a strategic goal in which fake accounts are central to the operation, and we remove such networks from our platforms. Pursuant to our Community Standards, we do not allow content that shares or asks for private information, either on our services or through external links. This includes content claimed by the poster or confirmed to come from a hacked source, regardless of whether the affected person is a public figure or a private individual. We recognize that private information may become publicly available through news coverage, court filings, press releases, or other sources. When that happens, we may allow the information to be posted. Additionally, if we determine that hacked materials are being posted by or at the direction of a foreign government influence operation, we remove that material.

Posing as "hacktivists" or civically engaged personas to spread hacked or fictitious leaks is a practice we have regularly observed in influence operations. We expect that this tactic will remain a potent tool to manipulate public debate—either through releasing hacked materials wholesale, claiming to possess them to sow uncertainty and force people to prove a negative in the absence of evidence, or publishing distorted documents while claiming their authenticity. This can be particularly challenging to counter in the time-pressured context of election news reporting.

As an example, in 2023, we disrupted an Iranian CIB network that ran a series of fake "hacktivist" personas and offered to publish allegedly hacked documents in certain countries. We also took down a for-hire network run from the United States and Venezuela that targeted Honduras with a fake hacktivist persona called "HondurasLeaks." As detailed in our most recent Adversarial Threat Report, a number of recent Russian operations engaged likely witting and unwitting people to create content and amplify their campaigns, including in Armenia and Europe. These operations often also target journalists and public figures to get them to pick up these narratives and give them credence. This could include seeding hacked or forged materials with unwitting opinion-makers and politicians.

Generative AI-created multimedia claiming to be hacked materials can further add to this challenge. While we have not seen evidence of this technology being used by known covert influence operations to make hack-and-leak claims, we all need to remain vigilant to monitor how generative AI might enable this centuries-old tactic of forging evidence to advance one's strategic goals. That is why we encourage influential figures and the public at large to remain vigilant to avoid playing into the hands of deceptive operations attempting to manipulate public debate. In addition, it is important for political campaigns, candidates, public figures, and media outlets to keep their information security up to date because they represent attractive targets for hackers. As part of our effort to help strengthen account security among these high-target groups, we have run a series of in-app reminders directing people to our security and safety features.

Additionally, safety enhancements like Advanced Protection on Facebook offer security tools and additional protections for candidates and their campaigns, as well as local officials. Through this program, we help accounts on Facebook that may face additional threats during an election cycle adopt stronger account security protections, like two-factor authentication. The program also provides additional security protections for people's Facebook accounts and Pages, including monitoring for potential hacking threats. This allows us to more quickly detect potentially suspicious account activity by monitoring for attempts to hack the account, such as unusual login locations or unverified devices. Should candidates or election officials have concerns about the misuse of our apps or the appropriateness of labeling or handling of their communications, they may always contact us directly via email or via our Meta Support Pros.

Question 12. **Ahead of the U.S. 2020 federal election, Meta promised to label state-controlled media on Facebook and Instagram, as well as in Meta's Ad Library. However, research from the Center for Countering Digital Hate in 2022 found that 91% of posts containing content from Russian state media about Ukraine was not covered by this policy and did not display with any labels.**

a. **What steps is Meta taking to ensure that it is properly labeling most or all content from foreign state-backed accounts?**

We know that social media plays an important role in connecting people with each other and with information they might not have encountered otherwise. While this enables people to discover content that brings them value and to build communities across the globe, it also means they may not have the context they need to engage with such content critically. As we have previously detailed publicly, we have developed a transparency-first approach to balance the integrity risks posed by state media outlets against the risks of over-enforcement. Beginning in June 2020, we started applying labels to media outlets in certain countries that are state-controlled, which we define as being "wholly or partially under the editorial control of their government." We wanted to provide greater transparency into these publishers because they combine the influence of a media organization with the strategic backing of a state. We believe that providing people with knowledge about whether a publication is under the influence of a government better equips them to make informed decisions about the news they consume.

We developed our approach to SCME labeling alongside more than 65 experts around the world specializing in media, governance, and human rights and development. The input we received from these organizations was crucial to understanding the different ways and degrees to which governments exert editorial control over media entities. We know that governments continue to use funding mechanisms to control media, but this alone does not tell the full story. That is why our definition of state-controlled media extends beyond just assessing financial control or ownership and includes an assessment of editorial control exerted by a government.

We look at several factors that may indicate editorial control by a government. For example, we rely on evidence of structural mechanisms through which governments can exercise influence. This includes evidence of direct ties to government, such as whether an outlet is owned by state institutions or state-run companies or whether their governing bodies consist of people appointed by government authorities. Notably, it also includes evidence of indirect ties, such as whether newsroom leadership share affiliations with the government—there is an increasing trend of regimes outsourcing ownership to "puppet owners"

to obscure the intensity of their influence. We also look for evidence of governance mechanisms by which outlets can preserve their independence. Additionally, we consider country-specific factors, including press freedom, and consult open-source research conducted by academics and leading experts. If we determine that there are enough protections in place to ensure editorial independence, we will not apply the label. If we designate an entity as state-controlled, we manually identify and label the SCME's Facebook Pages and Instagram accounts and restrict their ability to advertise in the United States.

We expanded our SCME labeling program in 2022 following the Russian invasion of Ukraine to impose additional restrictions on Russian content. In addition to labeling hundreds of Russian SCME Pages and accounts, in multiple languages, we take enforcement actions globally on all of them in a multitude of ways. We prohibit ads from Russian SCMEs and have demonetized their accounts, and we refused an order from the Russian authorities to stop the independent fact-checking and labeling of content posted on Facebook by four Russian SCMEs. And in September of this year, we expanded our ongoing enforcement against Russian state media outlets. Rossiya Segodnya, RT, and other related entities are now banned from our apps globally for foreign interference activity.

In addition, we have globally demoted content from Facebook Pages and Instagram accounts from Russian SCME outlets, making them harder to find across our platforms. For example, we downrank posts from Russian SCMEs in Feed, and we do not recommend posts from Russian SCME accounts in Explore and Reels. We have also made these accounts harder to find in Search, and we demote posts that contain links to Russian SCME websites on Facebook and Instagram. We label these links and provide more information to people before they share or click on them to let them know that they lead to Russian SCME websites. On Instagram, Stories pointing to a Russian SCME website are placed lower in the Stories tray. We also label these Stories to let people know that they lead to a Russian SCME, including their Spanish language outlets. To provide even more transparency into these outlets, we launched "nudges" that prompt users to confirm whether they want to share or navigate to off-platform content from these outlets. Importantly, these measures enable users who express specific and clear intent to find these outlets and view their content where it can be fact-checked and viewed alongside counter-speech. If people still choose to reshare these posts to their Stories, we will place those Stories lower in the tray. By providing this additional transparency, we aim to give people more context if they want to share direct links to Russian SCME websites or when others see someone's post that contains a link to one of these sites.

Our approach to SCMEs is actor-based, not content-based, meaning that all content from labeled Russian SCMEs is demoted. We have also taken additional steps to enforce our Community Standards and Community Guidelines, not only in Ukraine and Russia but also in other countries globally where content may be shared.

While Russian-origin attempts at covert activity (CIB) related to Russia's war in Ukraine have increased, overt efforts by Russian state-controlled media have decreased as a result of these measures. Recent research by Graphika shows posting volumes on Russian SCME Pages went down 55 percent and engagement levels were down 94 percent compared to pre-war levels, while "more than half of all Russian state media assets had stopped posting altogether." We have seen Russian SCMEs shifting to

other platforms and using new domains to try to escape the additional transparency on (and demotions against) links to their websites.

Question 13. **In July 2024, Director of National Intelligence Avril Haines highlighted the Iranian regime's role in provoking anti-Israel and anti-American protests in the U.S. and has previously highlighted Iran's role in attempts to undermine U.S. democratic institutions.**

- **a. What actions has Meta taken to address the presence of Iranian influence operations on social media since the DNI's announcement?**
- **b. How is Meta differentiating between accounts of Iranian government actors who do not enjoy the right of free speech and those of American citizens who do?**
- **c. As anti-Israel and anti-American protests sweep the country, which at times become violent, what are Meta's policies to promote public awareness in instances where it has identified Iran's role in fomenting such activity?**
- **d. How does Meta ensure public notification of these accounts, when they are discovered?**

Iran is the second most frequent source of foreign interference; we have disrupted 30 global CIB networks since 2017. As one example, in Q1 2024, we removed a network of nearly 50 Iranian-linked Facebook and Instagram accounts, as well as Pages and Groups, for violating our policy against CIB. This network, which targeted Israel, included several separate clusters of activity and used fake accounts to create fictitious personas posing as Israelis in Israel and abroad, manage Groups and Pages, and post content. This operation had presence across the internet, including on Telegram, YouTube, X (formerly Twitter), and TikTok, likely to backstop its fictitious personas so they appeared more legitimate and could withstand scrutiny. The individuals behind this activity posted primarily in Hebrew about news and current events in Israel, including criticism of Hamas and supportive commentary about Israel. We found and removed these clusters before they were able to gain a following among real users.

Most recently, we published information about our investigation into APT42 (also known as UNC788 and Mint Sandstorm), an Iranian threat actor known for its persistent phishing campaigns across the internet targeting political and diplomatic officials, and other public figures (including some associated with the administrations of President Biden and former President Trump). After investigating user reports, our security teams blocked a small cluster of WhatsApp accounts posing as support agents for tech companies. Some of the people targeted by APT42 reported these suspicious messages to WhatsApp using our in-app reporting tools. Those reported messages enabled us to investigate this latest campaign and link it to the same hacking group responsible for similar attempts aimed at political, military, diplomatic and other officials, as reported by our industry peers at Microsoft and Google. We have not seen evidence that their accounts were compromised, but encouraged those who reported to us to take steps to ensure their online accounts are safe across the internet. Out of an abundance of caution, and given the heightened threat environment ahead of the US election, we also shared information about this malicious activity with law enforcement and with the presidential campaigns to encourage them to stay cautious against potential adversarial targeting.

As described in our earlier responses, we conduct independent investigations to identify what is—and is not—foreign interference. When we investigate and remove these operations, we focus on behavior rather than content—no matter what they post. Additionally, we continue to monitor information coming from

our industry peers, our own investigations, and user reports, and we will take action if we identify attempts by malicious actors to target people on our apps.

People who want to interfere in elections rarely target a single service or platform. Cross-industry collaboration, transparency, and reporting are essential to preventing and discouraging these networks from engaging in harmful conduct across the internet. That is why we publicize our takedowns of CIB for all to see, provide information about them to third parties for their review, and share relevant information with researchers, academics, and others, including the Congress. In 2017, we started publishing detailed reporting on our work to detect and counter security threats on our platforms, known today as our Adversarial Threat Reports. We also publicly release threat indicators we identify on our GitHub platform.

Question 14. **Since 2017, industry has widely attributed to the People's Republic of China an online influence network dubbed "Spamouflage" promoting PRC narratives and harassing opponents of the PRC government. Recent public reporting has identified examples of the influence network employing inauthentic social media accounts to influence political discourse in advance of the U.S. 2024 federal election.**

a. **What steps is Meta taking to identify these inauthentic profiles and to delete them?**
b. **What is Meta doing about content originally published by those accounts, but then shared and amplified by real people? Will that content by removed from Meta platforms as well?**
c. **Who is responsible at Meta for determining authentic from inauthentic accounts? What does the process look like?**
d. **How is Meta engaging with the U.S. government, including the Intelligence Community and law enforcement, to share or exchange information on these types of operations when they may affect candidates, campaigns, or races?**

Spamouflage, originating in China, is one of an increasing number of CIB networks which widely spread their assets and infrastructure across many internet surfaces, rather than centralizing their activity and coordination in one place. Spamouflage has been seen operating on more than 50 platforms and forums, including Facebook, Instagram, X (formerly Twitter), YouTube, TikTok, Reddit, Pinterest, Medium, Blogspot, LiveJournal, VKontakte, Vimeo, and dozens of smaller platforms and forums. This campaign was run by geographically dispersed operators across China who appeared to be centrally provisioned with internet access and content. It included positive commentary about China and its province Xinjiang and criticisms of the United States, Western foreign policies, and critics of the Chinese government including journalists and researchers. Our investigation found links to individuals associated with Chinese law enforcement. Because this is a prolific and persistent, though ineffective, influence operation, we continue to look for, work with other threat researchers, and block Spamouflage's attempts to come back online. As an example, in 2023, we took down a cluster of accounts linked to Spamouflage that was reported by researchers at Graphika to have used AI-generated newsreaders in their videos on social media platforms including Facebook, X (formerly Twitter), and YouTube. These early attempts at using AI-generated videos were quickly identified and exposed.

As described in our earlier responses, we are constantly working to find and stop coordinated campaigns that seek to manipulate public debate across our platforms. Our Community Standards prohibit CIB; we

do not want organizations or individuals creating networks of accounts that mislead people about who they are or what they are doing. When we take down these accounts, it is because our investigation has identified deceptive behavior (like using networks of fake accounts to conceal their identity); it is not based on the identity of those behind the account or what they say. Still, people continue to look for new ways to mislead people, which is why we continue to take steps to make it harder for them to do so.

We also constantly work to stop the spread of misinformation. Even if an account does not violate our policies and is not removed, it is still subject to fact checking. We have built the largest independent fact-checking network of any platform, with nearly 100 partners around the world to review and rate viral misinformation in more than 60 languages. Stories they rate as false are shown lower in Feed. If Pages repeatedly create or share misinformation, we significantly reduce their distribution and remove their advertising rights. We make these efforts regardless of the viewpoint of the content or its author. Additionally, we continue to monitor information coming from our industry peers, our own investigations and user reports, and will take action if we identify attempts by malicious actors to target people on our apps.

We will continue to publicize our takedowns of CIB; share relevant information with researchers, academics, and others, including the Congress; and publicly release threat indicators we identify on our GitHub platform. For more information about our coordination with government and law enforcement entities, please see our response to Question 6.

Questions from Vice Chairman Rubio

Question 1. **In October 2023, days after the Hamas attacks on Israel, X (Twitter) took down hundreds of Hamas-linked accounts on its platform. Have Meta's platforms taken similar action?**

Meta has long considered Hamas to be a terrorist organization, and the group is banned from our platforms, as well as other terrorist organizations such as the Popular Front for the Liberation of Palestine (PFLP) and the Palestinian Islamic Jihad. We remove accounts that represent these groups when we identify them. Under our Dangerous Organizations and Individuals policy, we also remove glorification, substantive support, and representation of them when we become aware of it. As an example, in 2021, our teams detected and took down a cluster of activity linked to a covert influence operation we attributed to Hamas. After the October 7, 2023 attack, these fake accounts attempted to re-establish their presence on our platforms. We continue to stay vigilant and take action against Hamas and other terrorist organizations.

In the wake of the attack on October 7, as conflict-related content surged on our platforms, we also implemented a number of temporary measures across both Arabic and Hebrew markets to help limit the prevalence of violating content on our platforms. We quickly established a special operations center staffed with experts, including fluent Hebrew and Arabic speakers, to closely monitor and respond to this rapidly evolving situation in real time. This allows us to remove content that violates our Community Standards or Community Guidelines faster.

In the nine days following October 7, we removed or marked as disturbing more than 2,200,000 pieces of content in Hebrew and Arabic for violating our policies around DOI, violent and graphic content, hate speech, violence and incitement, bullying and harassment, and coordinating harm. As compared to the two months prior, in the three days following October 7, we removed seven times as many pieces of content on a daily basis for violating our DOI policy in Hebrew and Arabic alone. In the majority of cases, we remove the content before people even see it.

Question 2. **What is Meta's internal process for delineating free speech from nefarious activities of U.S. adversaries including the Iranian regime?**

At Meta, we are committed to free expression. Each day, more than three billion people around the world use our apps to express themselves and make their voices heard. We want people to be able to talk openly about the issues that matter to them, whether through written comments, photos, music, or other creative media.

At the same time, we do not want people to misrepresent themselves on our services, use fake accounts, artificially boost the popularity of content, or engage in behaviors designed to enable other violations under our Community Standards. That is why we prohibit inauthentic behavior—a variety of forms of deception, performed by a network of inauthentic assets controlled by the same individual or individuals, with the goal of deceiving Meta or our community or to evade enforcement under the Community Standards.

To determine what constitutes inauthentic behavior, we rely on expert investigative teams and look at the deceptive behavior (like using networks of fake accounts to conceal identity), not what they say. This is an adversarial space, and we always conduct our own independent investigations to identify what is—and is not—foreign interference.

We have removed over 200 networks of CIB since 2017, including networks from Russia, Iran, and China. Still, people continue to look for new ways to mislead people, which is why we continue to take steps to make it harder for them to do so.

Additionally, we label state-controlled media on Facebook, Instagram and Threads so that users know when content is from a publication that may be wholly or partially under the editorial control of a government.

Question 3. **Since the Director of National Intelligence's public announcement in July, have the Office of the Director of National Intelligence or other Intelligence Community agencies reached out to Meta to identify accounts of Iranian government actors for Meta to take action?**

In August, we announced that after investigating user reports, our security teams blocked a small cluster of WhatsApp accounts posing as support agents for tech companies. Our investigation linked it to APT42 (also known as UNC788 and Mint Sandstorm), an Iranian threat actor known for its persistent adversarial campaigns using basic phishing tactics across the internet to steal credentials to people's online accounts. This malicious activity originated in Iran and attempted to target individuals in Israel, Palestine, Iran, the United States and the UK. This effort appeared to have focused on political and diplomatic officials, and other public figures, including some associated with administrations of President Biden and former President Trump.

Some of the people targeted by APT42 reported these suspicious messages to WhatsApp using our in-app reporting tools. Those reported messages enabled us to investigate this latest campaign and link it to the same hacking group responsible for similar attempts aimed at political, military, diplomatic and other officials, as reported by our industry peers at Microsoft and Google.

The vigilance of these users to report the messages to us suggests that these efforts were unsuccessful. We have not seen evidence that their accounts were compromised. We have encouraged those who reported to us to take steps to ensure their online accounts are safe across the internet. Out of an abundance of caution and given the heightened threat environment ahead of the US election, we also shared information about this malicious activity with law enforcement and with the presidential campaigns to encourage them to stay cautious against potential adversarial targeting.

Historically, Iran has been the second most frequent country of origin of CIB networks we have taken down. While we have seen fewer novel Iranian-origin operations recently, we continued to detect and enforce against attempts by previous CIB networks to re-establish operations. Our work against Iranian foreign interference campaigns since 2017 has also enabled us to keep refining our understanding of their tactics and attribution.

We continue to monitor information related to Iran, and will take action if we identify further attempts by malicious actors to target people on our apps. We strongly encourage public figures, journalists, political candidates and campaigns to remain vigilant, take advantage of privacy and security settings, avoid engaging with messages from people they do not know and report suspicious activity to us.

Question 4. **In February, Meta announced a policy change to address the use of AI-generated content, in response to news that an AI-generated video of President Biden was allowed to remain online. Specifically, going forward Meta will now label AI-generated content as such with watermarks and disclaimers.**

a. **In the case of AI-generated content created by Chinese state organizations, will Meta follow this policy instead of taking down these accounts?**
b. **Will Meta consider labelling content as "Chinese government created" or "Russian government created"?**

Earlier this year, we announced changes to the way we handle manipulated media on our platforms and labeling AI-generated content. We have begun labeling content when we detect industry standard AI image indicators or when people disclose that they are uploading AI-generated content. If we determine that digitally created or altered image, video, or audio content creates a particularly high risk of materially deceiving the public on a matter of importance, we may add a more prominent label. This overall approach gives people more information about the content so they can better assess it and so they will have context if they see the same content elsewhere. We will keep this content on our platforms so we can add informational labels and context, unless the content otherwise violates our policies.

In addition to our approach to labeling AI content, as described in our earlier responses, our Community Standards and Guidelines prohibit CIB. When we find campaigns that include groups of accounts and Pages seeking to mislead people about who they are and what they are doing while relying on fake accounts, we remove both inauthentic and authentic accounts, Pages and Groups directly involved in this activity, regardless of whether the content is AI-generated.

The entire industry is still working to determine how to use and share signals, such as metadata or invisible watermarks, or otherwise automatically detect whether photorealistic content that people share is AI-generated. We look forward to continuing to work with stakeholders, including those in Congress, on these issues.