ROUGHLY-EDITED TEXT FILE

CENTER FOR DEMOCRACY AND TECHNOLOGY

Report Launch – Offensive Speech and Hate Speech Targeted at Congressional Candidates in the 2024 Election

October 2, 2024
9:30 – 11:00 am ET

REMOTE CART CAPTIONING PROVIDED BY:
Michelle Anderson, RPR, CRR, CRC, CSR(A)

* * * * *

>> Good morning, and welcome to the launch of the policy brief hated more online violence targeting women of color candidates in the 2024 U.S. election.

My name is Müge Finkel, and I'm the director of the Ford Institute for Human Security at the University of Pittsburgh. I'm very excited to be co-hosting this launch with my colleague, Dr. Dhanaraj Thakur, at the Center for Democracy and Technology, and we're joined by participants in person and in Washington, D.C., as well as participants streaming live, including our research partners at Koç

University.

We will first start with words of welcome from Dr. -- dean of graduate school public and international affairs at the University of Pittsburgh. And our gracious host in this venue before turning to Christina -- the founder of the hashtag she persisted, which is a digital global NGO focused on intersection of democracy, women's leadership, and digital distortions.

We are very, very lucky to have Christina's expertise this morning to help situate our research in the field of policy, advocacy and global research at online violence against women in politics.

We will follow up Christina's remarks with our presentation on the top-line messages from our research. For our participants in the room, we have physical copies of the policy brief, while our participants online can access the brief at the Ford Institute and CDT websites. We included these links in the webinar chat.

We hope to have at least half an hour for question and answers following our presentation. We ask that you submit your questions using the Q&A tab on the webinar. One of our team members will be moderating. She will make sure that we will get to as many questions as possible, and when we can't take them at the moment, we will direct you to a more detailed information we have compiled about the study.

We will also have the recording of the event available should you like to share among your networks.

Now without further ado, it is my pleasure to invite Dr. --

>> Thank you so much, Dr. Finkel, for the invitation to be here and to share a few comments, and for all of your effort in bringing us together around what is I think we all agree an extremely important topic.

The Graduate School of Public and International Affairs at the University of Pittsburgh is also just really, really grateful to be part of this important research, and especially grateful for the partnerships that are at the center of this work.

Our special thanks to the Center for Democracy and Technology and Koç University for their collaboration. This university, and I think all that is ahead for this work, is really a testament to the power of bringing together experts in epidemiology and experts in a wide range of other organizations to address the biggest challenges that are facing our society.

I think we can all agree that this research on online violence against women in politics is of -- and with its particular focus on women of color is so, so crucial.

As we see this phenomenon playing out in the U.S. and across the world, women and women leaders in particular have long faced the kinds of attacks and harassment not faced by their male counterparts, and as we've seen divisions grow in so many different societies, the rhetoric and the threats have grown and the mechanisms to share them has significantly grown as our online environment has changed.

I'm really grateful for the expertise of this research team and this partnership. The Ford Institute for Human Security at Pitt has been a leader of raising awareness on a whole range of human security issues, and this work conditions that tradition, and I think is poised to make an important impact on how we understand political and online violence, and especially its gender dynamics.

I'm really hopeful that over time it can also shape the remedies and the policies that we can use to address these issues.

Thank you so much to Dr. Finkel, to the whole team. I am really looking forward to hearing from you today, and I look forward to learning more in this session. My thanks to all who are joining

us in person here at the Washington, D.C. center, as well as online. Thank you so much.

>> Thank you very much, dean. We really appreciate having you here and having your support throughout this project. We're grateful.

With that, I'm going to turn to my colleague, Chris, and again, she is the co-founder and co-director of the hashtag she persisted in leading NGOs, and we are very lucky to have you join us this morning.

>> Thank you. Thank you. It's so wonderful to be here.

Let me just frame a few points that I think can help us in discussion. I myself am just now going through the data. It was embargoed until today, but I can see some similar themes and that I think are relevant to all of our conversations around these issues.

And one is we do have to have data-based conversations in order to inform an evidenced discussion, and we shouldn't have to work so hard for that in the way that data from the social media companies is now being restricted, severely, so that sort of is a topic we should dive into.

Many of those who joined us in the room this morning also had to walk up eight stairs because an elevator is not working, so we in this room are very committed to this topic, and so maybe our private industry partners can help out with that. So that's sort of one thing.

So any type of new evidence-based information helps us put context to the parameters of what we're seeing here. So let me lay out a few thoughts. One is it's really important to consider the unique experience of women in politics, and one positive trend is that we're now addressing that the online world and offline world doesn't have a lot of separation.

You know, there's a greater appreciation than, let's say, four or five years ago when it was really seen as mean tweets and women if you can't handle it, you know, then get out of politics was sort of the tone.

So now we have a breadth and depth of data that informs what's going on here. But we also need to further put it in the context of what this -- these types of attacks against women who are in politics, who are elected leaders, who are running for office, are meant to achieve, and that's where in she persisted we ground this in the context of what's the state of democracy.

Why is this happening? Is it because the world is growing so much more misogynistic and racist, or is it because the social media tools that enable this are aimed to benefit a set of anti-democratic actors who are using illiberal methods in which to foment opposition to the pro-democracy position of women.

So that's sort of a sticky issue to unwind, but that's where I think we have to consider the experience more deeply of the women who are targeted, because the intent here is, in fact, to distort and silence. And in the United States we're having a very distorted conversation even about this -- not only the social media platforms but the regulations, the policies, what's happening and why.

What we have seen, no matter the country or context, is that this has a silencing effect. I talk to more women activists, women who are somewhere in trying to influence politics who say that they don't want to be involved in public advocacy anymore. They are withdrawing themselves. They are getting off of social media.

And that's not equal because men are not doing the same things, so if we're going to have conversations around protecting free expression, we must also protect against what happens when self-censorship, when survival online means getting off online, getting yourself removed.

And so that is not as deep a part of the context of how we talk about these issues, so that's sort of one thing that I think we have to dive into. And what we've seen is the second issue, is that the women who are defending fundamental human rights are the ones who are attacked the most.

And so often with media and people we work with around the world, we're asked, you know, well, is there a partisan difference, is there a party difference? There is. There is in the distortion of gender as a topic as a driver. So that's sort of one thing that is also worth unpacking a little bit here.

And this is where the disinformation and the hallmarks of disinformation come in, because of the asymmetrical media ecosystem in which our discourse is had, and so the two prongs of the problem that we've been mapping in the United States and other parts of the world shows that not only are women leaders and women candidates and women in politics attacked more severely compared to their male counterparts, but it has a distorting effect on the perceptions of their leadership.

So it's like a forever chemical in the damage that is being done here, and the second prong of this problem is, then, the way that the anti-rights agenda that's unfolding, anti-transgender, questioning women's equality, attacks against DEI, they are meant to serve a certain agenda that is about rolling back rights, and women on the right are not defending those, are not talking about these issues in the same way.

They are not talking about their sexuality in the same way, and so when it comes to image-based abuse and distortions and deep fakes, we see then the imbalance of who is targeted and how they are targeted.

I want to add one additional point to this conversation. We as those who are think tanks, academics, advocacy organizations, we have to come up with new words, right, to describe what's going on and to really capture the essence of it.

And so one of the phrases that has become I think the preferred framework for this is technology-facilitated gender-based violence, referred to as TFGBV. I like to remind ourselves, those of us who are working in this kind of world, is that we need to really be a bridge to understanding with policy-makers and the public.

And TFGBV is a little bit -- needs some tweaking, right? What is going on here is really an outgrowth of anti-democratic behavior and political violence. So if we center the conversation as gender-based violence as the root of this, I think we end up with a skewed analysis of what's going on.

So gender-based violence is a convenient tactic in which to tap into misogynistic communities and grow them, but that's being done through the hallmarks of disinformation.

So you can't separate -- we shouldn't separate out what we think is actually happening in human reactions to women's rights and the growth of women. We should see the progress in the world, though, as the vulnerability right now of certain audiences who want to stall that and want to bring it back, want to challenge it.

So it's a rich conversation, but we need to broaden our language too and think about this as gendered disinformation. I think in the United States we're very caught up in whether we can even use the term "disinformation" anymore, but when I look at data from whether it's Brazil, Moldova and the mapping we have done around the world, these are influence operations.

Let's be very clear. That is meant to undermine women with real people but with the hallmarks of distortion digitally, and that can only happen at scale with the tools of social media and with the lack of trust and safety within the social media companies.

So I'm interested in having a conversation around the infrastructure that then is the threat surface in which these kinds of attacks are organized. Less interested about the misogynist next door, because he's always going to be there, and he's always going to be trying to undermine things, but how does this then get weaponized to the scale that it is is because of the broader fight for

fundamental rights.

And so that's a different conversation. I also like to remind ourselves we need to have strategic communications. We need to be that bridge of understanding to the public. We shouldn't talk in code by using an acronym that then is literally inside language.

When we talk about algorithms, this is simply decisions that are made by humans. There is a human in the loop here, and so that would be the last point, let's broaden out, before we dig into the data, what are some of the solutions and what are some of the remedies that are fit for purpose.

So that first point, research, data access, not sure how we're going to get it because we need to leverage power against the very powerful digital monopolies that are really telling us that they will do whatever they want.

Secondly, we need short-term and long-term solutions, and I think we'll get into this more in discussion. The short-term being how in a state of emergency do we help women who are still active in politics manage their images online? How do we protect against the attacks that are meant to undermine all of women's leadership? Not just the women who are targeted.

Because it is meant to create an unfair playing field in which their qualifications, their appearance, their laugh, their sense of control of themselves and their emotion, is now fodder for debate.

And if we put that in a framework of political violence, that is orchestrated. So back to my first point, this is why GBV is not the root. There's an orchestration here when we're talking about using attacks, incitement of violence and hate to change democratic process to potentially suppress voters and to distort the image of women.

And so short term what do we do? One, the research is not our public message, so don't do that to women. And this is an interesting I think moment when we are in an engaged debate in an election how we talk about these issues responsibly, and then how outside of an election we also grapple with some of the stickier issues.

So what we can see in the context of the United States is that more conversations around historic race that this is, or the barriers that women face, create an electability myth in voters who are not necessarily deciding whether who they are going to vote.

And so we need to do women a service by putting the context of what their everyday experiences is into what can help gain ground, what can help create the offensive digital armies.

Now we'll never be able to compete in platforms that have no trusted safety and aren't complying by their own rules.

You can't do a positive mean-based effort in which to counter that, but at the same time we have to operate with the tools as they exist and be able to support women in office.

So the research, the harms that are being done, are not the message. We have to pivot in more clever ways.

The long term is we need a variety of solutions that creates a more serious atmosphere in which we can get new rules for social media companies. I know everyone gets fatigued about talking about tech accountability, there's no progress.

Well, there's progress in other parts of the world, and the U.S. is being left way behind, and that's the Digital Services Act and what we've seen in Brazil in going toe-to-toe with a platform. We can leverage influence here when we create the political will and an enabling environment in which to do that, and we should do that in the United States.

So the last thing I'll say is if you didn't pay attention to the Department of Justice investigation that resulted in indictments over tenet media and the role of Russia today, which is -- was if you're in

Europe you wouldn't be able to see Russia today videos on Facebook, but if you're in the U.S. up until two weeks ago you would, right?

And so let's unpack that. What does that have to do with women? Why am I bringing it up? I'm bringing it up because if you look at the way that that Russian influence operation [indiscernible] and evidence around how it was organized, they were dependent on right wing influencers in which to carry those messages.

And those messages are anti-gender. They are distorting around the topic of gender equality, DEI, all of these narratives that are able to [indiscernible].

So our adversaries across the world know this. They have focused on gender as a tactic for many years, and the anti-democratic players in the U.S. also are taking that mantle up. So we have to be honest about some of how this gets organized.

There was a reporter who reached out -- and this is just one tiny example of the responsibility I think we all have, whether we're academics, researchers, think tanks, we're working on the policy side, we're journalists trying to figure out how to cover this issue. The role of mainstream media can either help or hurt.

How we try to create knowledge about the system, but at the same time realize we're in an active debate and an engaged debate, and what we do to expose can hurt or help.

And so the German reporter that reached out -- and this is one of many media stories that I don't do and I don't participate in -- was wanting to do an article about the Internet rumour that Michelle Obama is transgender. So this is one of the themes we've seen in the attacks that women leaders, members of Parliament, Prime Ministers, women who are perceived as strong publicly are subject with this kind of damaging character attack.

Which is meant to undermine a real conversation around transgender rights that is important for the world, for our country, but it's also meant to create the idea that the only power and authority for women comes if they are secretly men. So it's a double way.

Now why is that story not important for mainstream media? It is damaging. It is carrying these damaging narratives forward, meant to undermine women leaders across the world.

And so my conversation and back and forth with this journalist was why they shouldn't cover that, why they are part of the ingredient that main streams attacks and makes it difficult for any woman to compete in this environment.

So lots of issues I've thrown out so we can have a robust conversation, but that's where I would end, and thank you so much for this research and for the work that you're doing and the leadership that you provide to really ground this in the context of women in politics.

>> Thank you so much, Christina, and if we weren't intimidated before with this research, I think I am right now.

But it also gives us great energy to be kind of part of even if it's a small part that we want to play in these very important roles that we want to kind of carry the conversation forward, carry the conversation for productive ways, and you have helped us identify.

So with that, we are now going to take the next 25 to 30 minutes to share with you the top-line findings from our research, and with that, I turn to my colleague, Dhanaraj.

>> Great, thank you so much, Müge. Thank you for the points that you all raised.

So we're going to spend building on what Christina previously said, we're going to spend the next 30 minutes or so sharing details about the research that we produced together.

We're going to use a slide deck to do that, and we're going to take turns switching back and forth.

>> A true partnership.

>> So hopefully that's not confusing, and hopefully everyone online can also see the slide deck. Great. Thanks.

All right, so let us start by just saying a bit more about the background of our respective organizations. Starting with the CDT. So the CDT is a non-partisan non-profit organization based in Washington, D.C. and in Brussels. We primarily advancing on focusing civil rights and civil liberties in digital spaces.

We cover a range of different topics. Free expression, election processes, privacy data, surveillance and so on.

As part of the research team within CDT, one of our major objectives is to identify evidence gaps that can help inform policy debates, and this project that we discuss here, and it also builds on a stream of work that we have been doing for a while, is one of those gaps, because in working with different groups, advocates, other policy groups, such as She Persisted and others, there often is a lot of evidence around this issue around violence against women online and in public life, but particularly around this connection.

But particularly around this connection. And more specifically on the issue of women of color, which is the subject of this talk we're going to look at.

We primarily focused on civil rights [indiscernible] online, and as part of the research team, we want to identify these gaps that we shared in the model today. Let me turn to Müge to talk more about her institute.

>> MÜGE FINKEL: Sure. Very briefly the Ford Institute for Human Security is part of the Graduate School of Public and International Affairs at the University of Pittsburgh. We were founded in 2003, and we take pride in being one of the first academic centers in the United States devoted to the study of human security.

All of the cross-cutting threats to human security we at the Ford Institute choose to focus our research and our expertise on the three, and those are gender equality, food and water, security, and immigration.

Briefly on the research team, as you can see, this research team behind this policy brief is truly international and truly interdisciplinary. Again, a source of pride for all of us.

As you can see, the team includes researchers with expertise in public policy, political science, and computational social sciences, and we are most grateful for their contributions to this brief and our online collaborations because as Christina highlighted, while these issues that we're talking about today in particular the data from the U.S. elections are common across the world.

And there are many people like us doing similar research, so we are very lucky to have a team who have thought about these issues in a different context and come together helping identify some of these common threats, and as well as the differences.

So we are very happy to have this collaboration behind this work.

>> DHANARAJ THAKUR: All right, good. A bit of background then on the research itself. There's an increasing amount of work on this problem around the challenges and kind of the forms of violence, online violence that women in public life face, so women politicians, journalists and others.

There is also a body of research that looks at -- that seeks to understand the additional kinds of variables or forms of oppression that these women face online, and here I'm speaking about identities around race, gender, age, parental status, immigrant status, disability status, and so on.

The reality for all of us is that we hold multiple identities, and the reality online is that these

multiple identities become different forms of oppression, and this is particularly true for women politicians, women in public life that hold these multiple identities, and so part of our motivation here, which really builds on research [indiscernible] that has been done for some time now for many groups, activists and others.

Particularly activists and others from communities of color is that you have to take an intersectional approach to this problem. So it isn't just about gender. It also involves analysis around issues like race, immigration status, parental status, and so on, as I mentioned earlier.

So this issue of intersectionality becomes important, and as I said, the idea of approaching this from this perspective is not new and has been promoted and pushed for by many other groups before us. So we wanted to build on that, and two years ago released a report called "an unrepresentative democracy" that took that and focused on women of color politicians and [indiscernible] gender disinformation. This is what Christina was referred to earlier.

It's this idea that attacks on many of the women politicians online are a combination of not just online abuse, hate speech, but also misinformation, because it's hard to separate these two.

In reality, the campaigns that seek to undermine their legitimacy and seek to undermine their electability and so on combine these forms of attacks, and so we found in our previous work one of the [indiscernible] in the audience [indiscernible].

So building upon that and going further, that project was in the 2020 election, and now we have a new [indiscernible] here in the U.S., so we started to look at the state of play with that election now.

And so we do take a different approach with our colleagues at University of Pittsburgh, the Ford Institute and Koç University. In this project we did not look specifically at disinformation as we did in a previous project.

We also take a different approach by using language models, and we will get into more detail on that later, but that also builds then on prior experience and research at the Ford Institute as done on the Turkish elections with our partners at Koç University.

So in order to complete this exercise we focused on conversations or posts on Twitter, or X, formerly known as Twitter, and to do that we looked at a list of candidates running for Congress now, so we compiled a list of over 1,000 candidates.

Then we collected tweets, any tweet that mentioned the candidate's Twitter account was included in our data set, and this was done over a time period of this summer, between May and August this year.

And so based on the candidates, 1,031 in all, we were able to build a data set of over 800,000 tweets.

In addition to that, because of the candidacy of Vice President Kamala Harris, we also included mentions to her Twitter account in this data set as well, and so this is a combined data set that we analyzed using our language models. And [indiscernible] Ford Institute's website, along with more details [indiscernible].

>> MÜGE FINKEL: Okay, so I'll speak briefly to our methodology, and I say briefly with all intents and purposes because we have a very longer description of our methodology that is going to be posted on the website there for you to kind of ask all kinds of curious questions and hopefully that the report that we put together will address that.

But to start up, we developed a conceptual typology for understanding and classifying types of offensive speech targeting political candidates -- [indiscernible] about the point about the previous research done by CDT demonstrates for those of us who are curious about the field that there are a

lot of acronyms that are going around.

And every piece of research produces its very own acronyms. So instead of helping clarify the field to understand better, to measure better, and remedy better, we are actually kind of adding to the complexity of the concepts.

So for our own sake, we decided that for this particular research we would focus on offensive speech targeting political candidates. We define in our team offensive speech as words or phrases that demean, threaten, insult, or ridicule a candidate. Those are our four types of offensive speech.

And that may turn to hate speech, and for our study and our research we define hate speech as a subset of offensive speech. This is really crucial to kind of get it across as clearly as possible because you're going to see our top-line messages, and some of those findings may surprise you.

But I would like to remind you and I would like you to recall that there is offensive speech as the bigger category, that has those three types of functions, and hate speech is a very specific subset of offensive speech where a specific reference is made to someone's identity, including race, gender, sexual orientation, or religion.

Based on this typology, our team developed original LLMs. So I will throughout out my first acronym that I will qualify, the large language models, and these are natural language processing tools that computer...computational

social scientists are very good and very keen on developing and making them do their jobs better.

So we developed these original LLMs for binary hate speech, offensive speech, and offensive subtext, the four subtext that we mentioned.

For the hate speech subtypes, we use a model developed by the researchers at UC Berkeley, and again, the idea doing this while developing our very own models was necessitated by the fact that the existing models -- and there is a whole menu of them -- did not necessarily perform well with the tweets that we were interested in collecting and categorizing.

So we worked on developing our own ones and making sure that their measurement scores were to the taste of everybody who was curious about these. So they performed better.

So again, I will refer you to a much more detailed and you can tell I'm a social scientist not the computational social scientist to kind of address those questions that you may have much more properly to the link that you see on the slide here.

Okay, so to the meat of our study, right, we are here to discuss.

With our focus on the 2024 election, we compared the levels of offensive speech and hate speech that different groups of Congressional candidates are targeted based on their race and gender. We paid a particular emphasis on the congressional women of color candidates, as well as the U.S. Vice President Kamala Harris as a woman of color and presidential candidate.

So top, top, top line finding, overall we find that offensive speech and race speech targeted at Congressional candidates are not distributed equally, and that race and gender matter for the varied experiences of Congressional candidates that they have online. So to qualify we have the top four findings that we're going to unpack very shortly individually and give you more context.

But to qualify this statement that offensive speech is not distributed equally among all types of groups, as well as hate speech, our very first finding concerns a significant amount of offensive tweets that women of color candidates receive on average.

Reaching up to 20% of tweets for Asian-American and African-American women candidates. Second, we find that a similar trend holds true for hate speech, where African-American women

candidates in particular are targeted with significantly more hate speech than any other group that we [indiscernible] and we will give you those comparative numbers to kind of get you to see what we mean by significantly more than any other group.

Our third finding is that party affiliation is an important factor shaping the levels of hate speech women candidates receive. Again, listening to Christina, this conversation didn't happen before the brief. Some of the points that were raised, like the types of female candidates who speak for the fundamental rights and find themselves defending those as part of their politics may come into effect.

But at the moment we find that at least party affiliation is a significant factor shaping the levels of hate that we find candidates receiving.

Finally, it was impossible not to inquiry about Kamala Harris's experiences vis-à-vis the Congressional candidates.

We find that Harris received proportionally less hate and offensive speech. I hesitate because I think as many of you we also find this quite puzzling, and we acknowledge among the team a series of factors that may have played into this finding, but we will get to that as we kind of unpack these findings individually one by one.

>> DHANARAJ THAKUR: Right, so let's turn to that and start unpacking these findings in more detail.

So the first one we will talk about was on offensive speech. So again, just a reminder we're talking about speech -- think of it this way. Tweets that include some form of profanity, some kind of demeaning kind of language or trying to ridicule the candidate, et cetera.

They are not including references to identity. This is significant. We're not talking about identity here, and yet still we find that women of color candidates are more likely to be targeted with these kinds of tweets, and more specifically we are talking specifically about text within the tweets there.

Asian-American candidates, African-American candidates are much more likely to be impacted by this, more than one in five tweets that they are subject to includes this kind of language.

The overall average for candidates across the board, so regardless of race and gender, was around 60%, and so this is the level of targeting that white men candidates, men of color candidates and -- -- were subject to.

So the point is here that these two particular groups of women candidates, African-American and Asian-American, are more likely to be subject to this form of speech, right? And remember, we're not talking about race and other factors yet.

Overall, there is a gender gap. So if you just look at women in general, regardless of race and -- -- and so on, women are more likely to be subject to offensive speech. Based on the data that we have, a women candidate is more subject to this kind of speech.

What was also interesting in our analysis is I mentioned different kinds of offensive speech, profanity, ridiculing the candidates, et cetera.

The pattern of -- the proportion of these tweets across different groups [indiscernible] so the amount of tweets that contain some kind of ridicule or profanity that was subject to African-American candidates or white men or white women was essentially all the same, the proportions were all the same. So the vast majority of candidates were subject to this kind of ridicule, which in other words are tweets that don't necessarily include profanity but are trying to demean them or ridicule them in certain ways.

So for example, calling them stupid or one example was that you're dumb as a rock, or things

like that, right? Profanity was much less.

So it's interesting that across the board these proportions were similar, but however, women of color candidates were proportionally more subject to these.

The other category is hate speech. Here we are talking about any kind of tweet or comment that references things like the candidate's race, gender, religious affiliations, or something about their religion, sexual orientation and so on. So here there's a much more significant difference across the different groups.

And we're looking at starting again by looking at women of color candidates on the slide. African-American women candidates are subject to significantly more hate speech than other women of color candidates, four times as much, in fact.

Again, so if we look beyond just women of color, so beyond what we have on the chart here, and you compare this to, say, men of color or other groups, it's all significantly higher. In fact, it was -- the African-American women candidates were subject to was seven times as much hate speech compared to men of color, more than three times [indiscernible] and 18 times [indiscernible].

Now you could argue that with white men because we are talking about identity here they may be less likely to be subject to this kind of speech, but the key finding is a huge difference that -- the key finding I want to raise is there's a huge difference between the two, right?

So this points to what many researchers, particularly those from communities of color, already pointed out before, and I mentioned this at the start of the background of this research. It's not the idea that African-American women are subjected to this kind of speech online. It's not particularly new.

But I want to raise how proportionally it's a drastically different problem from the other candidates. It also reinforces research that already exists in the social sciences and political sciences around the kinds of experiences that African-American women have online.

Think of the work of people who talk about decision R. There's also research that focused a couple of years ago about the toxicity of [indiscernible] particularly for Black women online.

So this kind of severity around these kind of [indiscernible] is not new, but we are -- it shows that this trend continues right now in this current election cycle, and it shows the kinds of additional obstacles that some candidates face.

>> MÜGE FINKEL: Okay, as we continue with looking at the hate speech patterns, as you will remember, I said that we found out that the party affiliation is a significant factor shaping women's experiences. So here we are able to show you how significant it is, but before I kind of, like, have you look at it and say "this is really important," I want to kind of sneak in another interesting factor.

In our data set, we find that while on average Republican women candidates are almost twice as likely to be targeted with hate speech than Democratic women candidates. When we look specifically women of color candidates, we see this effect reversed and much higher.

Right here you can see that the Democratic women of color candidates are subject to hate speech 19 times more than the Republican women of color candidates.

And this kind of goes into the point that Dhanaraj made earlier, and we all know -- we kind of talk about intersectionality being the favorite buzz word at some level, but it also kind of comes into the point of this helps us to show how and why our intersectional identities indeed matter.

Because if you only looked at the women's experiences, we may not have seen that, but when we say what kinds of women candidates get this much, then we see it's significant to the point of, again, 19 times more than women of color Republican candidates. We also continue to find that race and gender-based hate specifically much more prevalent for the Democratic women of color

candidates than those who are Republican.

I'm sure we can all -- I hope that we all get to discuss what may be some of the reasons behind this finding, regardless, I think it is an important one that we need to recognize and think about it as we go through kind of some of the remedies.

Our fourth and final topline message, we focus on Kamala Harris as the first woman Vice President and the current Presidential candidate, and compare the proportion of hate speech and offensive speech she received during our time of data collection.

We did proportions that Congressional women of color candidates received.

You can see between the two graphs the numbers are surprisingly low for Harris in terms of offensive speech we find that Harris received 40% less than women of color candidates overall, and 45% less than African-American women candidates in particular.

In terms of hate speech, while tweets Harris received included hate speech compared to 2.5% for women of color candidates and 4% for African-American women of color candidates and 1% for Asian-American women candidates.

I kind of gave away the line previously, but we are, as a research team, we are also puzzled by these low counts for Harris, especially if you recall what happened in 2020 when Harris's candidacy was announced as President Biden's running mate.

We had studies demonstrating within the first 10 days the amount of hate tweets which she received were over the roof. So our findings puzzled us within that background, but we think that we have a good list of factors why this may be so.

For example, again, the -- we collected our data between May 20 and August 23, and Harris's candidacy was solidified on July 21, so there is a great chance that what has been happening since then might not be within the same trends of what we captured within that time period.

Similarly, again, we have all lived through the general positive euphoria that surrounded Harris's campaign, the positive energy, the initial days, as well as the fact that Harris herself does not explicitly discuss her gender or race to most audiences, and we think both of these factors may have contributed to the low accounting of the hateful and offensive tweets that our research was able to capture at this time.

>> DHANARAJ THAKUR: Okay, so summarize some of the findings that we did in our research so far. There are many many questions that we still have to address. Maybe that's a follow-on project that Müge is going to talk about later on.

Based on what we have here, and based on the disproportionate impacts that we see on certain groups of candidates, we want to just highlight some positive recommendations that several of us in the room and elsewhere have also pointed to before, but we think [indiscernible].

The main problem is that we have a situation where [indiscernible] women of color that could deter them from running for office or undermine their candidacies in various ways.

We look at social media platform in the area to address solutions and some of the previous -- I think Christina alluded to this a little bit before.

So one thing is around the candidate policies contact policies [indiscernible] have in place, and having clear language and guidance around how those systems address attacks, particularly based on when they are based on race, gender and/or other kinds of identities.

Because the reality is that racism, misogyny, other kinds of ableism and other kinds of forms of discrimination exist in reality. They also exist online. The social media companies, therefore, need to take that into account and realize that not everybody is treated online the same way. Not all political

candidates are treated online in the same way and some of these attacks are more debilitating than others.

Tied to that is this idea of transparency. So we often call for greater transparency. It comes in different ways. If we are talking about requiring companies to put in place better tactics and systems and guidance to address attacks based on identity, we also need greater transparency in how that will work. How effective are these new policies, how effective are these new designs that they put in place.

That requires greater transparency on the part of the companies to explain they did this to address this problem, now how -- to what extent has it actually worked, as opposed to just announcing that they put in place a new system, but we never know outside the company if this is [indiscernible] or not.

And there are many different problems on social media, but I'm talking specifically about the disproportionate facts that we see on women of color politicians.

And so step back, this is a reminder here, this is very important because I didn't allude to this, but just to make sure that we don't assume everyone knows this, but the idea of having equal representation across the various communities in the U.S., in our [indiscernible] are very important for policy-making. You need to have all of us represented when we are talking about development policy, tech policy.

What we have currently is under-representation. There are many reasons for that. Because of the history of racism and so on in the country, but where we see this online and where we can see tactics to prevent that and improve representation, we should. That's why I'm harping on these recommendations here.

Reporting tools. So a broad problem of online attacks against women, various forms of violence and gender-based violence and other forms of violence, reporting tools and reporting systems that are put in place by many different companies, social media platforms, but we also often hear from other researchers, academics and others, that these reporting tools fall short in many ways.

For example, many of the people that -- and by the way, this is something that was mentioned in our previous project when we interviewed women of color candidates. When they do report -- when they do make use of these reporting systems, they are never sure where in the process their report is, where their request is, and what is being done to address the problem that they highlighted.

They often do it as lacking -- as a poor means of accountability for the problems that they faced. So there are many other organizations that point to this as well, but we're just highlighting this here because of the specific problem we're talking about. We need these candidates and their teams and their campaigns need better reporting systems in order to address this problem.

Risk assessments refers to this idea that social media platforms when they are trying to put in place new systems, new forms of, you know, presenting their news feeds or new design features, they should have risk assessments to understand the potential impacts that this could have.

And here we place an emphasis on race and gender. So what kinds of impacts would new systems, new features, new [indiscernible] and so on have on women of color candidates, for example? Could this help them? Or maybe this introduces new risks.

And this is kind of a question that these platforms should ask themselves and should investigate before putting in place new features, because sometimes the features can have disproportionate impacts on people, and here we go back to the problem that we don't want to see this problem being made worse for women of color candidates in particular.

You already heard the notion of accessing data. So I will repeat it here, that we need as

researchers outside of these companies better access to data on the platforms in order to understand the problems themselves.

So in this study we were able to access data from X to understand what was happening, but bear in mind we always get just a snapshot of what was going on. My guess is that the problem is actually much worse than what we can identify, and this is just on this platform. The problem exists on all platforms.

But we have very little insight into what's happening, into who the actors are behind this. We can tell you there's a tweet, but we don't know if this tweet -- we don't know the extent to which this tweet was part of a formulated campaign or part of, like, an influence operation or the extent of which it permeates across different platforms and so on.

These are questions that are important to address, but as researchers we are limited because of the lack of [indiscernible] data. Where that data is sometimes available, it's made available at a very high cost. Like in the case of this study. Or it's done in a way that privileges certain researchers over others, and which essentially then it's the generation of science more broadly available to the public.

So we need better tools for this. Christina mentioned the Digital Services Act. In the U.S. we do not currently have such a mechanism.

These are some of the steps, not all of the steps, that social media platforms could take to better address this problem and help make people understand what part of ways we can [indiscernible] solutions into the problem as well.

And so we encourage companies to do better on these fronts, and there could be others that we could touch on in the discussion. And again, it's not simply focused on X. [Indiscernible] much more, in fact, we heard about [indiscernible].

>> MÜGE FINKEL: So next steps. In addition to taking part in the policy discussions and advocating, as Dhanaraj summarized stronger measures for social media platforms, our goal is to be able to contribute to knowledge production as concerns academic research on online violence against women in politics in and of itself, but not separate it from the larger and more important discussions as part of quality of democracy and quality of democratic governance.

So we do want to understand what happens to online -- to women in politics in online and offline platforms, but we also want to be able to bridge that as causing other important consequences for all of us.

To that effect, our first course of action is we want to produce a more comprehensive research report that expands on the themes that we briefly were able to share with you today. We want to unpack further and qualify the online experiences of women political candidates in this election. We have ample data. We have very smart counterparts and a research team.

What we didn't have is enough time and energy to pull the longer brief together in time for -- to predate the election.

So our first goal is to kind of provide a more comprehensive account of what we know, and we already know that there's some really interesting stuff that we want to pull from that.

A second course of action, we want to explore the effects of these processes on elections and election outcomes, but we want to do that in a comparative perspective, right? We all are following events as they are unfolding in Brazil as concerns social media regulations.

We have election data at national and municipal levels in Türkiye, and now in the United States all of these settings we all know have very different institutional setups and information. So our goal is to kind of engage them in a comparative analyses which will help us to contribute to

discussions on political violence and gender-based violence in politics both online and offline.

So those are going to be our immediate and short- and long-term steps.

>> DHANARAJ THAKUR: Great, good. We're going to wrap up shortly and move to Q&A, but before that, it's important that the large number of people who were behind this project are thanked. Müge already pointed out the research team and at the Politus Project and CDT, but many people here are listed on the slide and listed in the document that also helped support this project.

Many of my colleagues at CDT and also the Ford Institute, and then the project was supported through several grants, one from the Knight Foundation and then the Kabak Grant to the University of Pittsburgh.

>> MÜGE FINKEL: And if I may add one addition, I'm so grateful to have the dean back here. Having her support means a lot to us to kind of help us really further this study, making it an important -- it is important to all of us that we spend our days and nights, but it is so important to have our colleagues to acknowledge that this is an important piece for our school, for new generation of policy-makers, and for all of us to collectively thank.

And I also want to thank Megan who hosted us, got us the space. This is wonderful, and we are grateful, Megan. Thank you.

>> DHANARAJ THAKUR: Great, well, thank you. We can stop the presentation there.

[Applause].

So I just want to move over to Q&A, right?

>> MÜGE FINKEL: Yes, absolutely. We have our in-room colleagues. With apologies to online colleagues, we have actually coffee, tea and some goodies to keep our energy going, so we can do that while we're taking question and answers, and we have our -- one of our team members, Amanda Zaner, is moderating the online forum.

Again, if I could request that for the streaming participants, if you could use the Q&A button to submit your questions, and again, you also have access to the full report online in the two sites that we have included.

So Amanda, maybe you will let us know, but we will start with the conversation in the room.

So please, and thank you so much for joining us today despite the hiccup that we all climbed eight floors. We are grateful. Please?

>> I have a question.

>> MÜGE FINKEL: Please, do you mind introducing yourself, briefly?

>> So my name is Alia, analyst at [indiscernible] I work on free expression and [indiscernible].

Thank you so much for the report.

There is a question and Dr. Finkel we were talking about how some of the research findings related to attacks against [indiscernible] previous research findings a few years ago.

Is there a theory or any findings that would suggest [indiscernible] longer in office, the longer [indiscernible] is there a correlation or some sort of association between a person being in office longer, being in the social environment, and the level of race and sex-based attacks?

>> MÜGE FINKEL: It's an excellent question. So my answer is not going to satisfy you, but I will attempt. One of the things that we are very curious about researching is the effect of incumbency, right?

So we are going to dwell into it, but we do have that data for the U.S. election, and we also have that data in the Turkish election, so we are going to be able to do that, and we are curious about looking into it.

We haven't done that analysis for this study yet, but since you asked for the theory and the hypothesis, I think one of the things that the reference that I mentioned contradicts what you are and I would be thinking as a political scientist, right?

Like at the point of last election when Kamala Harris was named as the running partner, the amount of attacks that were skyrocketing was within the first two weeks, and this is -- this research was published I think the hashtag She Persisted quotes it, but it was in the New York Times and it was in the *Washington Post*. It's just skyrocketed. They looked at the first two weeks.

When we look at the first two weeks now, we don't see that, so it is a curious question in terms of how, why, and I can only suspect that there is, like, a situation of this particular election is I hate to say unique because every election is unique in its own way.

But it is indeed a different one, so we will be looking at -- we are extending our time period to kind of see whether that is indeed a thing, and we are hoping it will include some of the references in the longer report, but for colleagues who have been in this line of research, please.

>> One, I would look back at the institute for strategic dialogue's report in 2020 that was looking at existing members of Congress on Twitter. So there's some relevance to that.

I have a question, though, that relates to the data here, whether and how the deep fake video against Vice President Harris would have been coded or categorized within this data set. It may not have met the violence criteria that you laid out, which I think is valuable for us to geek out on some of these definitions because you've got demean, threaten, ridicule, and insult.

>> MÜGE FINKEL: [Indiscernible].

>> Okay, so this is where sort of when we look at what's the ecosystem or conversation around identity especially around race around believing women as they define themselves, and so if we look at -- I was just trying to pull up these numbers as we were talking. So on July -- the data set ends in July 26, or 23?

>> MÜGE FINKEL: No, it's May 23 to August.

>> Okay, so in the first I think week we were within the campaign with Vice President Harris there was a deep fake, the first major of the election, distorting her campaign video, that was re-tweeted by Elon Musk, that has generated 123 million views.

>> DHANARAJ THAKUR: Right.

>> So where does that factor in? Because if we're looking at sort of distorted imagery, and sadly even the parody label was taken off of that deep fake in order to share it to his 198 million followers.

So how does that factor in?

>> DHANARAJ THAKUR: That's a good example. There's a lot of attention paid to that because of the re-tweet and because it was -- there's debate about whether or not it was a parody, but setting that aside, does -- it wasn't essentially meant to be [indiscernible].

The thing is, as I mentioned at the start, the language model we use looks specifically at text. So what that means then is we would not [indiscernible] on an image, and then the implication of that is we could be in effect under-counting the degree of offensive speech and hate speech [indiscernible] but it also the same analysis applies to Vice President Harris, the tweets that are targeted at Vice President Harris.

So then we would not include these kinds of comments or tweets that [indiscernible] video, right? It could be an under-counting problem. It's hard to say the extent to which that impacts this finding between Vice President Harris and other women of color. Hard to say.

And to go out to other [indiscernible] I think the part of the challenge, the constraints we have

with the data on Vice President Harris in a particular time period, I would guess that if we looked at it now, here in September and October, it might actually go [indiscernible] Vice President Harris.

At the time her campaign had just started. You could argue there was a lot of positive commentary about her nomination, whereas now it's entirely different.

So again, these are questions that would require more analysis and maybe more data as well, yeah.

Yes?

>> In the research, did the candidate have to be tagged or did it get picked up in any type of hate speech or demeaning language? They had to be tagged?

Because my thought would be there's a lot of accounts, especially on X, that use posts or engagement farming, so they don't specifically say anything demeaning or -- like, they allude to it, and then in the quotes, that's where you see, like, user engagement with the hate speech, within the quotes, but nobody is tagged in the post.

>> Right, so here they have to be tagged. Because of the methodology we have to use a kind of standard approach to looking at all the candidates. All comments about all candidates, right?

So it requires that their account, their Twitter account be tagged in the comment. So the tweets that we included in the data set would say there's some level of -- I remember this you're dumb as a rock. It literally said at candidate, you're dumb as a rock.

It would be someone speaking to the candidate. It could include maybe two Twitter accounts in the same message, but if it went up to three or more, we did not include those because now we are referring -- it's hard to determine, like, who are they referring to in that situation.

So what that means then is that if someone said, like, hashtag Kamala Harris or just the word Kamala Harris without her account linked to it and going on to some offensive or whatever it is, that would be not be included because her account was not included.

Again, this could be an under-counting problem, right? But it's because of the [indiscernible] used that we [indiscernible].

>> MÜGE FINKEL: We [indiscernible] on the conservative end to kind of make sure that what we counted as were strictly kind of -- like, we could back it up with the language models that we -- it's on the conservative end for sure.

>> DHANARAJ THAKUR: But also it points at the need for consistency in the method, right? Because a Twitter account, we're clear that that is included and we're clear of who that account belongs to. For the candidates, anyway.

But if you use, like, names or derivation of names or misspelling of names, then you get into more subjective analysis of are they actually talking about this person or is this quote for somebody else? That becomes a more complex exercise.

>> But it's also been a strategy to evade rules by actors.

>> DHANARAJ THAKUR: Completely true. People may use variations.

>> MÜGE FINKEL: Please?

>> I'm [indiscernible] institution.

>> MÜGE FINKEL: Thanks for your question.

>> I'm a [indiscernible].

>> MÜGE FINKEL: Thank you so much for joining.

>> The distribution of [indiscernible] all candidates and describe it a bit? On the maximum end, you may not know necessarily which candidate was mentioned most frequently, but focus on the top number of mentions for a single candidate, which was the minimum? I was curious how that was

skewed?

>> DHANARAJ THAKUR: That I don't know.

>> MÜGE FINKEL: We have the data, but I don't know if any of us can probably recall off the top.

Maybe, Amanda, if you don't mind taking that question on and then we can fire up -- our colleagues at the Politus Project are online, and they offered to participate as we go along, so before you leave, we may have that information, which would be lovely. Thank you. Thank you for your question, and it is important because some of the things that we talked about and I think the hardest part of the team is when we had all the results and we had a five-hour meeting thinking which ones will make it to the four-pager.

And then you see the four-pager became an eight-pager, so there is a lot of interesting questions that we all want to ask, so it will be great to be in the longer version to include this portion.

>> DHANARAJ THAKUR: What we are -- everything we are presenting here, just to be clear, is averages, right? We're talking about women of color, this is the average across all women of color. Some women of color, like AOC, the proportion may be higher for them than others.

We did get into [indiscernible] because we wanted groups and talk about the narrative in that sense, but there's clearly people in research about specific candidates as well.

>> MÜGE FINKEL: And I'm sure you probably are thinking about the Republican women candidates, right? Well, the numbers are small, but they are very vocal, so how that definitely plays into the whole numbers of the tweets that they receive and averages.

>> I have two related questions online. The first, one asks: Given the context in the X platform ownership changing, and given that there are other popular platforms, why did this research center around X as a platform for studying?

>> DHANARAJ THAKUR: So this is a challenge many researchers face when it comes to looking at problems around violence against women online. These are just generally looking at social media which is that X in the past had made it easier for access to data on their platform.

Now this is much less so and has become much more expensive on the platform, but there's also value in looking at X because many politicians and journalists continue to use that platform. So I think in the comment that person put there that they user base is quite low across the U.S., this is true, but for X, 22 or 23%, but as you could tell, almost all the candidates that we looked at have a presence on X.

There are 1100 in the original database, but almost all of them are present there, which says a lot about the value that politicians see in X, and journalists and so on.

So there is value in looking at the conversations there, but they are completely right. In terms of popularity, it's not the most popular platform. Obviously things like TikTok in particular Instagram, Facebook and so on.

Those are much harder to study. Because of the restrictions around access to data, the restrictions around [indiscernible], the Meta recently took down pro-tangle, which is a tool that researchers used for some time to understand conversations of those platforms. That places an additional barrier.

So basically there are frustrations that we face and so we are forced to look at one platform or another. I completely agree with the point. I think TikTok would be a very important one to look at.

Many candidates are using it. Even if you don't look at a candidate amongst themselves, conversations of what candidates are happening at the most level on these platforms.

>> Do a companion study if they wanted.

>> MÜGE FINKEL: We would appreciate it.

>> DHANARAJ THAKUR: Yeah, we will -- since you opened the door there, just to reiterate the point that I think whatever companies are engaged -- Internet [indiscernible] researchers, it's good, but.

>> MÜGE FINKEL: But.

>> DHANARAJ THAKUR: Broad science as a whole needs access for researchers across the board, not just this specific tool.

>> Right, yep.

It's a different in the way of more methodology questions, but it was an interesting comment too in what is the ability to have a full picture of this ecosystem, cross-platform, and we didn't look at the hallmarks necessarily of what is authenticity around some of those accounts too, which would be sort of another -- not to say that this was all fake, but that's where the weaponizing of attacks happens.

Those cross-platform coordination, the narratives that are maligned, but the maligned behavior that we're seeing in the way that social media is organized creates, then, a difficulty in looking at one platform without the context.

I think there's other data here that enhances and shows some similar patterns that we're seeing across platform that really create that kind of ecosystem.

>> MÜGE FINKEL: I mean, it's also interesting to kind of -- we want to ask the question. We are not yet there to answer, but who does the hate -- concentrate on the recipient of the hate, and it's kind of a -- it wasn't an easy task by any means, but it is kind of the first order of question to ask.

The second level is who does the hating. Yes, [indiscernible] through the inquiry we are going to find out some of them are going to be hate fake and created purposefully, but there is another group of population who are very active on Twitter and who purposefully do this kind of tweeting.

So we want to kind of -- and that's why we're very lucky to have our partners at the Politus Project at Koç University. They have some of those questions answered based on the Turkish elections. It's like as a comparative study to kind of pull those tenets and to look at another country's election is our next level.

But we are all curious about who are the haters, right? It's a general profile. It would be interesting and very much telling, especially we want to -- if you want to step away from requiring remedies targeting social media platforms, but what other -- I'm considering what are the long-term remedies in order to kind of create empathy in a group of -- if there's an identified group of hating done, what other long-term solutions that we can all think about.

>> DHANARAJ THAKUR: Yeah, I think -- what you just said reminded me of a point in our previous study on the 2020 election which was that some of the people that were sharing these kinds of abusive language at women of color candidates and misinformation were actually the other candidates themselves. So I think it would be interesting, kind of an additional question to ask when you focus on who is doing this is to look at the extent to which actual candidates are involved in these networks that are part of these campaigns or at a minimum share in similar themes. Because we actually anecdotally saw where -- so there's an example in Texas where there's a candidate of a woman of color in office, a medical doctor, and there were lots of images that were shared to kind of discount or suggest that she was not, that she was pretending to be a medical doctor.

Some of those tweets came from her opponent in the other party, so I suspect -- maybe not to isolate an incident, but it goes to why it's important to understand who [indiscernible].

>> But not only the hate. It's what is the origin of hate, but what is that hate based on. This is where the conspiratorial kind of content around our political discourse, then that is the first trigger point of an emotional reaction that is distorted, and that's why I view these issues as political violence because political violence is orchestrated.

It's not spontaneous, and so one of the first steps is to erode trust in not only individuals -- [frozen] because otherwise we may be looking at a human dimension around what's shifting without understanding what's actually shifting it is based on an entirely unrealistic view of the world and how that, then, is fomenting.

>> Yes, I wanted to follow up on some of the conversation, just ask about the intersection of domestic and international security as well. As we look at data, like yours and other data in the States, how can it be used to assess threats [indiscernible] actual risks or violence.

We see actual violence against women candidates and women in office. How should security agencies be paying attention to your work and work that you're planning to do?

>> MÜGE FINKEL: I hope they are listening. That's the first to start, but Christina, you had mentioned while we were having coffee that through your research you came across a CIA database.

>> Secret service.

>> MÜGE FINKEL: Sorry, secret service database. I got my security agencies confused.

I think the violence piece is important, but how that violence kind of happens is where we are some of it is generated through online, and we know that even though these attacks happen online, they have real-life consequences. People who are attacked are real people.

Their security is threatened. They are sexually threatened. Their families are threatened, so the security threat behind this online kind of facade is absolutely real and physical.

So to that extent, I think there needs to be a lot more attention paid because that's also the disconnect when I started studying this area online violence.

It seems to be kind of like removed from gender-based violence to a certain extent, political violence to another extent.

And I think that's a little bit dangerous because it makes it, this type of violence as if it is less dangerous, which is absolutely not the case.

We have cases in Brazil, all over the world, where women politicians have been killed, have been raped, have been threatened, their families. So I think there has to be. That's why it's great to be partnering with advocacy and organizations on the ground who have a mission statement to kind of carry this to the lawmakers.

We in the academic world want to kind of give you the arms and the weapons to stick with the security theme, but it is -- I think it's like if it just stays on what we found, it doesn't do service. It needs to get to the second level.

Someone needs to hear it, and that's why we have partners who hopefully have the ears of agencies that are going to need to hear it.

>> I think the intelligence agencies in the U.S. and policy-makers have normalized attacks against women, and so we are not seeing the way that this is a core part of distortions, influence operations, and threats to our national security.

And the platforms, unfortunately, do very little. In fact, I -- just on the way over here. So definitely check out the Wired investigation yet around inside two years of turmoil at big tech's anti-terror [indiscernible].

So the effort after Christchurch call to come together and deal with actual terrorist accounts. If

they are not adequately providing protections and data for known terrorist activity online, the idea that we can get them to sophisticate their approach to gender, it's not going to happen with the platforms. This requires policy and leadership.

Period, and hopefully we will be past the election in a way where we can make meaningful efforts around these issues with the leadership that's needed in the United States right now for that.

>> MÜGE FINKEL: We have three minutes left at our disposal, so if there are any questions, please.

>> Very quickly, my name is Matt. I work at [indiscernible] down the street. Quick call out on technology. I want to commend how I think you found a really good middle ground between a conservative approach, making sure that the data wasn't following subjective analysis.

I can see how you can go in so many different directions, using hashtags, the way you were talking earlier about how people intentionally get around the reporting there by using misspellings, et cetera.

But I think that was really [indiscernible] I can't think of a better way to have done it, given all the different constraints, and also [indiscernible] subjective analysis, [indiscernible] 800,000 plus tweets manually.

That's my question. When it came to the close of 900,000 tweets between the Kamala Harris specific and other congressional women of color, how did the language model code? There are so many different ways it could happen. How did the language model define offensive versus hate speech? That's just a really interesting question that I would like to know more about because there is so much [indiscernible].

Speak a little bit about it, that would be great.

>> MÜGE FINKEL: Amanda, can you speak a little bit about it?

>> I can speak just briefly. So it took a lot -- so our coding team at Koç University, they took time to annotate a sample of X amount of tweets between the 300,000, depending on the total data set that we had, and they created [indiscernible] for what words and what language was considered [indiscernible] and whether it counted as profanities.

And thankfully there is some crowdsourcing involved, so that's where the UC Berkeley model and other different-based language models and these other language processing tools really come in handy because they help make the tasks a little easier for us to define language and use our own models to adapt to that.

And so they were able to define the parameters of what was hateful, what was offensive, and what types to then train their models and classify accordingly, and then track it and see how it was reporting.

And if it seemed to the human eye if it seemed off, they could go back and add [indiscernible] makes sense.

>> MÜGE FINKEL: So the only thing to add is that it was never offensive speech versus hate, right? Because hate was a subset of the offensive speech. So you had to be offensive and then kind of it was an identity-based then the model would take you to the hate speech model. So that's why we were struggling in terms of making sure that that was identified.

And again, the way that we identified you could hate a person based on identity is multiple identities were captured by the model, and that took you from the subset of the offensive speech to the hate speech.

So I will take -- I know we are off, but I want to kind of leave you with six numbers that may not come out, because I think it kind of to me as -- again, I'll frame myself as a social scientist.

It kind of, like, brings the whole problem together, and I will leave Dhanaraj to do the closing for us.

So for all these tweets that we have, our models have logged, 16% of them contain offensive speech. Of those 16%, 4.4 were hateful. Of all.

When we looked at the women political candidates, of all the tweets that women political candidates have, it was 18.4% were offensive, 9.2 were hateful.

When we then looked at African-American political candidates, we found that 21% of all of their tweets were offensive, and of that, 19% were hateful. So there is a huge -- you know, we can all go around and talk about politics being kind of a violent space, and at some point it's violent for everyone, but unless we ask these questions of who gets the worst of this violence, so again, those are my six numbers that are kind of making me understand what we're doing and why we are doing it.

So I wanted to kind of leave it with those that make sense, and thank you so much for me, and Dhanaraj, please.

>> DHANARAJ THAKUR: Yeah, no, just thank you. Thanks, everyone, for attending and thanks everyone online and asking these really good questions, and for the discussion.

Thanks to Christina and the dean for joining us and for hosting us as well. And thanks to all the collaborators on the project, the Ford Institute and the Koç University. There are a lot more follow-ups to do, and you alluded to that. We're looking forward to that.

If anyone is interested in continuing these conversations, please reach out to us.

>> MÜGE FINKEL: Please.

>> DHANARAJ THAKUR: Thank you so much.

>> MÜGE FINKEL: Thank you.

[Applause]