

[Captioner standing by for realtime captions] >> This is also a gentle reminder that we are going to be recording the session primarily for notes purposes. Please do keep that in mind. >> Thank you all so much. You have already contributed your thoughts and ideas into these different question pods on the screen. We are about one minute out from the start of the session. I would like you to please go ahead and review what others have written. Thank you for sharing with us where you are connecting from today on the map image right at the top of your screen. It looks like we actually do have the gamut today with our colleagues collect connecting in the room. From Europe to Africa to the United States. You so much for sharing that as well. It seems that we have a few comments on when you act on hate speech. Is a question about whether or not to use the same comments on the counter trust efforts for protection, and protection. And we also had a suggestion about countering hate speech and how it should be ongoing process. We're hoping that today we will be able to cover a little bit of this. And then I was also just like to share into the room the response under the third question about shutting the airways of Internet with love speech. And what the alternative to hate speech is. Again thank you everyone for sharing your thoughts and reflections. I would like you to

just be aware that at all times there would be a chat pod on the bottom right of your screen that will allow you to communicate with our production team. If you do have any questions that arise or any comments that you have the free to chat type them into the chat pod. And then finally we will let to keep you posted that we will have a Q&A session or question-and-answer session so if you do have questions that we are not able to answer in the moment we will save them and adjust them at a later time. So with that I would love to introduce our, I would like to introduce myself, my name is Janine [Indiscernible name]. And my colleague is on the line. We are the production team for today's session. Without further ado I would like to introduce Josh my clutter with the senior advisor for media and Internet freedom at USA printer for democracy, human rights and governance.

Thank you so much for calling into the webinar. This is a part of a series on information disorders and that could be misinformation, disinformation and hate speech. We did a bunch of projects at the DRG center that are based out of a BAA or a broad agency announcement project. We did a series of experimental pilot activities. This one focuses on the spread of each each and disinformation and the role of the social medium media platforms play in amplifying them. We had a previous webinar on one of the pilot it is. We did the one in December and another thing the link for them. You can click on it and listen to it on your free time. This webinar will be recorded so if you like it please share it. And if you want to talk further to any of us were participating please reach out to us. I would like to introduce the other presenters who are going to be on this webinar with me. You can see when from article 19. She is acting executive director for this organization. She is a leading freedom of expression MGO that defined this right for the digital age. I am also joined by Adam Caplan who is a senior media advisor with the office of transition initiative at USA de. He worked to support the media development and endeavors of OT ice programs globally. Why do we care about his speech? I think this is a

really issue and an important space for us to watch because of the online space and social media increasing over civil rights defends and others taking place. It is important to achieve your objectives in an open vibrant state that that open spaces being threatened by a range of bad speech. By bad speech I can speak of a large range from dangerous, weapon eyes, hate speech. A lot of it coming from bad actors. The manipulation of social media is such that toxic messages are algorithmic amplified. I put a definition out there for you and I think it will be able to talk about this idea more but algorithmic amplification threatens to erode the openness of the space. Out directly to the gained harassment and spent hate speech can drown out silence, censored the voices and online participation of some groups. If you look at this definition, the definition is algorithmic amplification is when content becomes popular at the expense of other viewpoints. This is what we're sitting on a lot of the platform. I have also included a link to an article about algorithmic amplification. It is actually if you go on that page there should be a click through for it. I guess is not there just to talk about those groups that are centered in silence because of algorithmic amplification some groups are disproportionately targeted. Women, minorities and the disadvantage. On the top you can see the freedom act from 2017 which an excellent report which talks about how social media is being manipulated by government to target political dissent. Politicians encourage and incentivize followers to harass, "enemies of the state" or flood social media and media with government policies. And a lot of this and that is called censorship through knowledge. Also women on social networks is disproportionately targeted with harassment. There is an international report from last year what they are calling a new frontier of human rights abuses that women are still feeling silence to speak out online with fear of abuse. How bad had the situation gotten there was a wake-up call to the world and to Facebook about his speech online and it is his potential to contribute to reward harms. There is ethnic violence against the Muslim minority in Charlotte. You can see in the pod the article about the country [Indiscernible]. They had been credible reports and credited incitement on Facebook. Since 2014 the violence really began in 2016. Since the beginning of mass Internet use which happened during that period between 2012 and 2014 inflammatory post against [Indiscernible] regularly appeared on Facebook that we have linked to a Reuters investigation that was written in 2018 where they found over 1000 armies post, comics and comments and the Burmese -- for those of you that don't know the community standards are Facebook those are the rules that what and what is not allowed on the platform so those are the rules of the road on Facebook and these are all on prevention. Summer posted by influential figures in the Burmese army with social media following. Some are accused of crimes against humanity and war crimes. For people working in context outside the U.S. and this is us, working with USAID this is a particular concern. This is particularly concerning for social media users in the non-US context. These companies are based in the U.S. and designed to work in the U.S. context but don't consider some of the local contextual issues when they operate outside of the U.S. I just want to point out we have a poll up on the screen and I don't know if anyone answered but the answer is , okay, so the ants answer is the U.S., the U.S. high committee for human rights did call out Facebook

for its role. So just to go back as I was saying Facebook is it really designed for how it operates outside of the U.S. context but context really matters and this is something that I'm going to drill on for little bit. The contextual issues that surfaced in this crisis in Burma included issues such as Facebook was Burma is Donna social media network. You might even hear from some people living in Burma that Facebook is the Internet. And that at the same time at Wednesday's working there they had no Internet in the country. Facebook had difficulty interpreting certain words or terms in the local language for example one particular racial slur in the Burmese language, [Indiscernible] can be highly derogatory and used against Muslims but it can also have a very innocent meaning. It can mean chickpea. So it has the two meanings. The contextual understanding of that really matters. There was software problems in Facebook which meant that many mobile phone users in Burma had difficulties reading Facebook so instructions for how to report worrying material and understanding in this context most people access the Internet through Facebook and on their [Indiscernible]. Four . Facebook had at this time of the 2018 reports it had no Burmese speaking content moderators. According to the Reuters report the company had just one such employee in 2014. A number that increase to four the following year. The company increase the number in the following year and promised to grow that to 100 Burmese speakers. It is hard to imagine how that is going to address the problem given the thousands or millions of posts people have in a day. They had 20+ million users am Burma. After this horrible incident took place Facebook commissioned that human rights impact assessment from a group called BSR and it came out in November of 2018 and it came out with a bunch of recommendations . It is an interesting document and I recommend looking at it because it does relate to a lot of the work that we do. It recommended that Facebook create a human rights policy, it recommended enforcing his own content policies or is community standards. It recommended researching the distribution characteristics of hate speech and act on those findings from that kind of research. It recommended Paco reporting and engagement with local stakeholders such as local NGOs that we work with. It recommended investing in efforts to end increase digital language. It prepared for this related to future development in Burma. This year Burma has general elections. Elections tend to be times with high spikes of disinformation, misinformation and hate speech. What has Facebook done? It is a greasy usage increased usage but imagine how difficult this is for community so large of 20+ million users. Facebook says that it is also looking at Artificial Intelligence to automate some of this process. Another development at Facebook is Facebook Supreme Court. Facebook announced in September that it was going to create something called an independent oversight board that was going to be like a Supreme Court and it will develop something like case law which will make and review decisions about what it takes down or what it leaves online. And it will be a place to appeal moderation decisions. Offer recommendations on how to tackle problem and it content in the future. It was launched last month and I just want to point out that the poll , do you think that Facebook could successfully use automation and a out to filter out his speech around the world.

Okay.

Someone changed their answer [Laughter].

This is just an opinion question. But I can see from a lot of the, that there is an overwhelming lack of faith or lack of confidence in the eyes AI or in Facebook. I would talk a little bit about the Supreme Court because the Supreme Court is also probably going to [Indiscernible] any challenges to any of those decisions that either Facebook content moderators for the Artificial Intelligence that they deploy will make. Regardless for both the Supreme Court or the AI is still going to be the problem of navigating the fine line between free expression and harmful speech, serving evidence arbiter for global platform, harmonizes a range of interests and visions of what freedom of speech looks like. Every country has a very different conception and different ways of regulating speech. There is also lack of clarity. No clarity on the resources, the board or the AI will be used to determine a speech or what specific local knowledge will be used for expertise or how the expert Likert expertise will be chosen. Additionally the focus on protecting freedom of expression by the oversight board that scope might be too narrow as their other such universal rights such as personal, freedom of assembly, freedom to vote that also impacted by content decisions. Also, I want to highlight a really interesting article by human rights watch that this might call for a much more radical re-examination of business models especially social media and advertising ecosystems. According to human rights watch and we will post the article in one moment it says that that model is a significant barrier to addressing digitally media harms. We will post that, it is in the web links. We're starting to see a lot of unfortunate responses that governance are making when he speech spins out of control. Some governments, even Democratic government are such shutting down social media when he speech talk speech spin out of control. In Sri Lanka in 2018 the government sent down social media during the state of emergency during anti-Muslim riots in the [Indiscernible] province. There is even some scholarships that we posted here that says that shutting down social media doesn't reduce violence, it actually feels it. I want to conclude by saying the human rights impact that Facebook commissions actually provides quite an interesting vote for what tech platforms and what local civil society can do. I am excited to turn the mic to Quinn to talk a little bit more about that role and how her organization, article 19, has tried to grapple with it. Civil society groups, media independent groups have a constructive role in guiding social media forms to understand the context for hate speech and misinformation in a non-US context and to use that information in waste to improve how social media platforms feed their algorithms and moderate content. I will turn it over to Quinn.

Hello everyone. I am Quinn and thank you Josh for the introduction and [Indiscernible]. Do have a quick poll for people to take right now?

Next one. The one about the countries. Yes. So we have question here on the whole for those that are not aware [Indiscernible]. [Indiscernible - echo]. >> We have a few people here. I think the wisdom of the crowd is coming through. Event, in terms of looking at the impact of data and data analytics at and the answer is all of the above. [Indiscernible]. [muffled]. We were fortunate enough [

muffled]. How can you identify and counter the speech and online in Kenya. [muffled] actors know that the information environment can be manipulated to their own end by exportation X [muffled] in which information is shared by these platforms. What you get to see [muffled]. Why , during the last election there was a significant amount of outside interest in the political context audio [Indiscernible - audio cutting out] through the propagation of armies in the independent election. There have been signs that these kinds of networks were still active and we are still being used for certain political ends. This project is looking at understanding the information on the online media [muffled] to identify what is the different information identifying desk looking at identifying methodologies or unifying or classifying this type of hate speech particularly happening in Kenya. The reason for that is because these algorithms often rely on what are called signals of authority in terms of deciding which content should be algorithmically amplified in terms of what people get to see. These algorithms have incredibly blind to local content and their often making poor decisions in terms of the kind of content that is being shared based on the lack of that context knowledge. We look at the quality of information and the signals to drive automatic decision-making in online [muffled] the way that we decided to do look at this particularly Yen and Kenyon

-- the census was looking at different tribal and ethnic questions in the country that would be tied to political participation in the ability of people to get access to services going forward. We look at this particular system is at a time, because there was so many political implications, August 2019 without there was a fair likelihood that there would be significant amount of disinformation a hate speech against tribal identity and ethnicity, to test our method of trying to identify the speech in a certain different environment. [muffled] so a key component here was using this event to actually see if we could develop a model of information to the local context and categorize the speech in the particular period. Some of the key findings from our research that we did was a number of steps. The first was research, what was happening during the Kenyon [Indiscernible]. One was looking at a pure algorithmic approach by which we had an automated system that was trying to detect speech during this particular event. Had a blended approach which combined local knowledge and local expertise with an algorithmic effect especially of hate speech and content that was being shared on Twitter. Some the key findings from this particular [Indiscernible] was during the census pair there was significant hate speech that was amplified again by digital explicated needs means. The algorithm did not understand context

multiple [muffled] therefore one of the things that was completely key was actually understanding that context and how to build context into the way that the systems are evaluating that in the way that we went about doing that was actually looking at developing what are called indicators They're looking at standardized indicators in terms of the way that moderators could evaluate individual pieces of content , categorize them in terms of the kind of content that was there so that people could say that this content was had racial overtones or disc content was attacking people on S&S of the or it was anti-LGBT. We try to come up with a standard set of indicators so we

could see better content [muffled] at scale. The goal is to try and improve the overall, though way the algorithm treats this content so it isn't a one on one trying to take things down one at a time or [Indiscernible] but that is being done on a broader more comprehensive approach. One of the pieces of research that we employed in this was public network analysis. This was done by a company by [Indiscernible name] which actually looks at methodology was close to basically taking every individual piece of content that is shared and tried to categorize and class dump based on where the acidity between the different users on Twitter are to try to create classifications and groups of individuals are based on the kind of content that we share. And then using that. We're trying to determine what the connections were between the different user groups based on the kind of content that they were sharing. If you look at the slide here there are two distinct comp logs. This is looking at the different kinds of content user and the content being sharing. On the green side there are the people to share a lot of content based, talked about different Kenyan politics. Interestingly the blue is actually the international [muffled] for people outside of Kenya. You can see that there is actually a lot of it does on the green side of this map indicating that there were quite a lot of international political interest and [muffled]. Based on that we also looked at what is the level of automated activity. What is the level of the Army of information that operate in a coordinated way automatically. There was a significant level of potentially automated activity sharing content specifically around the Kenyan context. And also specifically around non-[Indiscernible] group. It was basically botched goods. [Indiscernible] through automating and propagating hate speech or disinformation in terms of the content that we see being shared in the automated [Indiscernible]. I think we have another, another [Indiscernible] here. What type of for an engagement was there with the Kenyan engagement. Were people just following, active, what kind of content with a sharing. What would you expect from that kind of [Indiscernible] that people were having? So we have one [Indiscernible]. There was a lot of different disinformation being shared [Indiscernible] but one of the things that we also discovered in terms of looking at our research was there were a number of suspicious names that came up in terms of suspicious locations that information was originating from. We actually, three of the domains that we analyzed in the research were actually domains that were affiliated with the Kremlin or focused on Russian politics. Not necessarily the kind of domains you would think would necessarily be focusing on a very domestic Kenyan . That was Russian insiders in particular. There also seems to be a number of domains that had been suspended by the provider because they were trying to mimic popular news sites existing legitimate new sites. There was one that was mimicking CNN and the size in particular were trying to propagate this information in terms of saying things around the Kenyan to incite a speech. Often times we were propagating hate speech themselves. This content would often be attesting one political party and it was clear when we analyze this that it was a part of a broader network to spread this information or to mimic local international media to get more legitimate need for their political point of view. In some cases these domains were actually suspended by the hosting providers because they

were illegitimate. This shows the extent to which there is increasingly foreign political interest in these events particularly in the global network analysis that we did. The next phase of the work after we look at [muffled] what do we actually do about it? The things that we did was the [muffled] identifying misinformation, [Indiscernible] using indicators and content the fires -- and that we worked with local civil society and media in Kenya that was a co-working pop up. That pop up based on the network analysis that was done brought together journalists, human rights advocates, researchers and fact checkers representing over 20 civil society organizations and journalists to try to codesign standards or address this information on social media platforms. We asked them the question of what kind of content are you seeing and how are you going to classify that content according to your own local knowledge of what would be considered hate speech or device of content or racist rhetoric for things that will lead to incitement to violence. Though standers we asked the question where the social media global standards for content moderation fail, how can civil society that is connected only negatively -- we are looking to give civil society and fact checkers greater [muffled] and then advocating through the social media platforms are proving the way that they deal with content moderation and the divisive hatespeech in particular. That was one of the key issues that we were looking for [muffled] providing people with a way to develop standards, [muffled] and be directed at the social media platforms themselves by helping change the way they address these issues. To give you a quick overview of what this kind of standards looked like what we asked people to do is take a sample piece of content from twitter. [muffled]. We identified a number of participants in the pop up including understanding unfair content takedown so things that have been taken down that were legitimate and should have stayed up. Understanding the role of automation in that. Looking at more effective flagging systems for the harmful content and the content that would rise to the level of incitement and the international standard. Monitor and manage in [muffled]. From those needs we looked at the content that was being shared and developed localized standards in terms of developing a common understanding and common definition of discrimination and information and how they manifest in Kenya. Developing clear and locally formed indicators were content would fall under these standardized definitions. And then using the localized standards to document social media [Indiscernible] they do not align with the Kenya content. When things that are slipping through the net should be caught. [muffled]. You take a particular piece of content. In this was [muffled] and the response was [muffled]. Take that content and then you ask people at the national level in Kenya to identify what kind of speech is reflected in that context to try to identify the way and standard that should be applied to the evaluation. This one identified filtered on ethnicity, [muffled]. The reason to develop the standers is if we go back to the platform you can use this new standard in evaluating the content to improve the way that you your automated systems are looking at tackling the hate speech standard on it a one-to-one [muffled]. The purpose was the civil society and that journalists were trained to identify communicators, to better counter hate speech and to improve their advocacy of the platforms themselves to improve better decision-making in their country. This was

a really useful pilot project that we were working on with PSA help. We learned a lot of ways of approaching this kind of content in terms of building a stronger and more coordinated response that isn't based just on fact checkers but is actually looking at trying to influence the fundamental system, electronic system for information sharing behind it. Thank you and I really look forward to any questions that you have.

Great. Thank you. So thank you Quinn. I am Adam and work for TI. I will be speaking through three examples quickly. I understand we are running short on time so I will last through this pretty quick. Feel free to post your questions if stuff goes by too quickly. The three manifestations that we're looking at is Miramar, [Indiscernible] in Ethiopia. This represents three different phases of programming for OTI. For more on that you can reach out to me and I can connect you up with mission people who can tell you where it is going from here. Our bargaining program is right in the middle of, we're doing that in close collaboration with the mission and other actors on the ground. It is an interesting case of localized platforms which I will do in a moment. In Ethiopia the work is just beginning. We don't really have a full sense of all of his involvement. We're trying to figure out what our mandible interest our. Next up. From these three examples commonalities, the principal commonality is hate speech online predates hate speech off-line and those on long-standing grievances. These are deep, see the conflicts that are not going to be addressed simply with the application of technology. The expectation that algorithms are going to save us is probably not a safe one. It also exists in non-online forms and what we have seen at the anonymity and the vital capacity of online has actually deeply exacerbated the problem. Stuff has gone, the volume has increased and the offensiveness has gotten worse. Much of this hate speech present superficially as organic. Emerging from a lot of different sources. This is not what we understand to be the main driver. It is Astroturf, but they grasp. These are concerted efforts by political actors seeking to present their work as organic when in fact it is deeply focused and pushed in a specific political effort. Ethiopia is the platform of choice which presents challenges and opportunities. In Bosnia in addition to Facebook where seeing a strong presence of hate speech on local platforms. Will talk about that in a minute. Next slide. Differences. With respect to our analysis the EC if the open example of, sorry, let me start with Burma. As Jost delved into Burmese example has been well studied. It did include considered efforts by the time a dark, and focused effort to push this stuff. They started early and were organized. They were systemically and they were successful. They were able to create a space where violence against [Indiscernible] was a reasonable thing for both individuals and the state to prosecute. This is slightly different. As a result you are saying a wholesale genocide and genocide in Burma. In Ethiopia with one exception the Ethiopia violence has been located on Ethiopian campuses. This would tend to make sense as the predominant digital users are youth and the largest population. There are examples of other [Indiscernible] namely a gentleman named Mohammed using twitter to actively incite violence of the events in 2013. However in Bosnia what we're saying is the plethora of hate speech across all platforms and the levels of violence relatively low but quite specific making it very difficult to make a

connection between actual violence and content online or violence sparking offensive content in both directions. Facebook has been publicly shamed as Jost mentioned into their Burmese case. However it is unclear to what degree another example might be used. The Burmese interaction with Facebook to create a comparable level of concern I the company in this issue. What we have seen is if it isn't widespread violence it is wholesale contravention of the communities and Facebook are not documented that Facebook is not going to be interested particularly in countries without high economic impact for the company. Next slide. This is what we understand and how we go about the approach. Monitoring has its place. It is important and it is significant but not as we understand as an avenue to mitigating the overall volume of hate speech in this specific place. This is beyond what we believe of OTI or any manageable objectives. You are not able to turn the stuff off. It is too deep seated and will continue. That said media contains campaigns and literacy effort can and have been demonstrated to change behavior. Scale is a huge problem. There was a question earlier about why don't we flood the zone with love speech. I have yet to see a USA mission offer the resources or any donor offer the resources necessary to grade up comparable level of content the flood the design. It would take a huge effort. Not that it is not possible but it is not necessarily something that omission will be compelled to think is a readable source of resources. That said what we have seen is with rigor, diligence and publicity Facebook has been able to be shamed into acting. We can talk a little bit about what that action actually looks like and whether it might be a reasonable approach but it has happened. What we have seen is with monitoring and a combination of media literacy and mediation training local stakeholders can be engaged and inoculated so that violence can be mitigated in their environment when hate speech seeks to incite it. These are examples from former where we interacted with local religious leaders and political leaders and demonstrate to them that this next round of hate speech around a specific event that we can forecast an election, a census for example, is likely to spark violence. In to be ready and we have seen officials be able to step into the middle of that. Maybe I will leave it there and we can go to questions to keep it tight. [Laughter].

That was tight and actually incredibly informative and helpful. I want to thank our presenters and I want to thank our participants. This is that time what would give it to you. We have been collecting your questions throughout the duration of the session and you will see the questions that we are answering posted on the chat beneath the flight. If we haven't had an opportunity to answer your questions we can address them via email.

The first question that is out there is the one from, here do we defined real-life impact of hate speech. I will start off with a really easy answer and leave it up to Adam and when. What we're really talking about here is violence. We are seeing increasingly and there is no definitive proof that there is some leakage between hate speech and hate speech online and violence that we're saying there I want to recall in the pre-social media era in Rwanda there so much hate speech that it was the first time that you saw a media outlet that was brought

before the international criminal court, the [Indiscernible name] which targeted individuals and call them derogatory names, cockroaches and pointed out the addresses of individuals. That is what we're talking about we talk about real-life impact of hate speech. Adam or Quinn do you want to chime in and offer you wisdom?

I think for me the point, thank you Blair for that question, it hinges on cards that . This is an incredibly different thing to do. I have had issues with the expectations that we can. The thing that I suggest to OTI programs and others that I talk about this to you do not need a causal definition. If you can demonstrate that there is a correlation or these demonstrations are correlated with violence they will be continue to be correlated with violence. That is enough. Don't expect a perfect causality to be a necessary condition. Act on correlation.

The fact that you and high commissions to human rights actually call Facebook out which is a unique situation. I think it illustrates that there is heightening concerns. Quinn? Did you want to add to that? Okay.

Sorry. [Indiscernible - audio cutting out].
[muffled]

their very clear standards under which each can be categorized in international standards and general comments on [muffled]. If that is the real-life impact. One of the things that is becoming more interesting to look at in terms of hate speech is actually what we're starting to define as violence. In particular hate speak targeting women or minorities particularly LGBT a lot of people have made the argument that it is vitriol and the psychological pressure that is directed against these individuals are the fact that they are actually clear violence in and of themselves. That is when [muffled] what kind of violence are we trying to present. It is a specific manifestation or drive people off of the space or keep people away from the boating base because [muffled].

Great. Another question. From Rebecca. Which identity groups or targeted by hate speech and Kenya were people with disabilities targeted during the census.

I will turn that over to Quinn. If you point out if you look at the slide that was one of the categories that was listed in terms of what our civil society and media partners were looking at. Quinn, do you want to respond to that?

I can go back to that slide, to answer your first direct question. Disability was one of the categories of hate speech. There was a large volume of hate speech directed at people for gender or and . They think the biggest volume that you saw were based on ethnic and tribal groups. [muffled] that was the majority of the [muffled]. There is a vitriolic subset that was focused around
the [Indiscernible] group.

This is Lisa. Can you give an example of the contextually specific forms of hate speech that were not been caught by existing algorithms

and how algorithms could be altered to identify them. How did the platforms respond in which social media platforms did you target. This is for Quinn. I think you addressed this before but maybe holistically you could address it.

I think one of the ways that the slang terms that people use are not pick up by social content moderation is referring to people as different kinds of animals in a local dialect or insects [muffled]. That is a classic example of content that the systems don't pick up or [muffled]. There is an abysmal lack of content moderation that happens on the platform. Our work was specifically looking at twitter and Facebook predominantly because in East Africa and [Indiscernible] and companies are opening offices in Kenya to serve as a regional approach which is actually [Indiscernible] billion-dollar companies but they don't have a lot of [Indiscernible] in different markets around the world.

I think there are probably some really interesting examples in your activity which were around who said the thing especially in cases of post around of the PDQ LGBT issues. It seems to be coming from the foreign political accounts that we identified in network analysis. A lot of it was the content that you looked at was domestic and it was coming from the domestic [Indiscernible]. So the majority of the divisive content ultimately broke down on basically people that [muffled] and then has some of the content that was being amplified and shared by some of the actors that were trying to [muffled] and those are some of the issues from the Russian [muffled].

To add, something I forgot to mention with respect to the Bosnia context what we were specifically looking at was hate speech on local platforms, local news aggregators and comments within the common strains of those aggregators. What we found at the local level we had significantly more traction with local actors and with the advertisers on these local platforms and we were able to significantly moderate content. The way this relates to this question however is once we were able to identify terms that were offensive new terms were created. These things do not a static. If people want to express hate they will find a new way to express hate. So expecting that we're going to be able to perfectly to an algorithm that will be useful for ever is not a reasonable assumption. This requires human intervention constantly.

Thank you Adam.

Absolutely 100 %.

Absolutely. Quinn, Adam thank you so much and participants who joined I want to thank you. There were a lot of questions which unfortunately due to time constraints will not get to but you can email those to me at

jaymac@usaid.gov. Thank you again for attending wherever you are

[Event Concluded]

Again everyone thank you everyone for your participation. You will see a few poll questions on the screen for you to please go ahead and responded to and furthermore at the bottom of your screen you will see that we have listed all of the different web links that were presented in this session. If you would like to access any of those web links click the name of the article or the link you would like to access and then click browse to be taken to the new page. Thank you so much for your attendance and we hope that you have a great rest of your day or evening wherever you are.

[Event Concluded]