ROUGHLY EDITED TEXT FILE

Center for Democracy and Technology
Future of Speech Online:
AI, Elections and Speech

September 16, 2024
2:45 – 5:15 pm ET

REMOTE CART CAPTIONING PROVIDED BY:
Mary Birnbaum, Voicewriter

* * * * *

>> ALEXANDRA GIVENS: Hey everybody. It is so nice to see you here. Welcome to the eighth annual Future of Speech Online hosted by the Center for Democracy & Technology and Stand Together Trust.

My name is Alexandra Givens, and I am the President & CEO of the Center for Democracy & Technology, a 30-year old nonprofit, nonpartisan organization that works to protect users' civil rights, civil liberties and democracy in the digital age.

I am delighted to welcome you to this incredible event where every year we tackle a major topic related to free expression online. Last year, we focused on generative AI and its implications. This year we are focusing on how online speech impacts elections, in ways that are good and healthy for democracy – and some that are less so. With more than 60 elections happening worldwide, and our own general election in the United States just 50 days away, over 2 billion people will head to the polls this year. The stakes are tremendously high. Digital technologies have been critical to making access to information about these elections widely available including by sharing candidates' messages, helping small and upstart campaigns reach voters with ease, and connecting us to the many events on the campaign trail that we can't attend in person. But there are challenges and harms as well. Researchers, elections administrators, and voting rights experts have raised significant concerns about the use

of digital tools by malicious actors to stoke hate and harassment, create and target false and misleading information, and undermine trust in electoral processes. These efforts have real impacts on democratic participation. They may even lead to violence. At CDT, we work to protect free expression and to advance free, informed and fair elections. These goals always intersect, but never more so than now as so many people around the world head to the ballots.

This FOSO, we seek to explore this intersection between supporting free expression and protecting, securing, and deepening trust in our elections. Free expression is necessary to enable voters to share information about candidates, express their desire to vote for one candidate over another, and advocate for change. Suppression of speech is the first resort of authoritarians seeking to consolidate power. We need only look abroad to see evidence of this, as political unrest following disputed election results in Venezuela has led to the imposition of internet shutdowns by the government, as well as other governmental tactics to silence dissenting voices. At the same time, bad faith actors are increasingly sowing fear and division by disseminating false or misleading information about our electoral process or in a manner designed to influence those processes, both within our borders and beyond. Academic research that seeks to better understand our information environment and strengthen our electoral systems has come under increasing attack, hampering our ability to understand and combat adversarial threats. Layoffs within key Trust & Safety teams have challenged companies' ability to enforce their elections-related policies consistently and equitably, and provide necessary support to election officials seeking to boost trusted information. Meanwhile, local newsrooms continue to struggle with tight margins to make high quality news about the electoral process available to voters, particularly in Spanish and other languages. Together, these factors work to undermine our information environment, sapping trust in institutions and increasing cynicism , and undermining people's access to trusted information as they exercise their right to vote. For these reasons, we must act to support rights to free expression that are the lifeblood of democracy, while ensuring the safety and security of democracy's most essential process, elections.

So how do we do that? CDT has been laser-focused on this effort. In the past year, we've led groundbreaking work on combating the unique risks new generative AI technologies pose on our electoral environment. Today, our Elections & Democracy Team is publishing a new report with Proof News stress-testing popular chatbots on how well these systems respond to questions from voters with disabilities about how and where to vote. Our AI Governance Lab has been diving deep into the creation of rights-respecting best practices Our Research team has published landmark research surfacing the unique and disproportionate ways federal women of color political candidates are targeted by disinformation campaigns, often on race and gender lines, which has become even more relevant as the U.S. presidential race heats up.

Now, as we gather with all of you over the next two days, we'll hear from disinformation and voting rights experts, free expression experts, technologists, companies, and more about what's at stake this election and how each of us has a role to play in upholding our democratic rights.

We'll begin with Renee DiResta, who has been at the forefront of the research and advocacy field that has been working tirelessly against the erosion of our shared reality. Her research into influence operations and the actors who undermine our democracy has led the way in explaining to the public the impacts of rumors and propaganda in the wider world. It has also caught the attention of political actors that have sought to suppress her research and silence her advocacy for their own partisan ends. We're privileged to have her here today.

After Renee, we'll gather a panel of experts moderated by our very own Kate Ruane, Director of the Free Expression Project at CDT to discuss what's at stake this election season and how we can best guard against threats to individuals' access to the ballot and the information that enables it. And to close out today's agenda, I'll be back on stage to chat with two leading company voices Roy Austin, who leads the Civil Rights team at Meta and Ginny Badanes, who leads Microsoft's Democracy Forward initiative, about how they are thinking about the election in this critical stretch ahead.

We're then back again tomorrow, solely online, for more great discussions. At 12 pm Eastern tomorrow we'll discuss how to create systems that are resilient against disinformation with speakers from Witness, the Meta Oversight Board, and Wikimedia.

After that, CDT's Becca Branum will lead a panel about the online speech questions the courts have tackled in Murthy and beyond and how that will shape how companies interact with governments and elections experts to make information about the election available. Finally, we will close out with a panel on how we can create robust data access mechanisms to enable researchers to study the impact technology has had on our elections. It's a packed agenda; we hope you'll join for all of it and spread the word online.

A few housekeeping notes as we get started: We are recording and live-streaming today's session on CDT's YouTube page and via Zoom. We'll go till 5pm ET and for those in the room this afternoon, we invite you to join us for refreshments after the official program concludes.

For our in-person audience, we'll hope you ask questions and you can raise your hand and have the moderator repeat the question for our online viewers.

Online, folks can ask questions either using the Q&A function on Zoom or you can submit questions via email at questions@cdt.org or via Twitter using #cdt questions, #FOSO2024, and @CenDemTech. CDT's Communications Manager Elizabeth Seeger will be monitoring these channels and will pass those questions to our team.

With that, enough of the opening. We will begin with having Renee come up to join us. Thank you all so much again for being here.

>> Hey there. It's an honor to be here today and I hope in 20 minutes to set the stage by making connections between some of the critical themes that we will be discussing as the tech policy community at this event focused on AI elections, and speech. As we approach the 2024 elections, we find ourselves at a pivotal junction for democracy. This is the first US presidential contest since an election that despite being free and fair erupted in violence. This is not an isolated moment of political tension for the digital age's rapid flow and information both [indiscernible] and misleading has directly changed how people engage with elections, both within the US and around the world and yet the political climate within the United States feels particularly fraught.

Polls as recent as a month ago find 30% of the American electorate remains unconvinced that Pres. Biden was legitimately elected. These views were shaped by media and commentary, including from the former president himself, but also quite significantly through viral amplification chains and the improvisation of social media influencers algorithms and online crowds.

The challenge that confronts all of us in this room today, from civil society to academia, to government to tech platforms, is how to best do our part to respond to a crisis of public trust in our elections, at a time when new technologies that are rather adept at disrupting trust in what we see and what we hear are becoming more and more commonplace.

This crisis of trust is not singularly caused by technology nor by social media nor by online speech. But the design and incentive structures of the online communication structure are contributing. Social media platforms are powerful messaging tools that can raise awareness that can galvanize crowds to act, we can see it used to overthrow tyrannical governments yet increasingly campaigns rely on the platforms to shape narratives to mobilize supporters and galvanize potential voters. Of course, the same tools that spread political speech and enhance freedom of expression can also be exploited to disseminate misleading or false information.

One way that this happens is via the now ever-present state-sponsored trolls, sometimes creating false personas, sometimes surreptitiously paying real voices to speak on their behalf. We have already seen Iran engaged in a hack and leak operation in the election cycle and have seen multiple DOJ indictments of Russian actors using a variety of tactics from fake domains to purchase influencers.

Troll operations are usually at least minimally impactful so long as platforms [indiscernible] to disrupt them and not all platforms have. The more pressing and impactful challenge [indiscernible] when it comes to balancing free expression with accurate information as the rise of diverging narratives lead to conflicting views on the

fairness of our elections. These often take the form of rumors, unofficial obligations highly plausible, spread from person-to-person through friend to friend. Unlike misinformation which also, which often involves demonstrable mistakes, rumors emerge from uncertainty, they spread rapidly from person to person before facts are established, before we can know what is true. This makes them particularly potent in shaping public perception when something goes viral the rumor is going to go viral and the facts as they are determined to weeks later are often not. Social media is structurally perfect for this [indiscernible] provocative claims to reach accounts we have come to trust so we discuss and share claims among similarly minded friends online. Social media researchers who do real-time election observation in 2020 like my team then at Stanford Internet Observatory or the other teams that participate in the election integrity partnerships are rumors not information that go viral over and over again. Allegations about ballots or voting machines or voters that seemed plausible with facts unclear in the moment.

These allegations are often quite deliberately picked up by political influencers who frame and amplify them using sensational rhetoric. Big if true. Even as they rely on comfortably familiar tropes. The dead voter. The bussed in voter. The illegal voter. And so many claims about mail-in ballots in the week leading into the election followed by a sharp pivot to allegations of rigged voting machines on election day itself. These highly repetitive claims often created or amplified by incentivized hyper- partisan influences and in some cases by the former sitting president of the US himself were then curated and boosted by social media algorithms that key off of signal attention and push the claims out to a further more receptive and yet distinct audience. The people who are in these audiences in turn distribute claims across all platforms so what happens on one platform makes its way across the entirety of the ecosystem.

In aggregate, many of these claims initially began as rumors became part of a deliberate effort to preemptively delegitimize the election, to lay the groundwork to undermine the election with terrible impact to certain individuals, particularly like ordinary election workers people like Ruby Friedman and Shane Moss, who were caught up in lies and turned into villains. The audience in social media is not passive. If it dissipates directly and shares the rumors which has the effect of deeply investing people in the claims. Rumors boosted by real participants and turned into viral trends by incentivized political influencers is a new and uniquely participatory bottom-up form of political propaganda. It reaches people where they are and comes from people they trust. The net effect is the creation of competing versions of the truth and a build up of community lore, and the entire cinematic universe in which different realities of events transpire. This has contributed to factionalism and splintered realities which raises the stakes for election integrity, and this is going to happen again in 2024.

The challenge of balancing election integrity and free expression is clear when you see this as a rumor mill rather than misinformation or facts, platforms in fact did try to tackle

that information during the 2021 election. They had many policies for doing so. These are not always enforced uniformly but platforms recognized and acknowledged the threat of persistent delegitimization of premature claims of victory. They ran war rooms and moderated content as they saw fit. We spoke to them occasionally at the election integrity partnership when content that seemed to violate their policies began to go viral. After the election we observed that they seemed to have overwhelmingly erred on the side of maximizing free expression. When they do decide to moderate viral rumors most of the time an overwhelming amount of the time the post simply receives labels. The platforms attempted to add more clarifying information. In other words the platforms appeared to respond primarily through counter speech.

Counter speech is foundational to the American experiment. It's long understood to be the most desirable and freedom-maximizing remedy for countering that information but today it struggles to keep up. There are structural challenges in our new communication ecosystem. Online factions are most likely to share rumors about election integrity and are not often interested in nuance or reason to debate. The design and incentives of social media platforms often surface the rumor but rarely the correction. Some algorithms are worded with more strident language, rage inducing claims not the kind of bridging content that might be better suited for coming to a common or simply less acrimonious consensus about the facts.

Meanwhile, the individuals best suited to know the facts in the case of elections, the state and local election officials, are often ill-equipped to get their messages out in the age of network communication. Influencers work constantly to maximize their reach and grow their followings. To tap just the right meme for the moment to inspire their followers to participate. Election workers do not. The professionals best suited to respond to rumors with factors not always aware of what rumors they should even be responding to, and that is because they are focused on doing their jobs running elections not becoming election influencers.

Media and fact checkers are also participants in the counter speech effort, yet they suffer from a crisis of trust among audiences who are now conditioned to believe that everyone outside of this multiple of influencers and media are lying to them. Compounding this, fact checkers are often dependent on platforms designed to get their assessments out in the world as well. Usually via labels or interstitials. These interventions that those who demonstrably traffic in election rumors and lies have now spent the better part of two years reframing as big tech censorship. You are being asked to believe that a legal content label is censorship as well and the members of the same political machine that set out to undermine confidence in the free and fair election of 2020 have spent the past two years working to ensure that network responses [indiscernible] to rumors and lies is a risky undertaking. They did this by reframing communication channels as a cabal and a field as a complex breaking the collaborative bonds that had begun to form in 2020. And they have left American elections less

secure as a result. Structural challenges and partisan residue from 2020 continue to plague us but there are also new challenges unique to 2024.

Since 2020 we have seen significant shifts in the social media ecosystem. With the users on both the right and the left migrating to platforms with content... sorry content moderation policies most aligned with their values. This may alleviate some anger and consternation about specific moderation calls, but does not do much to restore American society to mutual trust or a shared reality. And of course, there are the newly democratized technologies like generative AI, which simultaneously enable the creation of unreality and can be used to further [trust] and belief in the real. Generative AI is not just a technological innovation. It is transforming narratives true and false that are built humbly from deep fake videos to AI generated articles. We now live in a world where digital tools that can enhance free expression can also blur the line between reality and fabrication with ease.

We know that the state actor adversaries [indiscernible] in our elections have adopted these tools. Open AI now joins the method in releasing quarterly reports. Here too however foreign adversaries still largely lack on the distribution front.

Once again, the most impactful use of a novel technology to shape a landscape will come from authentic users. Discussions about the risk of AI in elections focus a lot on the theory of sensational fabricated moments. The idea that a deep fake will swing the election. I'm sure everyone in this room gets that media inquiry constantly.

However, generative AI is already being used to generate misleading content from audio defects to LLM powered persona accounts to fabricated news articles. It's becoming just a part of the political conversation.

We have seen the technology used explicitly for manipulation in elections elsewhere in the world, with fake audio occurring in campaigns in Slovakia and the UK. One thing that generative AI may be used for in the US 2024 election as a means to fabricate evidence to support the kinds of viral rumors that I talked about. Evidence of ballot fraud. Evidence of people doing things with ballots that they are not supposed to be doing, the sorts of rumors we saw over and over again in 2020, now with the capacity to produce evidence to support the false or misleading claims. The evidence will be believed or disbelieved based on viewer trust. This is how technology has played the most profound role in the American political conversation, and we have also seen it shape complex understandings of atrocities and conflict elsewhere in the world. Once again, the core issue is trust and that is not a problem with a simple technological fix.

In 2024 therefore, we are both confronting unresolved tensions from the last election, while simultaneously adapting to technology that is uniquely capable of producing new ones. In both of these realms, there is a clear need to strike a careful balance between protecting free speech, a cornerstone of democracy, while safeguarding the democratic

process. It is understanding to feel a sense of paralysis. Yet doing nothing is not the right call. Communication platforms overrun with manipulators yield actively misinformed publics, which undermines democracy and by extensions undermines whenever most cherished rights related to free expression for the private platforms that serve as pseudo-public squares have really recognize they do bear some responsibility for minimizing manipulation, harassment and election delegitimization. Many companies that offer generative AI tools have recognized the responsibility to ensure their products are not misused. Their participation in the development of transparency and provenance and mechanisms which enable users to be more informed about the content they are seeing is a critical step in our public adaptation to [indiscernible] reality. But this is just a first step. We can see more steps. It is time for the FEC for example to recognize the importance of content and paid political speech such as campaign advertisements particularly those that use an opponent's voice or likeness the FEC might also step up and weigh in on the importance of paid participant disclosures although that is possibly the subject of another talk.

The tension between these two priorities of free speech and election integrity has never been more apparent. The question is how to strike the right balance and how do we protect the integrity of elections without compromising the rights that define democratic participation? and how can we use technology to safeguard voters rights to free and fair elections while also ensuring the protection of speech that supports democratic participation?

As we approach the 2024 election we stand at a crossroads, one where the integrity of our democracy hinges on how we address the challenges ahead. Viral rumors, deepening factionalism and the growing influence of AI aren't just technological issues. They are social challenges that strike directly at the heart of our democratic values. But we are not passive observers. And the power to shape the future is in our hands. Together we can collaboratively and inclusively define the digital landscape in a way that serves the public interest. We have the opportunity and the responsibility to ensure social media platforms and generative AI technologies are tools that strengthen democracy not weaken it. We can create systems that protect election integrity and preserve systems that form the foundation of democratic participation.

Look around the room today. There's so much expertise, experience and insight gathered here. People from within platforms, from civil society organizations, governments and more. We can forge a path where technology works for democracy and where free speech bolsters the electoral process. Over the next few days we can get into specifics about how to use design as well as AI to depolarize, improve counter speech and bridge between groups of Americans to do more to restore consensus. Platforms can shift incentives through design. They can encourage engagement with diverse viewpoints, not just those that provoke outrage or tribalism. Regulators can ensure that we achieve the kind of transparency necessary to understand what is

happening on powerful private platforms. We can jointly resist the pressure of political machines and continue to take action toward achieving the kind of network counter speech that institutions and freedom loving democracy activists must prioritize today. It is up to us technologists, legal experts, policymakers and civil society organizations to make these changes to build platforms that foster understanding rather than division to ensure that technology incorporates the values that we hold here and to make sure in 2024 every vote counts and every voice is heard. We have the tools, the knowledge and the collective will to get this right. So let's use the next two days to challenge assumptions, to explore new strategies and to work together to ensure that technology is a force for strengthening democracy, because the time to act is now.

>> Everyone, I would like to call the folks on the first panel up.

Thanks so much everyone. Thanks specifically to Renée. I cannot imagine a better call to action or a better way to set up this panel. Truly a wonderful talk. Thank you so much for being here and for your words and for your research. We really appreciate it.

So and also thanks so much to everybody else for joining us today. We are excited to delve into the topic of what is at stake quite a bit in this election. What challenges we are facing as we approach the next presidential election in the US. Joining us for this discussion we have a truly incredible panel of experts. We have Cathy Buerger [indiscernible] at the dangerous speech project Washington DC based NGO that is the relationship between speech and violence. We have Tim Harper, my colleague, CDT senior policy analyst for erections and democracy. As an expert in election policy Tim has driven policy and advocacy efforts across industry and civil society. He most recently served as a content policy manager at Meta in political advertising and prior to that he was senior elections policy analyst of the bipartisan policy Center where he was responsible for election administration policy developments.

Right here is Dave Toomey the senior advisor. And last but far from least is Laura Zommer. Among her many accomplishments Laura was the director of the first fact checking organization in the global South that strives to empower individuals with verified content in Spanish, actively engage with diverse communities and build a resilient defense against the impact of mis and disinformation.

As we know are at this point about half the population living countries that are going to have a national election or already have in a very new future we are going to have one here probably less than 50 days I think so there will be a presidential election here and also state and local races Senate all the way down to [indiscernible] do not quote me on that whether that is disinformation you can ask him later but for now that is to free expression rights are necessary to the electoral process. So citizens must have the ability to communicate their opinions, receive reliable information about candidates and advocate for change in order for democracy and elections to be meaningful.

But at the same time as we have heard already there are those that seek to manipulate the information environment to their own ends, spreading disinformation to deceive and mislead people about candidates, about issues and even about the time, place , manner and qualifications to vote. These manipulative efforts can have a real and harmful impact on our democracy. We have seen this just recently as a racist and unsupported claim about Haitian immigrants, many of them asylum-seekers in Springfield Ohio, made an appearance in the US presidential debate last week causing schools and festivals to close. What issues on the horizon to be exacerbated by new technologies like new AI and how the safety concerns play out in Spanish-language committees especially. We will also explore what we can do to inoculate ourselves against these threats while supporting free expression.

First of all, I'd like to start with you. In your work you monitor disinformation that can impact voting and coordinate with experts and allies to respond. Can you tell us a little bit about what you have been seeing so far in your work, what is new and what is old in terms of these campaigns?

>> Yes, I think it's important to set a table this little bit. The spread of election disinformation is not new, it has been going on for decades if not centuries. And it is designed to intimidate voters to spread fear and create barriers particularly directed toward communities of color. So there's a lot of this that has been going on for some time but the rising use of social media has caused this information to spread like wildfire. So it spreads exponentially. It becomes further entrenched where it's almost like everyone is starting to see it and it gets more publicity and in the last few elections information spreads rapidly by high-profile [indiscernible] it gets and it just starts and goes from there. The problem now is that it is spread to the mainstream media. And it is spread to it as having real-life consequences affecting policies, it is affecting for example you have seen more states since the last election that have either proposed or enacted legislation that are creating more barriers, making it harder to use other ways like voting like vote by mail and things like that, and as you have all seen it leads to physical harm and violence. We saw this on January 6. This is not directly related to the Haitian situation in Springfield where there are bomb threats happening all because of false information spreading.

So there are several election disinformation themes that are happening and recurring both in the last couple elections and this election. Renée actually teed up a lot of this already, which is great. They are anchored by the big lie, the notion that Joe Biden didn't win the election, the false notion that Joe Biden did not win. The anchor by which a lot of the notions are spreading but there's a lot of narratives that we see. They include false information about the security of ballots and false information about the security of drop boxes. [indiscernible] Also false information about the use of mail-in ballots saying they are rigged and do not work. And the postal workers are tossing them, that kind of stuff. Preempting claims of fraud before the election has occurred. You see this in 2020

and you see it now. Always on the midterms of 2022 was increasing harassment online harassment of election officials this is a real problem because election officials are civil servants were trying to run election safely and securely in most elections are run really well even during 2023 it was amazing how well they were run safely and secure but when you start undermining that, election officials are leading their jobs and sometimes election deniers are taking their place. They are getting election deniers who are contacting election offices because they are getting disinformation and they have documents from the election workers and they are overwhelmed. If you have election workers who are there to keep elections safe and secure, if you are disrupting that you are undermining the way voting works and you are making our votes less secure [and undermining democracy] as well this year probably the most aligning trend we have seen this is touching on the Springfield situation is the Huge rise in anti-immigrant rhetoric. A lot of disinformation spreaders are using anti-immigration rhetoric to spread disinformation about voting. The idea that ineligible voters are registering or trying to vote illegally. This is generally not happening, but it is getting a lot of attention. It's getting to Congress now and campaign ads and things like that.

So there are real-life consequences to this. As we also saw on January 6th it can lead to real violence and of the most disinformation that we saw in 2020 was after the election was over as soon as the networks called for Biden all the stuff the steal stuff picked up. It all ramped up like crazy. And unfortunately a lot of people who were seeing this or involved in it actually were influenced by it. And I think we'll get into it further and touch on it as well. You add AI to the mix and it has the effect of turbocharging all the themes and narratives. There can be greater volume, more targeted, so I think [indiscernible].

>> Tim, you want to pick up on that?

>> Sure, so I think there are two pieces here. One is what we are seeing and one is what we are anticipating or worried about seeing. And I will say those are very different things so far when it comes to generative AI's role in the election. There are a number of ways in which generative AI has been hypothesized to be used in new ways that can further spread the misinformation that have long existed. I will emphasize that mostly generative AI can be used to spread existing narratives. It is not [manufacturing narratives] that we have seen so far but the first thing that I will highlight is that generative AI can create new ways of spreading existing narratives. For instance, we are worried that generative AI could be used for hyper- targeting localized misinformation. So what I mean by that is, take for example the kind of publicly available data set of people's phone numbers in a publicly available set of people's voting location and sending someone a text message that says hey so you know, voting at peace auditorium has been [canceled] because of [indiscernible] that can be done at a faster scale and a larger range using generative AI like we have not seen in the past and also used for something like translating information, disinformation campaigns into foreign

languages with more accuracy and kind of in community specificity than it's usually capable of, this sort of disinformation can be translated into foreign languages but typically it is time-consuming and costly in a way that generative AI can make that a lot easier.

But I would say in kind of addition to those things separate from disinformation, it also poses risks of misinformation. And this is something where CDT just uploaded that report today to which looked at responses from five different chatbots, Meta Google, Mistral Gemini is Google's and open AI and... Another which [indiscernible] that said, the results found that a vast majority of information has some sort of insufficiency over 60% of the information we found had some sort of insufficient answer. In some instances that was because of straight up misinformation about things like voting registration deadlines and laws for voting online. In many states people with disabilities are able to return a ballot electronically or online. Maybe the chatbots got that information and corrected it which could lead someone to have incorrect information about how they can fill in their ballot. They also incorrectly stated that voters could in some states do curbside voting when they could not, or in other instances saying they could not vote using curbside voting when they could. That sort of misinformation is something we are very concerned about because it is not a theoretical disinformation campaign but good like the average voter seeking information on their polls, it could impede their ability to exercise their vote.

I would say separate from the kind of things we are concerned about generative AI causing, there's also a broader societal risk here which is that fear of generative AI is in many instances now kind of the worst outcome, because in part due to kind of the spread of information that people are aware that generative AI can be harmful they are becoming more afraid that other information could have been created by generative AI and therefore might also be unreliable or misleading which has led people to have less trust in facts. In general. Which is often referred to as the [liars dividend]. This kind of your generative AI playing into elections and the outcome being decreased trust in elections is I thinking the most militias that we are experiencing in this election something society governments in the companies all have a responsibility to try to ensure the public that despite theoretical risks we have been seeing less of this than we anticipated so far which I think is unlikely optimistic [indiscernible] this conversation

>> You are usually a ray of sunshine. [Indiscernible] I am going to make things worse then because most of the time I feel like we are proceeding in a conversation where we are assuming the conversation is happening in English. But not all conversations in this country, in fact many of them do not happen in English. So I would like to turn to Laura because I am wondering what you see in the communities that you represent in the Spanish community. What is the Spanish-speaking community, what are the similarities between what we have already talked about, what are the differences, what is happening in the Spanish-speaking world?

>> Yes, for sure you mentioned that I have been running [indiscernible] Argentina for more than a decade since 2022 we have a project in the US [indiscernible] may be to take action into better focus on this information in Spanish or other languages than English and the reason why we are doing this is because no one did it. Before. And that's a problem. Obviously we can pay attention to what the platforms are doing or not or what academia is doing or not, but the main problem is there are not enough resources. There are not enough clever or smart minds focused on tackling this problem. And what we are trying to do is just to put together [these efforts] taking into account local media that in some cases have just told reporters or academics in different universities to start to pay attention to these because in 2016 in the US was discussing the problem of disinformation targeting Latino communities. But [without strategies or investment] that the US [indiscernible] is pretty small.

And then maybe what I can bring to the table is, it is not necessary that different are completely different narratives and this is not something that's just happening with Latino communities in the US, if we see the main narratives around the world we see global patterns and we have been discussing the same happened with violent speech. The patterns are the same. Bad actors, the ones who are creating disinformation to earn money, are not necessarily interested in innovation. What they are doing is repeating what was successful. What was efficient. If something engages people, anywhere in the world, r copy that adding a hyper local component. And the question I promise to answer now to the point is why it matters to create content in Spanish in a country where 62 million people are Latino and 42 million people answer in the census that they speak Spanish at home. and in some cases they choose to get their news in Spanish. If we don't have a good offer for sure we are letting [bad actors] happen and winning all the battles that you can imagine. And the problem with that is that all the media, all the small media or all the Spanish-speaking media are struggling with resources and also are struggling with the algorithm from the [big tech]. Because if they, the US American disinformation experts were worried about [crowd handle] when they shut it down the problem with the information which is in English [indiscernible] wasn't good enough even to monitor disinformation. Because disinformation in Spanish in this country is not just one. Mexicans in Texas are talking about some topics one day that are not the same that humans or Venezuelans are talking in Florida. Or the discussions from Puerto Ricans or Salvadorenos in New York so for sure we need to figure out better tools that include human knowledge, language knowledge, slang from different countries or regions, cultural knowledge, all these big narratives, that are globally are really [successful] when they include or an actor or an element from the food or something that makes that appealing for you. Then what we need to create for sure is high quality content, high-quality journalism in Spanish, cultural relevance [indiscernible] that not speaking English is not smart, it's just that they don't trust the mainstream media in this country because they don't feel represented. They don't feel they are necessarily addressing them. It is not enough to translate from English to Spanish.

There are some topics that no one is covering. No one is paying attention. That obviously there are efforts that we are doing are pretty small. Now we are 108 media organizations working together. We meet once a week for an hour just to discuss the false narratives in Spanish that we are receiving through the communities. Technology is useful but we also need to build trust. If we don't build trust in that community, that community is going to continue just watching YouTube and getting their news and WhatsApp from families and friends that no one knows who they are. Yeah. Just start from the beginning, start doing high-quality content not just for media but also from not-for-profit, also from the government. How much content in Spanish your own organization has available in Spanish. And when I'm talking available, it is not just in the website [indiscernible] videos from YouTube and [cards from] whatsApp. And that is their choice if we do not give them that, someone else is doing that and it's probably going to be necessarily people who call it what it is. Sorry that I made it so long

>> No, that was incredibly helpful, what you are saying is we should basically listen to people and meet them where they are. I think that's absolutely true. Cathy, I wanted to turn to you, too because in your work you study the kind of speech that moved to people accepting or committing violence. I think that is slightly different or adjacent to a lot of the speech we have been talking about so far but I hope you could talk about what we are seeing online with respect to the election right now, some of the narratives that have taken hold and how technology can impact it.

>> Yeah, thank you and you're right, it's a little different than thinking about access to getting to the pole and voting. We see elections as these particular moments in time that make dangerous speech more effective. So there are things that happen in the world like pandemics and elections and conflict. Already kind of feeling the sense of precariousness and then when they hear there is another threat it is more believable so for the past 10 months we've been working with a team of researchers to document dangerous speech in the United States on a range of topics. And when we are looking for speech that we consider to be kind of dangerous that we can convince people to accept violence as an option we tend to see similar rhetorical patterns that we have talked about that are around the world. Dehumanization is one that a lot of people know but there are other ones, too. So for example any speech that suggests members of another group pose an existential or mortal threat to members of the in group, that is something that can make violence acceptable because it makes violence seem not like an offense of reaction it makes it feel like it is offensive like there is this threat semi-reacting to it is just me defending myself even though we know the threats are often not true.

So in this election we have seen a lot of dangerous speech as I'm sure you have all seen either in the news or in your own kind of communities, dangerous speech targeting immigrants calling them criminals, calling them invaders. Suggesting that if the state of the US continues that everybody is in danger that the country is in danger. We have

seen a lot of dangerous speech targeting trans communities and their healthcare providers, the rhetorical patterns around that focus on the threat posed to children. Again a really common hallmark that you see around the world is that this other group is a danger to our children, and how do we protect our children from them.

>> And family values in the Latino community as well.

>> Absolutely and of course we see dangerous speeches focused on clinical parties suggesting that members of the other party are a threat to democracy, that if the election does not go our way, that there's going to be a democratic apocalypse. There's going to be a Civil War. And we see the rhetoric on bullfights and that is something to keep in mind that when you have a huge portion of the population that is convinced that if the other side wins that we are at a democratic apocalypse, that is not a great place to be heading into an election.

So then we think about the role of generative AI in spreading this quickly as you are saying. And potentially creating kind of the proof that people can point to when they are believing some of these rumors. As something that is concerning obviously.

And I think the other piece of this is that it is super accessible. It is really easy for people to do this. This does not have to be done at the campaign level. This can be someone on their phone while you could be doing this right now sitting in the audience, creating some kind of meme using an AI tool. So I think just the accessibility of that, in the way it can permeate and become kind of saturating, it can saturate narratives where we feel like well everybody is saying this, that is a real part of the rumor spread that you talked about earlier. So that is kind of the primary effect that I think we see.

>> And if you feel that way is that necessarily true? Like if what is creating this is not actually real people doesn't make a difference does that inform anything?

>> Yeah, that's a really interesting question and something, so at the DSP we do kind of two buckets of work. Part of the work is looking at these dangerous narratives that are spreading and part of the work is looking at productive responses to it. How do we challenge them? So we know the speech is there, what do we do about it? So a lot of my work, I am an anthropologist by trade, is interviewing counter speakers, people who make a habit of responding to speech when they see it. Why do they do this? and a lot of our work more recently has also been looking at potential counter speakers. People who think this is a good idea but I have trouble getting going. And this idea of AI, when I come to think of it as the specter of AI has really come up for them because you hear them kind of talking through the decision-making process about, should I engage should I not, where should I engage. In the idea that we don't even know who is right. We don't even know if this person actually believes what they are saying. Is it a bot that is posting it or a person who is saying it because they are trying to start trouble or chaos or they have other intentions. The little kind of emotional friction in the process of doing counter

speech is a big deal. Because it is not an easy thing to do to put yourself out there in that space. So even that little bit of wondering, kind of is this authentic content, I think and has a big impact on the robustness of democratic conversation that we are seeing online.

>> It's a really interesting question of who even is your audience if you are engaging in counter speech are you trying to engage the person who said the false thing or convince somebody else

>> And that really differs depending on who you are talking to some people who do counter speech all the time will tell you it doesn't matter if it is a bot because you're not speaking to the person it's hard to get someone to change their mind and in general you should be speaking to the thousands of other people who are reading but I think still when people are starting to get engaged you are going to get bogged down in the idea, how do you change someone's heart mind? How do we do that?

>> I think one important thing is the idea of immigrants eating dogs and cats, and we don't even, [neighbors] and we start seeing that narrative like the narrative of this election eight years ago. Eight months ago. Then at the beginning we saw it is not just an immigrant doing a crime. They also test some narratives related to migrants [burning] the city and that probably didn't work so they stopped using it, and then they start with this narrative that we are seeing daily saying they are coming to the US to illegally register and vote and change our democracy. And what is not a surprise to me, but it is a surprise that someone can be surprised is that platforms know that. They try to create this platform or say during the debate they are eating dogs and cats [indiscernible]. And then I think for sure it's not effective just to focus as Renée mentioned on facts or disinformation as a specific way to address violence or the threat of public debate, because they are much more than that. Their prejudices. They are biased. They are all [indiscernible] folks helping to spread or just putting down. Being a non-for profit, dealing with this information in Spanish produces high contact. I need to pay for platforms to make our content visible. [Indiscernible] if we are looking for democracy. If we are treating the companies as companies doing business then it makes sense.

>> You lead me to the next part of this discussion, which is you mentioned a few times a few things like we actually can do all of your work is designed to inform people to provide reliable information in the language they are speaking what are the other things that people can do in the face of some pretty intractable difficulties? In both your work or people who are engaging online as regular everyday folks? This is for everybody.

>> One thing we work on is set up similar to what you are doing, is we have an operation where we [indiscernible] partners and we [indiscernible] about 250 [indiscernible] on how to deal with it and whether it be topics or anti-immigration rhetoric or how false information about certification of ballots. We also put out a toolkit [indiscernible] that does not engage with certain things. Don't share or Retweet anything

like that, it just spreads it. Use resources and things like that so we have some guidance but it's hard because people see stuff and it looks really real and they are not sure where it is coming from. Not anyone is going back and doing a reverse search of everything. But looking at trusted sources [indiscernible] not sure where it is coming from it is probably a good chance it is not accurate. That is some guidelines and guidance we give out.

>> I guess I would say there are a few different groups of things that you can do if you are an average person. The first is to protect yourself online. Digital hygiene is really important. Some of these tools make it more likely that you are going to get phishing attacks which are more persuasive and targeted to you. Making sure you have two factor authentication that your passwords are using a randomized password generator, that you are making sure that you are changing the passwords regularly. Those sorts of things make a big difference in securing your online identity which can't be overstated.

The second thing is in addition to those things you can build your resilience to seeing mis and disinformation by doing things like going to fact checking websites. By doing reverse Google searches of an image that you think might be misleading. If it is actually not representing the thing that it said and it came from a previous news event years earlier say for instance an instance of someone doing something shady with ballots that may have been an image that they are using for the purpose of spreading ballot information. That being the case, in addition to building your own resilience you can also debunk things you see online [indiscernible] record misinformation to social media platforms [indiscernible] existing policies if it violates those you can learn to look at the policies of these companies and determine when you can and should report them. And then also you can go as was said [indiscernible] information most of the major online platforms have a place where there is authoritative information about voting [indiscernible] center on Meta the authoritative election panels on search ended in addition to those snap has one as well. Many of the platforms offer sources [indiscernible] fact checkers [indiscernible] that there's things you can do to find accurate information and you should probably anticipate that it will be spread by people that you know and trust and not just by shady sources through advertisements online.

>> I can probably add, try to have as much of an [informative diet] as you can. Your own bias is not necessarily helping you to deal with this. And ask questions to the ones that you need on this information. The better way to address that in a way that can make that person not necessarily change their mind but probably change their behavior to continue sharing is to better understand why they trust or why they believe strongly in that. And in the case of the Spanish-speaking communities in the US and outside you probably know the WhatsApp played an important role [and new research] shows that while white Americans use 50% together news Latinos or Spanish in the US use WhatsApp 70%., And then all the ones trying to better serve Spanish-speaking communities probably need to have a main focus on WhatsApp. What we are trying to

do is we have a chat more people can ask questions send us pictures videos images they suspect to have in lots of cases we already have [indiscernible] because the misinformation [indiscernible] they repeat it and we already debunked that in the past people receive immediate answers, but if not, our team searches for that, we ask our partners in the states or city what is happening, if they listen before to the same narrative and when we create the article and the video we re-share that to the people that ask it. Ask that person to re-share it in their groups. And that is time-consuming. But at least it is the way that we found outside the US that is a pretty successful way. We are taking care of the questions that people ask us to make them help us to spread the content and in lots of cases debunk things that they trust.

>> The thing I will add really quickly is I think from a counter speech perspective trying to redefine what we think of as effectiveness with that we can often think too narrowly that counter speech is only effective if we stop the person from posting the bad stuff. Right? But in reality there are so many other things the counter speech can do and that research has shown it's much better at doing than getting someone to kind of change these beliefs that they already hold. There are probably a lot of people online who already believe whatever you believe that you are trying to get someone else to believe but do not feel brave enough yet to enter that conversation. And as opposed to what research shows us about the bystander effect in person online it works differently. It's an interesting finding that bubbled up through counter space research that when one person does something one person enters a conversation it does not dissuade other people from acting. It actually makes it much easier. It's easier to be the second or third person who would come in and say yeah I also think this is wrong. So even if you are just speaking to other people who might already believe what you believe, we can hold norms against disinformation and against dangerous speech just by convincing other people to also express those beliefs basically.

>> I have one more question after this but I wonder if humor helps?

>> it depends on what you're trying to do again. Counter speech can do lots of different things. Sometimes you are trying to reach a really large audience and make something visible that was hidden before. If you are trying to kind of change someone's mind or their behavior mocking them is not a great idea but if you're trying to make something go viral to get a lot of attention again I think really trying to think strategically about what you would like to accomplish is then matching it up

>> I think probably that…

>> Eating dogs and cats with memes it's like okay if someone was given this... let's make it like a satire immediately. [I did not measure] how that was the end of someone changing [indiscernible] or not but many more people are aware that there is someone just not paying attention or seriously.

>> Thank you. So my last question before we turn to questions from the audience is what are all of you on the lookout for postelection day, what should we be thinking about?

>> I think we are concerned about, this election is close [indiscernible] a lot of the narratives we have seen are [spiking] where the election is going or how the ballots are counted I think some issues with the certification process so that [indiscernible] in 2020 so I think we are concerned about things like that one thing I think we do have a better hand now on the kind of narratives we are going to see. A lot of narratives are recurring. Having said that, that especially in this election this year there could be some we are not seeing. So we have got to keep an eye out we are already doing some postelection like the last thing I want to do, but I think we have got to prepare for it because you saw what happened in 2020 and in subsequent elections, so some of it in 2022 so I think some of it's just going to be...

>> My two points on that is we don't consider the day after the election in November. We consider February. And we are planning on that. And we are recommending [other organizations] do the same. And our fear I discussed with you when we were preparing the panel is audio with AI in WhatsApp just a day or two days before the election that we do not know is going to have an important effect or not. And probably it is how big it's going to be? [indiscernible] gender or narrative related to communism or the narrative that at least for Spanish or Latino communities make differences due to cultural and historical social values

>> I think for us in the violence prevention world, the election is just the start I think. It's not like we hit the election and all of a sudden we all go back to being friends and agreeing with each other and having nice civil discussions. So I think that everything about the day after an election or the week after, whatever we get to decision time that you are going to have probably almost half the population being very upset about the outcome. So how do we think about speaking, again, holding the discourse norms strong and trying to be able to have conversations in that space and learn to live together again in a way where we can talk. And that is a really high bar. To be nice to each other online. It's not easy. But hopefully continuing, having people not be scared away from discussions. I think there are a lot of people at this time who are kind of checking out of conversation and saying I don't want to be part of this I cannot be part of this. in trying to recognize the real threat to democracy that that is. in finding a way to maintain those as we go forward.

>> Not to double tap on several so far but I think what we considered post election I think what really matters November 5 is not the end of the election it's not the. Postelection where all of the certification auditing and results reporting occurs which is also incidentally one of the most honorable periods of an election that is particularly so in the last few years where the percentage of voters who vote by mail has increased.

That means there are more ballots being counted for longer period of time after the election then we've almost ever received in their history. The reason that creates owner abilities is because a lot of results change. And while there were I think very intentional information campaigns during 2022 alert voters about that that was going to be a big difference during the 2020 election and COVID I think a lot of awareness has become a loss of a focus because we have become very focused on the issues du jour being primarily generative AI, but that the primary primary risks have been forgotten so anticipating there will be mis and disinformation [indiscernible] ballots by mail are still changing. So California does not certify its results for like 20 days after the election. So there's a long period where a lot can change [indiscernible] ongoing litigation. I think given how close the election currently looks that it's a pretty high chance that there is a lot of litigation during that time period. So I anticipate that that will be the case in addition to the post election period being vulnerable to recertification. There's also obviously the January 6 and post January 6th . Where the votes are counted by Congress and the electoral count act comes into play [indiscernible] I think there are vulnerabilities to kind of create opportunities for [indiscernible] political violence. [Indiscernible]

>> Thank you all. Audience. Can I turn to you for questions?

>> My question is, we have seen the situation unfold in Brazil between an electoral judge and I think [powers] and Elon Musk who has no powers but is influential and it made me think of the 2022 Brazilian elections and how I believe at the time major platforms were directly collaborating with rural judges in Brazil and taking down the resin requirement I think and 2448 hrs. Can you talk a little bit about that's an example but there are examples like that across the globe about the tension of notice and takedown in the moment of an election maybe in a new democracy and how that is [indiscernible] free speech and the spirit of democracy maybe?

>> I am just, a clear, I'm a voice, I'm against what happened in Brazil and [I never] present that case as a successful case. Some colleagues in Latin America represent it as a successful case due to the result. I am not that person. And mainly because what happened was because it wasn't the government but it was the judge. That would happen online that at least in America that is against the convention [Spanish phrase] derechos humanos. It surprised me and a lot of my colleagues in the US are much more open just to accept that we can just put down content without discussion. And I don't know you all on this, but I think we need to find other ways to improve the quality of public debate and what we are trying to do is explain some [indiscernible] adding information adding context to try to help people better understand but not necessarily asking anyone, not the Pope, not God to decide for us what we need to see or not online. And in that case what it was was an organization , mainly progressive organizations, trying to define a strategy against Bolsonaro that was pretty much useful

for the judiciary system with the laws they have and they are allowed to do that in Brazil to take action in a way that at least from my perspective is against human rights.

>> I think we have time for one more question. Who wants to go?

>> Dan [indiscernible] from the national Democratic Institute. I am curious, you talked a lot about kind of post election locations kind of immediate, but I feel like listening to you a pretty big shift generally with whatever happens as it happened in 2016 and 2022 had huge changes by the platforms generally positive engagement I would say immediately and then it did not last necessarily I think we are seeing kind of a negative trend but any general strategies particularly on platform engagements you know, assuming the election is somewhat stable and we have some positive outcomes in terms of in terms of safety and engaging with them on content moderation kind of issues?

>> I guess I will... [indiscernible, several voices]

>> We need to talk about X. [Indiscernible]

>> X is in a different category. There's been engagement from all of us at the platforms of winning news on these platforms that have rules on their books to address election disinformation. The enforcement of it is erratic. [Indiscernible] high profile users. We have engaged the platforms, they have made some changes that we requested over the years. There has been a pullback it seems in some of the actions being taken, contents that they were pulling back on it, Elon Musk basically took over twitter and said I'm not going to enforce it anymore, some of the other sort of followed suit to a lesser degree. I think Elon Musk gave the ability for others to pull back on that. That's not what we like to see. We think effective content moderation. If the platforms are enforcing the rules they are quite good. I think it would go a long way to address these. We are engaging a team to address it before and after the election. I do not know how it's going to turn out.

>> I would say to post election period is an opportunity for all of us across society regulatory and the companies to hold what happened during the election the cycle is interesting there's a lot of smaller platforms [indiscernible] the twitches of the world that are encountering their first major test which is also the case for major generative AI companies. Each of those industries and smaller companies are going to have a reckoning on this election cycle and I think that there is an opportunity to hold them to account. Now how we go about that I think is a question across society. We at CDC put recommendations for AI developers for improving and maintaining which [indiscernible] policies are in place. Many of the policies put in place for elections are also very short-term. So we might have an escalated pathway for election officials doing a short period during the election to be more durable and complete. So not only a question of platforms going wrong but also encouraging them to maintain the policies for the long-term and that is something we can also do together.

>> And on that note, excellent note to end things on, thank you all for a fascinating talk. We are two minutes away from our next panel hosted by CDT's Alexandra Givens. I hope you all hang out for that while we make some changes and if there are more questions for the panelists to stick around, we are going to feed you and do drinks afterwards, so please stay for discussion afterward as well.

[CC standing by]

>> Welcome to our fireside chat. I'm here to introduce them. He prayed a lot this year about the historic elections this year, the risks posed to them in some interventions we can take, so important actors in the information ecosystem are companies that facilitate the creation and distribution of this information. The rise of generative AI is driving rapid changes across various sectors and the impact on the ecosystem is a pressing concern. How are the technology companies and technologies themselves reshape the way information is shared and consumed and what challenges will arise for government campaigns and the public CDT CEO Alexandra Givens is going to dig into the questions about Facebook and Microsoft about how these companies are addressing the issues and what best practices they can offer the industry. Joining Alex Israel as vice president of civil rights and Deputy General Counsel at Meta a role that is first of its kind in the tech industry and one that's incredibly important for Meta prior to joining meta-Roy was a partner at law firm with the law firm of Harris, Wiltshire & Grannis LLP. He also served as an Honors Trial Attorney with the Criminal Section of the Civil Rights Division where he investigated and prosecuted for more than a decade hate crime and police brutality cases around the country; Deputy Assistant Attorney General Civil Rights Division, U.S. Department of Justice where he supervised the Criminal Section, and the Special Litigation Section’s law enforcement White House Domestic Policy Council’s Deputy Assistant to the President for the Office of Urban Affairs, Justice and Opportunity where he co-authored a report on Big Data and Civil Rights.

We also have Ginny Badanes, General Manager, Democracy Forward Project at Microsoft, an initiative within Microsoft’s Technology and Corporate Responsibility organization that focuses on addressing ongoing challenges to the stability of democracies globally. The initiative includes efforts to protect elections, political parties, campaigns, and NGOs from cyber-enabled threats. The Democracy Forward team also leads Microsoft’s work to improve the information ecosystem, which involves combating disinformation, expanding news distribution, increasing media literacy, and working with community-based programs and newsrooms to use technology to expand their reach. Badanes has spent her career at the intersection of politics and technology. Badanes has been her career at the intersection of policy and technology. Before joining Microsoft in 2014, she was vice-president of political services at CMDI, where she

advised presidential and senate campaigns in their efforts to leverage data and technology to improve their finance and treasury operations. Alex, I'm going to hand it over to you. Thank you so much.

>> Lovely. Thank you so much for joining us. So we took the previous panel about how we have to be careful on not overhyping the risks of AI but they're also very traditional legacy issues around the impact of the information environment in our elections but also people are calling this the first AI election were generative AI is much more available and present in our consciousness than ever before so I do want us to dig in on this piece of the portfolio and how your companies are thinking about it.

Tim on the previous panel gave us a little bit of an overview of what CDT and other advocates have been focused on but I'm curious from the perspective of the roles of the companies you represent how you see the risks manifest. If you think people are over concerned or under concerned and what you are paying attention to. Ginny, do you want to start us off?

>> Great, thanks for the invitation and great to see so many familiar faces. Hi everyone. So as we approach the election cycle we already talked about the 2 billion or four billion people who are going to vote this year, while we didn't start having the conversations about defects in elections we found we could not avoid it because a lot of people are really concerned. We would go into a meeting about something entirely different with someone from a government that had an election and they would immediately ask us what are you going to do about the threat of defects in our elections. So this is a year and a half or two years ago that we really started thinking about what is our responsibility? we are not just a technology company, leading AI company, and investor in a leading AI company, and we need to address the challenges folks have if you asked me what my concerns were two years ago, they are different than they are now and as we have [indiscernible] we have not seen AI drastically impact the events of any of these elections. It has been used but it has been used nefariously at times it has been used in meme generation and humorously and effectively in some cases but yes it has not been the big disruptor. But here's the thing, we are what? 50 days out, if I'm doing my math right from the election. And we don't know what is coming. We know that nationstates have interests in disruption. If you look at the report from ODNI or some of the warnings from US government as well as companies like ourselves in Google and others we know obviously people are looking to disrupt the election. We also know they have capabilities and AI is one of the tools in the toolbox to be disruptive so could we see some kind of defect of audio at the local level to disrupt a small election I think that is still possible. So as a technology company again who is in a position of looking at these and having a role there we are going to stay vigilant and be concerned but to the other point that I think Tim still a lot of thunder hear to the other point he made we also want to be really cautious that we do not overhyping the threat so much that people do not trust anything in the whole liar's dividend takes over. So it is really the balancing act

of how do we raise awareness to talk about the challenges and make sure people know this is not the only thing that could happen. It could happen in your local election. Make sure campaigns know it could happen to their candidate and have a thought about how they would respond if it did. Do they have an action plan? We spent the last eight months starting in Europe and working our way across the world where the big elections are happening trying to train political campaigns and candidates and stakeholders around how they would respond if this were to happen. So those are still pockets of concern we have fortunately I am less concerned than I was six or eight months ago but we are remaining vigilant and going to keep on this, again through not just the election but others as I mentioned through the post election period where we think that technology could actually be used to reinforce some potential agendas or conspiracy theories that are coming out.

>> [Indiscernible]

>> Can I just say ditto?

[Several voices]

>> I have more questions.

>> First of all, let's talk about the big news: you promised me a fireplace. I do not see a fireplace.

[Several voices]

>> Look, let me also just start off with thank you, thanks to CDT for all their efforts. And wonderful, one of the people we turn to and I get information from and get ideas from as things we need to be doing. also let me do a quick shout out to two members of my team who are here, Bobby Hoffman in the front row is leading our voting work for the civil rights team, and then Manar [indiscernible] who leads everybody else on the team. Look, we have not seen the problems I think a lot of people expected to see with AI. We have seen some stuff. We saw what happened in New Hampshire with the use of a Joe Biden voice. AI is getting better. The defects are getting better, the audio is getting better. The ability to do real damage seems to be getting better but we have not seen it yet. Look, this has been a matter for Facebook now, Meta, for close to a decade now. Something we have been looking out for, something we have been trying to stay ahead of. Put in protections for the election, for the vote. But at the end of the day, it is the same old, this is since the founding of America misinformation and disinformation and misleading people for political power or for other reasons. And this has been called by Emmett Clay, our global vice president, an accelerant. So the possibilities are there.

On the other side of that, AI can also help us deal with the issues. So AI can help us content moderate at a scale [indiscernible] AI can be a way for us to identify, and I think

we will get into some of this, identify the defects, identify the foreign actors, those who are trying to do harm to our elections.

Not as bad as I think a lot of people expected. We must still stay vigilant.

>> I think it is helpful to think about two different types of risks that can be exacerbated by AI. One is the defects issue. Misleading AI images the other is coordinated and authentic activity so when people are using, they can do this without AI for sure and I for a while now but using AI to increase the prevalence of fixed social media profiles. How easy it is to generate an entire conversation between both in convincing sounding English-language for example with hyper- targeting etc. and I think it's helpful to talk about those two things because the threat factors and interventions are going to be slightly different.

Let's do the defects piece first and I want to start with you Ginny because one of the efforts that boast of your company's participation in that you were involved in is the tech report this year on 22 companies to coordinate on deceptive AI content the security conference can you talk about a bit about the purpose of the initiative and how it's going now.

>> Sure. We are now 27 companies. We gained a few after the conference. I would start by saying the intention of it was the pull together the leading technology companies both on the creation side of AI as well as distribution side and those who serve in both capacities to lean into the moment and take a little bit of responsibility and acknowledged to the public and to governments we see this as a possible threat. We hear you. Again going back to all the feedback we have been getting. And we are going to lean into our responsibilities on this in many ways. I feel like probably a lot of you are like yeah that is table stakes. I don't think we have had that kind of collaboration around previous threats to this when it comes to the information environment so I think coming together and joining together was powerful in and of itself but of course that is not sufficient not enough we also put together a series of eight commitments about what we were going to do as companies to address it.

Now here is the tricky part. We are very different companies. We have different projects. It was really hard to get to a place where the eight commitments are going to look the same for each of us as far as which is the most important, how we are actually going to play them out so we each individually executed on the eight commitments separately but we came together to agree on what the commitments look like and you can look it up I don't have to give you like a whole pitch on the tech accord but we do think it addressed the main issues people were concerned about both from the creation to the distribution side and on the detection side what we were going to do from a policy perspective again not having the same policies but if we were to have policies we would enact them and on the societal resilience side of it which is a huge component I'm sure we will talk about in many ways is probably the most impactful thing we can collectively

be doing to create a resilient public when it comes to this. That is what we did in February and since then most companies have put out follow-up blogs detailing how they are individually executing the eight commitments. And we are at the seven-month mark of the tech accord signing. Our colleagues as well as a colleague from Google will be testifying on this this Wednesday. In front of the Senate Intel committee. So I think you can expect to see some updates from people about, coincidentally... About how they have been executing against their commitments and how we are sort of thinking across industry about what obligations are.

I don't think it was the perfect agreement. I do not think those kind of things actually ever see the light of day when you're dealing with that many companies and that many interests in a charge political global environment with lots of different pressures from a variety of actors it is going to be hard to come up with a document where all of us are like this is exactly what I wanted. But I will say we made some progress in something I hope we build on as we continue to work across different companies on the key issues.

>> Meta-is navigating hundreds of millions of uploads on any given day across your platform

>> So can you draw on what from a detection standpoint what Meta is doing for AI content

>> Yeah, I'm not even sure if hundreds of millions cover it, it might be billions. The number one thing is doing is [add transparency] [indiscernible] others have to be able to look into what we are doing in things like that so that's a really big effort internally we are [indiscernible] I assume most of you understand is basically doing your internal checks, having your expert internally an expert run through things and I think we heard through one of those with a paper that CDT just put out on accessibility. That is the kind of thing you have to do. You have to ask your AI the hard questions, see what responses it is getting so that you can make sure you are protecting people from bad information, false information or whatever it may be.

You know as a company we are dealing with inauthentic coordinated behavior, we are dealing with coordinated inauthentic behavior. Sorry, let me get my acronym right here. And that is at levels of hundreds of countries, nation-states and individuals constantly trying to either break our systems or to make them create false information for others. Look, it is just incredibly important that we deal with industry, we deal with government and civil society organizations, and deal with everybody trying to get this thing right. Because right now there is no, the rules are being made by the companies essentially. So the companies need to have responsibility and one of the pieces of responsibility I will give a shout out to the fact that my team exists is having a team internally that is constantly looking for these issues looking for these problems. So you want the external pressure and also have a representative here from oversight. You want that kind of external pressure. But you also need people internally, every single day, seeing the

decisions that are being made and having an opportunity to weigh in to the decisions and be able to weigh into the decisions without fear that you can't say what you honestly think about your own products and [indiscernible]

>> Ginny, can I push you for a second on the detection piece and where we are in the state-of-the-art right now? Let's play out the scenario you just described, which is that some video or audio file comes out a couple days after the election. The local elections [indiscernible] are desperately trying to prove it is a defect. How easy is it? And [indiscernible] detection?

>> Detection is a really tricky topic, and it is actually where a lot of us started ... A lot of the companies I will not give you a whole history lesson but if you go back five or six years there was a lot of resourcing and work putting into building out detection mechanisms which is the idea that we would detect bad AI with good AI. And to be clear, detection does work in certain circumstances and it is an important tool in the broader world but it is not a panacea, it does not solve all things.

An example that I use often when people are white why not? Why can't you just go through is a picture of Trump after the first assassination attempt where he is surrounded by a Secret Service agent is obviously a picture that we have all seen, but there is a version of it that circulated where the Secret Service agents are smiling. And depending on which classifier you ran it through, you might get a mixed result because it was an original picture that had been emulated by AI which is different from a picture wholly created by AI so there are gray areas and distinctions you have to start thinking about when you talk about detection.

All those caveats aside, detection is an important component. So one of the things we have done is that we have partnered with an organization, a nonprofit called true media. And what they do is pull together a bunch of classifiers that they built themselves in some like technology companies like us provide them either built to detect AI that came from our own products or just ones we built ourselves to try and determine if something is AI edited or enabled or created and he then works with civil society journalists, governments in some cases where they can submit through their account, a picture and image whatever a video and it will give them sort of a scorecard back. It's not going to be yes! This is AI. Is not going to be that simple, you have to understand how to read it. That's where the complexity comes in. But having that journalistic civil society did not exist six months ago. They did not have an organization to turn to. I know other organizations like Witness who are out there doing this work in the global South and elsewhere have similar projects where they will work with front lines of society groups and they help with his detection and analysis. We also have a form that any political campaign election official who believes that they have a deep think of themselves or the candidate can submit it to just one place it makes it nice and easy and we will run some analytics and analysis on it and try to give them some feedback and then of course if it

violates any policies we will take appropriate action on her own platform. So there are things in place now that were not there before. That doesn't mean that every local candidate knows about it. It doesn't mean that every local official has any idea how to respond.

So with 50 days left I fear that we have not collectively gotten the word out as much as we can but there are mechanisms now where people can start to take some kind of control if they believe there is a fake out there and detection is a component of that.

>> That's great. Let's turn out to the coordinated inauthentic activity and you touched on that Meta has had policies on this for a long time can you think about how your thinking of enforcement in 2024 if you want to respond to concerns raised on the previous panel about a pullback of what we have seen in previous years in terms of this work?

>> And let me just say, add on to the answer here: watermark, metadata, there are things and share those among industries so that everyone can search the images to make a determination whether they are fake or not. Let's add on to the fact that the fact checkers, and what we are doing there, we have over 100 fact checkers so when a fact checker lets us know through their research something is fake, if it affects like time place manner of the elections, then it will be pulled down if it is less than that it will be demoted significantly. So hopefully it doesn't spread but when it comes to coordinated inauthentic behavior, look, we are, I want to say since 2017 we have pulled down close to 200 groups organizations that we believed were operating in coordinated inauthentic behavior on the platforms. We have taken down 700 [indiscernible] 400 of which were white supremacist groups on the platform. We are constantly looking for this. There has been no pullback on that work. On the idea that we do not want coordinated inauthentic behavior on the platforms. We don't want that. We have seen nationstates Russia, China, Iran being the leaders in this space and we're going to continue to look out for that stuff and continue to pull that down so all those things are important. But look. I mean a reason why I come to CDT and meetings like this is we do not catch it all. And we have a direct line. I will give you Bobby's home phone number.

>> This is a privacy organization

>>... Well privacy for some but not others. Truly we need to hear from people and people are seeing things. We miss what's coming in. Look, we have the whole issue of What'sApp and the number of people, and the encryption. We don't see it. So some of these messages we need others to come and tell us that it is there, that it's happening. So that we can act on it but honestly in those spaces I have not seen any kind of pullback with respect to that. We don't want to see that on the platform.

>> Ginny, can you talk a little bit about the role of industry coordinating on these issues for information sharing? This is another area where I think civil society advocates are

worried that the companies may not be working together quite as much as they did in the previous election and I'm curious what your sense of how that is going.

>> It looks different than it did in previous cycles but I will say really the point that he was making about the provenance and watermarking when it comes to how we are as an industry going to get to a place where we can give consumers better indication so they understand what they are consuming, that is something I think we agree is essential and need some level of consideration across the tech ecosystem. If we are out there applying provenance or metadata to images created through Bing image creator and feeding it over to Meta and they are not reading it and therefore distributing a label that if we are not talking to each other it is kind of for naught we need to have some level of best practices that we agree on, we need to [indiscernible] we are pleased to see that meta-join CP to a Google has joined Microsoft as a member Adobe is a founding member and we're starting to see movement in the directions of Providence label making watermarking. Again 50 days it is not all there. These things take a while. I know everybody hates that in tech, but if we look back it is kind of the wild wild West when you used to enter your credit card number into [indiscernible] that seems insane, people will look back on this time and say it's kind of bizarre that we as the industry were not providing more information about the images people were seeing online, we [indiscernible] have not decided exactly what that says, I think if somebody can agree it is wholly generated by AI there should be some indication of that. There should be something that says this is AI generated but if it's added by AI what if you made the sky blue or, does it require the same kind of label, these are not the same kind of things there are a lot of good conversations happening by people on a fairly regular basis so it is a cautionary thing where I say there's progress we are moving in a direction where you are seeing progress day today but it's not fully resolved and of course the tech industry is involved as well so building this while moving forward in the technology is changing around us. We are probably contributing to the change of technology, but still. We are making those adjustments so it's progress, but it does not look the same as it has in previous cycles but of course everything changes over time

>> So that is on the AI watermarking labeling side of the bed moving over to coordinated inauthentic activity or where we see the threats of foreign interference but also it comes up domestically as well. How strong do you think information sharing is there?

>> I would say the analyst-analyst sharing is quite strong, when we see indicators on our platform about nation state interference and through our research identify another company may also have similar indicators, we will share appropriate indicators with them. That conversation is still continuing and strong and it's frankly just not about elections or politics. That is a more general way the industry works together when it comes to cyber security protection, which I think is important. We are a little bit less in the conversation just given where Microsoft sits in the ecosystem around coordinated

inauthentic behavior. I know we have colleagues at LinkedIn who do have conversations like that with their peers and other organizations. It is a little bit less of an area that we play and in some of the other tech companies but we do share information and talk with one another, again strongly at the analyst level.

>> Anything to add to that?

>> No, I mean we all recognize the dangers here [indiscernible] on the information sharing side, we just took down a number of Russian organized groups. We are going to continue to do that work. And then we also do a quarterly report. We have these reports that go out there, and one thing we want to hear at the civil rights team is are these reports telling you what you need to know or if you are reading reports

>> We are reading them.

>> But are there other questions that the reports need to be answering? Because it is something that is public everybody can read it and see what is happening and we want to make sure that we are providing information people actually want

>> I would love to build on that actually because when I talk about things changing, one thing that has also changed is we are all talking more publicly about these things in a way that you don't need to have private conversations anymore. The fact that the transparency reports are coming out so often from Microsoft you would never see a threat intelligence report from Microsoft on nation state behavior in previous election cycles but now that is something we do on a regular basis. So we have also come around to the very clear idea that more transparency in this space is good, and making sure the public hears it and they understand what is coming is important as well so that's one of the things that have changed. A lot of us are just telling everyone about this information, not just each other.

>> I want to take us now to a different topic. Laura spoke passionately in the last panel that we need to focus on not [English-speaking] communities and CDT has done a whole body of research looking at the effectiveness of large language models in non-English-language. It is hugely important because when content analysis for moderation purposes, for threat detection is happening, if the models are not working well in English obviously their significant vulnerabilities for communities. So I would love to hear from both of you and it would be great to have you start on how well you think the platforms are doing in non-English content moderation analysis and threat detection.

>> I think the answer is mixed. Four are mainly AI, English, Mandarin, Arabic and I am missing Spanish. Are the four main languages. Of AI. the way AI works, I think I'm speaking in a room that understands this, the more data and information you have the more likely your AI is to be better. So the more languages we have , the better AI is going to be. We do deal with dialects and regional talk, and everything else. That has to

be figured out. I think languages other than those four, I think we are continuing to catch up and figure it out. It's incredibly important that we get this right. And I think we recognize that. But I think the situation right now is that the language we have the most data on is the language where we are going to be able to do the most work, preventative work, stabilizing work and moderating work. So I think that is where we are right now. I think the other thing, honestly , encryption is something which is difficult, because we heard a lot of complaints from people about potential Spanish misinformation in the last presidential election. As we dug into it and looked at it, it was actually kind of, it was really kind of person to person, family member to family member on encrypted apps, what people were complaining about which is not something we can actually go after as a company. If we are going to respect people's privacy to use that. So we have to think of other ways. Of making sure that people get truthful and honest and useful information. It can't necessarily be at the front-end level of the companies. When they do none of the conversations that are happening.

>> Right, the other lever. You can't do anything about the pathway of distribution. What you can do is help boost a trusted information voting resource center.

>> That they, the voter information Center, in figuring out what is trustworthy information and getting people to actually read it. The problem is people don't trust anything. And there's confirmation bias where people only trust the things they believe are true to begin with. So anything else is false. So you are in this mess right now. Of getting people to understand and believe when you say something is actually truthful. We put screens up. We tell people this is likely false information, and what does it do for half the population? They read it and say that is what I want to read because that's the truth they are trying to hide from us as opposed to trusting any source.

>> How do you think about partnering with organizations in those communities to help boost the trust and raise awareness of those types of resources? Or if it is off platform it is less your responsibility and other communities

>> We help to work with over 100 fact checking organizations. I want to say 40 of them are not English-speaking. By working with these fact checking organizations, we are saying we trust them. We will continue to do so. Again, we are dealing with the world right now where people do not like the outcome, they are not going to believe what you are presenting and we will continue to work with organizations and continue to find ways to try to get the good information out there. It's just we are in this battle right now where people can't decide what is good information unless it makes them feel good about the information.

>> I want to pull us now ... I am getting the time ways from my friends in the back but the trust organization the issue we alluded to earlier about the a accuracy of chat bots [indiscernible] when they are being integrated into a search for more search results how do we make sure they are returning authoritative information we talked about the report

CDT show today about overwhelming numbers of inaccuracies and sometimes affirmatively misleading information about voting with a disability but of course there have been other studies proof news and others just on the general quality and information is being returned.

Can you talk a little bit about how your companies are thinking about that? Is that low hanging fruit to try to fix quickly or is it more complicated? Ginny?

>> It is more complicated than that. It's an area we have been focused on for quite a while when we released what is now copilot which at the time was bing chat it was a marriage between open AI model and a search index. And I had a lot of really smart people explained to me grounding and what we did to ensure the result that came out was really a synopsis of what was on the website, and showed me a lot of really great research that showed that we were not going to run into the problem of coming up with false information within the results. And I would say that the team has done a lot of [indiscernible] work and the copilot response is pretty strong. We've gotten pretty good grades on the things that have tested it but when it comes to election information we did not feel like it was worth the risk where if one in 10 times it gave you accurate information or slightly misleading information that felt like that was not okay, not an acceptable threshold. So we have since shifted if you ask a copilot about election information is going to ask you to do a normal search and that is because that was too risky. I hope that we get to a place in a year or two years. I don't know how long it will take us where we can stand behind whatever the answers are but when it comes to high-risk situations such as an election it is not worth taking the risk just to be able to prove our product can do it.

So while we do actually feel good with the results we have come up with through red teaming and testing for now that is the response we have. Similarly on the image creator for example you should not, and I feel like I am testing a group of experts [several people speaking] you should not be able to create an image of one of the presidential candidates for example. Those feel like necessary precautions that at some point in time I would like for us to be able to lift if we have the proper restrictions around it but until we can get to a place, this information is too important to risk so that is the position we take and now we will continue to improve. We are not perfect but I think we have the right policies to keep the public safe to the extent that we can and I know we can talk later about it but the importance is making sure people have access to authoritative information when they do the queries and that is the most important thing.

>> I mean, this is hard as hell, it really is, and when we talk about authoritative information we send people to secretaries of State's offices and some of the secretaries of State's offices are trying to be really inclusive and trying to put things in multiple languages and some are not. So even calling that authoritative information, while it is true it is not necessarily information the population needs. An example I use right now is

playing around with one of our generative AI tools, and the question was, show me an American family [indiscernible] And you get answers from that and the four answers we got from it were for white families. completely ignoring the fact that Native American families were here in the 1800s completely ignoring the fact that African-American families free and enslaved were here, and multiple other races were here. Your colleague [indiscernible] really understands this from Wikipedia so much of what is on the Internet was put there by certain groups of people, and that is where the information is coming from so asking generative AI and AI to come up with accurate answers to questions, you are immediately starting from a point of bias. And the question for every one of our companies is what do you do about our bias? What do you think about the bias, is it because of information sources? I don't have an answer for you as to how we deal with this other than to say this is really really hard. And that we need more and more diverse voices and people weighing in for us to try to get this thing as close to right as we can.

>> That is a wonderful note to end on. I am going to wrap us up [indiscernible] so terribly. You had a packed agenda today the guests will be staying around for the cocktail reception so will be able to cover things then. In the meantime please join me and think of these wonderful guests. I was asking if Kate is going to be back up to wrap up the conference. I'm going to take the mic for now. We are grateful for all of you who joined us and those of you who joined us over the livestream thank you. We are happy to have you and we are back online tomorrow at 12 o'clock Eastern. We look forward to continuing the conversation then. Thank you.