UNITED STATES COMMISSION ON
INTERNATIONAL RELIGIOUS FREEDOM

HEARING ON COMBATTING ONLINE HATE SPEECH AND DISINFORMATION TARGETING RELIGIOUS COMMUNITIES

Wednesday, October 21, 2020

10:00 a.m.

Virtual Hearing

P A R T I C I P A N T S

COMMISSIONERS PRESENT:

Gayle Manchin, Chair
Tony Perkins, Vice Chair
Anurima Bhargava, Vice Chair
Gary L. Bauer
Frederick A. Davie

C O N T E N T S

- - -

P R O C E E D I N G S

CHAIR MANCHIN: Good morning for our visitors and guests and members of USCIRF and commissioners, and to our dear friends and colleagues from across the pond, good afternoon, and welcome for attending the U.S. Commission on International Religious Freedom's hearing on "Online Hate Speech and Disinformation."

I am Gayle Manchin, serving as chair of USCIRF, and I would like to personally thank, and on behalf of the commissioners and USCIRF, our distinguished witnesses for joining us today to offer their expertise and also recommendations.

The U.S. Commission on International Religious Freedom, or USCIRF, is an independent, bipartisan U.S. government commission created by the 1998 International Religious Freedom Act, or IRFA.

The Commission monitors the universal right to freedom of religion or belief abroad using international standards as our policy, and we make our policy recommendations to Congress, the

President, and the Secretary of State. Today, USCIRF uses its statutory authority under IRFA to convene this virtual hearing.

During the past two decades, Facebook, Twitter and other social media platforms have emerged as an invaluable tool for connecting people around the world. However, we all now know how social media sites can be easily used to amplify hate speech and disinformation about religious communities and mobilize real world violence, discrimination and hatred.

Vile rumors or conspiracy theories that might have previously spread through a village or town can now be shared online and make it around the world before being debunked.

The algorithms that power platforms like Facebook and Twitter reward extremist discourse by incentivizing users to post provocative content that will receive attention through likes and reshares.

There is no definition under international human rights law of the colloquial terms "hate

speech" or "disinformation," but hate speech is typically understood to mean speech that prejudices a specific group.

International human rights standards require states to prohibit the most severe forms of hate speech, specifically any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility, or violence.

Disinformation, sometimes referred to as fake news or propaganda, means false, inaccurate or misleading information intended to cause harm. Disinformation and hate speech are interrelated and can overlap. To use an analogy: hate speech loads the gun, but disinformation pulls the trigger that transforms digital hate into real world violence.

Social media companies ban certain types of hate speech and disinformation from their platform. Twitter's hateful content policy, for instance, bans the promotion of violence, threats and harassment against people in religious groups and the dehumanization of people based on their

religion.

Facebook similarly prohibits "attacks" based on religious affiliation in its Community Standards, defining attacks as violent or dehumanizing speech, harmful stereotypes, statements of inferiority, or calls for exclusion or segregation.

In a welcomed move last week, Facebook announced that it would ban as hate speech content that "denies or distorts" the Holocaust. This policy change was in response to the global increase in antisemitic incidents.

Twitter and Facebook also ban, flag, or counter certain types of disinformation, but neither have blanket policies against misleading information on their platforms. Despite these policies, the volume of hate and disinformation being shared online is astonishing. Facebook, for instance, removes--removes three million pieces of hate speech a month, which means more than 4,000 an hour.

Today, we will explore the complex role

that social media has played in fomenting conflict--as well as hate, violence and discrimination--toward religious communities. We will consider how the United States government and social media companies can better contribute to combating the digital spread of disinformation and hate speech.

I now turn to my colleague, Vice Chair Tony Perkins, to further discuss content moderation and highlight some contexts of grave concern to USCIRF.

Thank you, Tony.

VICE CHAIR PERKINS: Thank you very much, Chair Manchin. I would like to join in welcoming all of you to today's hearing.

To enforce community standards that Chair Manchin outlined, social media companies rely on a combination of artificial intelligence, or what we call AI, and human analysts to wade through content to identify and remove inciteful statements or disinformation prohibited by community standards.

Notably, disinformation is not always removed, but instead may be downgraded or

corrected. Even as we wrestle with establishing a clear objective definition, as Chair Manchin made reference to, there is recognition that identifying hate speech involves a great deal of nuance, context and linguistic expertise. Relying on machines to recognize it remains a challenge.

Social media companies have also struggled with having enough content moderators who speak local dialects and have the expertise needed to proactively identify hate speech around the globe.

Now a recent audit of Facebook noted that their content moderation efforts remained, quote, "too reactive and piecemeal," end-quote. And harmful content continues to spill through the cracks.

Critics of current content moderation efforts have urged social media companies to move away from their current "whack-a-mole" approach and develop early warning policies that preventively flag situations where violence and atrocities are likely to occur.

There is also concern that the

overreliance on content removal can lead to online censorship that restricts fundamental freedoms and actually drives extremist views. That's a very, very tightrope that they are walking.

Around the globe the spread of false or misleading information through social media is causing real harm by catalyzing violence and brutality toward religious communities. For example, in India where Facebook has more users than any other country globally, its WhatsApp platform is used to spread hate speech and false information against religious minorities.

Now, in Pakistan, Facebook is used to target, frame, and accuse individuals of blasphemy, leading to detention, disappearances, extrajudicial killings, mob gatherings and even public lynchings.

Now government-sponsored hate speech and disinformation is particularly perilous as it fosters a dangerous culture of hate and religious intolerance where both online and offline abuses are condoned. The Russian Federation employs a very sophisticated disinformation network that

targets religious minorities with sensational allegations designed to create fear and animosity against them.

Jehovah's Witnesses are depicted in state media as dangerous and subversive, often with ties to western interests.

Government news programs accuse religious minorities of ties to revolutionaries in neighboring Ukraine and depict peaceful Muslim groups as terrorists. For example, in May 2019, a close advisor to President Putin published an op-ed claiming Americans and Israelis were plotting with Ukraine's President Zelensky to deport ethnic Russians from Eastern Ukraine and replace them with Jews.

In Iran, the government uses social media to spread anti-Baha'i propaganda while systematically harassing and jailing members of that community on the basis of their faith. Iran's Supreme Leader, Ayatollah Khamenei, regularly tweets antisemitic vitriol from his official Twitter account, while at the same time restricting

Twitter access for his own citizens.

Iran's security apparatus regularly uses Instagram and Telegram to threaten members of Iran's Sufi community and followers of spiritualist Mohammed Ali Taheri with physical harm.

I will now turn to Vice Chair Bhargava to further explain what is being done by social media companies in response to online hate speech.

VICE CHAIR BHARGAVA: Thank you very much, Vice Chair Perkins.

The proliferation of hate and separately disinformation on social media has been a central concern in how religious communities have been targeted in so many countries that we at USCIRF engage and monitor.

Burma is one prominent and horrific example. On August 25th of 2017, the Burmese military launched a genocidal campaign against the Rohingya people, who are predominantly Muslim. Burmese military units have been involved in indiscriminate killings of civilians, mass rape, and arbitrary detentions and arrests.

More than 740,000 Rohingya refugees fled to camps in Bangladesh while another 120,000 are displaced internally.

In Burma, Facebook is preinstalled on many mobile phones, which has allowed users to access the Internet. It has also led to a misperception that Facebook is the Internet, and this has enabled hate and disinformation to go viral rapidly. The United Nations Fact-Finding Mission for Myanmar concluded that Facebook enabled Buddhist nationalist and military officials to spread "hateful and divisive rhetoric" targeting the Rohingya.

In August of 2018, Facebook, after significant pressure, blocked and removed the accounts of 20 Burmese individuals and organizations, including General Min Aung Hlaing and Buddhist monk U Wirathu. Despite these efforts, Facebook admitted in 2019 that it, quote, "can and should do more," end-quote, specifically noting its failure to prevent the platform's use to "foment division and incite offline violence."

During the first half of 2020, Facebook claimed it took action against more than 330,000 pieces of content in Burma. Yet reports continue that groups promoting hate and intolerance continue to use the platform and the Burmese military reopened a Facebook page in June.

USCIRF commends social media companies for increasingly taking down content that contains hate speech or disinformation. USCIRF urges those companies to take steps, however, to support measures of accountability and to allow information to be used in investigations.

Many social media companies rely on artificial intelligence, as Vice Chair Perkins talked about, to remove content, and in doing so, the companies should make sure that any such content which could be important evidence of violent crimes or hate can be used in efforts to bring perpetrators to justice.

Facebook recently rejected a request by The Gambia to provide information relevant to the pending case against Burma for genocidal charges at

the International Court of Justice, or ICJ.

Facebook has asserted that it has shared evidence with the U.N. investigatory mechanism for use in potential criminal prosecutions, but the head of ICJ has reiterated that Facebook has yet to release evidence that relates and underscores the seriousness of those crimes.

Facebook must release evidence that could be used to hold responsible Burmese officials who committed alleged genocide against the Rohingya. Facebook's inaction is not only a disservice to Rohingya victims demanding justice; it also fosters wider impunity. Those who spew hate online, whether governments or non-state actors, may think twice if they know that social media companies are prepared to share their statements for use in future criminal proceedings.

Thank you, and I look forward to hearing our witnesses' views on these global and concerning issues.

I will now turn the floor back to Chair Manchin.

CHAIR MANCHIN: Thank you so much, Commissioner Bhargava and Vice Chair Tony Perkins.

As we all know and certainly aware from the comments that have been made this morning, the issues are large and looming, and we are so honored to have the speakers with us today that bring with them a great amount of expertise and some recommendations of how we can move forward in the future.

So we now look forward to hearing from them. We have four speakers, and I'm going to generally introduce them, but please know that on the website are very complete bios of all of our distinguished guests.

And our first panel or our panel will be David Kaye, who is a clinical professor of law at the University of California, Irvine. From 2014 to 2020, he served as the United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.

He is also the author of Speech Police: The Global Struggle to Govern the Internet, in

2019.

So, David, we welcome you and look forward to your comments.

MR. KAYE: Thank you, Chair Manchin, and thanks to the entire Commission for the opportunity to join you this morning for me and for you, I know not for the United Kingdom, for this hearing today.

In my remarks, I would like to highlight perhaps four specific areas, and I will try to keep it brief, in part because the comments that you, Chair Manchin, and your vice chairs and commissioners, have already made really drill down and highlighted the many issues involved with respect to hate speech online and religious freedom online.

So what I would like to do is simply highlight four different areas, and I'll start with a general question of the sources of law, the sources of decision-making that should be at issue in this area. I'll say a few words about the companies, I'll say a few words about governments, and then I'll say something a little bit more

specific about the U.S. government. And I'll try to do this all in just a few moments, normally devoting an entire course to these kinds of issues, but I know that you've already, you've already taken the course and are practitioners in it.

So the first point that I would like to make is, is actually a point that is drawn from the bipartisan history of the U.S. commitment to international human rights law. The United States ratified the International Covenant on Civil and Political Rights in 1992, upon the strong recommendation of President George H.W. Bush.

And the ICCPR, which is drawn from the Universal Declaration of Human Rights, protects in Article 18 everyone's right to thought, conscience and religion--the freedom of those principles.

In Article 19, it promotes and protects everyone's right to seek, receive and impart information and ideas of all kinds, regardless of frontiers--thus, it's an international right--and through any media. And those two rights together, along with the permissible limitations that you

mentioned already, Chair, in the context of Article 20, with respect to incitement that's based on national, racial or religious hatred, and also the restrictions on expression that are permissible under Article 19, that must be necessary and proportionate, those principles, those principles of international human rights law, I think are the essential principles for us to be thinking about the rules that these multinational companies, companies that are not simply bound or enjoy the protections of the First Amendment in the United States, but also operate in countries around the world where their users and the people in the public who are impacted by them also enjoy these fundamental human rights.

So the first point I want to make is that there is a source of law and a source of decision-making that can be relevant to us.

The second point I want to make about companies, so to focus on companies in particular, is that while we often talk about the content standards, that is the hate speech rules or the

religious freedom rules that they should be implementing, which I fully subscribe to, I think we should also be thinking about rules of transparency. The companies, as I think some of the comments have already indicated, are rather opaque both in their adoption of rules and their enforcement of those rules.

And one of the problems that we have, both as researchers and you as commissioners and as advisors to legislators and to the executive branch, to policymakers, a very significant difficulty we have is understanding not just the rules but how the rules are made and how they are implemented.

And I would strongly encourage the conversation, if not today, over time, to focus on that particular issue of transparency and the importance of companies being transparent about their work. Otherwise, the conversation is entirely asymmetrical where the companies have all the information and we only have sort of the shadows of the information.

The next point I want to make is about government responsibility. Now I think the commissioners have already indicated the problem and the problems that we see around the world with respect to governments inciting violence, governments promoting discrimination, and we have seen this in many parts of the world.

We've seen not only the spread of anti-blasphemy laws, the spread of hate speech, anti-hate speech laws, and the spread of disinformation or fake news laws, which at some level clearly have a basis in law and policy.

However, around the world, those rules are very often used as a tool against minorities, whether they're ethnic minorities or religious minorities or others, and I think one of the focal points, certainly for the Commission--I think historically the Commission has paid attention to these kinds of issues--but because these laws are now being applied with I would say extra-vigor with respect to online services, I think the issues that we need to be focusing on, to a certain extent are,

is government regulation serving to promote religious freedom or is it serving to undermine religious freedom?

Very often governments will make demands of the companies to either take down content or take action against users or accounts that are deeply problematic, and I think part of the effort moving forward should be not just focusing on the companies but also focus on the governments that create a hostile environment to religious freedom and freedom of expression.

And then the final point I'll make, and I really beg your forgiveness if I've gone over time, is to say a word about the U.S. government. So the United States government, in part because of our historic protection of both religious freedom and religious tolerance and our historic protection of the freedom of expression, both of which are found in the First Amendment, the United States has a strong role to play internationally in promoting these values--in promoting them in governments around the world, in promoting them in the United

Nations and in promoting them with respect to companies.

However, and there is a however here, to the extent that the United States removes itself from the institutions of international governance, such as removing itself from the U.N. Human Rights Council, or to the extent that the United States seems to privilege one set of rights over other rights, even though--and I want to emphasize this point--human rights are interdependent.

Freedom of expression depends on freedom of religion, depends on freedom of assembly, depends on nondiscrimination. All of these rights are connected to one another, and so as we move forward, I think it's important for the United States to reengage with the institutions of international law to regain a credibility that frankly has been lost over several years, not just the last few years, but over several years, to reengage domestically by making the human rights conversation not only a conversation about what they're doing over there, but about what we do

here, and make it a conversation about how we interact and bring international human rights law, which applies to governments, also to the table when it comes to companies.

So with that, again, thank you very much for your time. I hope that my Internet hasn't been too unstable for this. I know I've cut in and out here and there, but again thank you very much for the opportunity.

CHAIR MANCHIN: Thank you very much, David. Yes, we heard you loud and clear and certainly appreciate your work and the information you've shared this morning, and after our speakers we will come back with some questions.

So now we move to Susan Benesch, who is the Executive Director of the Dangerous Speech Project. She serves as the Faculty Associate of the Berkman Klein Center for Internet and Society at Harvard University and teaches human rights at American University's School of International Service.

Welcome, Susan, and thank you for being

here.

MS. BENESCH: Thank you very much, Chair Manchin and all of the commissioners for inviting me to offer you a few ideas alongside my wonderful fellow panelists.

Since Vice Chair Perkins and Commissioner Bhargava have already described the woefully numerous cases of hateful content and disinformation targeting religious communities around the world, I will instead describe for you some of the patterns that my colleagues and I have observed in that kind of content, and then I'll move as quickly as possible in my brief remarks to some ideas for how I think this problem can be diminished.

Chair Manchin mentioned the Dangerous Speech Project. That is the research group that I founded after observing, about a decade ago, strong and striking similarities in the kind of rhetoric that malevolent leaders use to turn one group of people violently against another one.

It is extraordinary how one can observe

patterns in this kind of content across, across countries, across cultures, languages, and of course across religions and religious communities. This kind of rhetoric, which I named "dangerous speech" for its capacity to inspire violence, has been used all too effectively in again a great variety of countries and languages against myriad groups.

It is language that may express hatred, but--this is really a key point--it is defined at least as much by fear as by hatred since what is most powerful in it is that it is designed to generate violent fear of other people since violent fear in turn makes violent reaction seem defensive and often morally justified.

This language is again all too common around the world at present, and it often targets religious communities as you have already heard.

Let me just note a few striking trends in it that we have observed in numerous countries and contexts online in recent times.

First of all, in some cases, this rhetoric

suggests that there is something inherently wrong with the religion, most often Islam, as you know, and therefore with its followers. This familiar language has spiked just in the past few days in the aftermath of the appalling murder and decapitation last Friday of a French high school teacher, Samuel Paty.

We have also seen a closely related tendency to conflate criticism of a religion, on the one hand, and disinformation about it with, on the other hand, criticizing or dehumanizing its followers.

Content ostensibly describing a religion serves as a dog whistle for attacks on the relevant religious community. This kind of content tends to surge in the aftermath of events, like the murder of Mr. Paty or the Christchurch massacre in New Zealand, when gruesome images of killings also proliferate online, so that social media companies find themselves occupied with trying to remove that.

This often means that they, therefore, do

not focus sufficiently on the hateful and false content that targets religious communities even in the times when that content is most abundant, when it is rife.

Another important trend that we have noticed is that rhetoric against religious communities often overlaps with xenophobia and the language of invasion. This language is a major hallmark of dangerous speech.

It is particularly common in the sometimes self-described manifestos of people who carry out massacres aimed at particular groups, and this kind of language, this language of invasion, this describing a threat and an either current and future invasion by another group of people, is very powerful since it suggests, often convincingly, with the help of disinformation, that another group of people pose an existential threat.

Again, this is terribly powerful language since it convinces people, often, that they must protect themselves and their families and their religious communities and their children by

committing or condoning violence.

I would be remiss not to mention also that hatred and disinformation are sometimes directed at religious communities from within by their own leaders. We have seen examples of this related to the COVID-19 pandemic quite recently, as in Bangladesh, Myanmar Burma, and other countries where clerics encourage their followers to attend large public gatherings, telling them that devout people were immune to the virus, and that those who were warning the same people to be careful were insufficiently devout or even atheists.

All of these types, these different types of content that I have described, circulate online of course. Now, I'd like to offer some ideas for countering them effectively.

First, I strongly endorse the points that have been made by commissioner--well, in fact, by all of the speakers thus far, particularly the concern expressed by Vice Chair Perkins and David Kaye, that we must protect freedom of expression vigorously even while finding the most effective

ways to counter hateful content and disinformation.

So here are some suggestions. I know my time is running out so I will simply sketch them and look forward to your questions and do my best to answer them.

The first is to work with social media companies to explain which content is dangerous since although I have pointed out that there are striking uncanny similarities in this kind of content from case to case around the world, understanding individual examples of it often requires an understanding of local context. And this is one of the main reasons why companies have failed in the past to act with sufficient efficacy and precision against such content.

Since it is so highly context dependent, often it's not at all obvious which content is dangerous and which is not, and how dangerous it is in its context. Companies need to make quick decisions in the event so they must have access to high quality information in real time.

This means building ties between companies

and reliable sources of information before there is a sudden surge of such content online. It must be preventative work, not reactive work, and indeed a great deal of content moderation is reactive as the common use of the term "whack-a-mole" to describe it suggests.

Second, it is important to choose the right means of responding to harmful content. In many of the discussions, in the vast majority of discussions on this topic, only one kind of response or policy is discussed, and that is what the companies call take-down, or removing the content. However, that cannot be sufficient by itself and in some cases may not even be the most effective response.

Alternatives include what is often called demotion or downranking, which means making content less, less visible to fewer people. Companies, including Facebook, are using this alternative quietly much more in numerous cases, including specific cases where religious communities have been targeted over the last year.

They are doing so, however, as David Kaye pointed out, in this case a completely opaque way. We might even say that the curtain is very dark so the use of downranking is even more mysterious and opaque. So it is absolutely essential to develop a system for oversight regarding the responses, the policies that social media companies are using to respond to these and, of course, other kinds of harmful content.

This, of course, also cannot be reactive. This must be the work of policymakers from a variety of sources, including but not limited to government.

It is also absolutely vital for making the right decisions to find sources of information and also speakers who are influential within the relevant community. That is to say there is another possible response that we, in fact, that the Dangerous Speech Project have been studying, a response to harmful content, which is to respond to it--literally what is sometimes called counterspeech.

This can apparently be effective, especially if the counterspeakers are people who are influential within the relevant community. Not nearly enough has been done to experiment with that possibility.

I'll now conclude by mentioning two more steps that are absolutely essential in my view, and currently almost entirely missing from content regulation by tech companies.

One is, as I've mentioned, oversight of which content they choose to remove or otherwise regulate. Here I want to point out that we, those of us outside the companies, know almost nothing about those decisions at scale.

As Chair Manchin mentioned, Facebook takes down millions of pieces of what it considers to be hate speech on a daily, weekly, and monthly basis. We know only what Facebook and other companies do regarding individual specific pieces of content when there is a public controversy over such content, such as the innocence of Muslims, to take an example related to a religious community.

We don't know on a given day what they are doing with the other approximately two million pieces of non-spam content that particular company takes down.

The second final step that is absolutely vital in my view is robust study of the effects of various interventions so that they can be chosen on the basis of data, not merely groping in the dark without knowledge of the actual effects of these steps.

With that, I'll conclude, and as I said invite your questions. Thank you so much.

CHAIR MANCHIN: Susan, thank you so much. And, yes, obviously we see that this is not, this is a complicated situation certainly with a lot of web going out in many different ways, both literally and figuratively.

And so it's now my pleasure to, as we move over to London, to hear from Dr. Shakuntala Banaji--excuse me--Dr. Banaji. Such a pleasure to have you.

Dr. Banaji is the Professor of Media,

Culture and Social Change in the Department of Media and Communications at the London School of Economics, where she also serves as Program Director for the MSc Media, Communication and Development.

Thank you for being with us this afternoon from there, and please proceed. Thank you.

DR. BANAJI: Thank you very much.

Thank you, commissioners, and thank you, esteemed colleagues. You set it up in such an interesting and eloquent way in terms of the debates on hate speech and violence across the world, but I think my case study, which is going to mainly focus on India, will be quite relevant and also will illuminate some of the points that you made.

I'd like to start today by briefly outlining the socio-political context of hate speech, harmful content and violence in India, and looking at the links between them.

Amongst many in India and in the diaspora, there is a deep and widespread social prejudice

against Muslims, Christians and Dalits, and I think it's advised for us to look way beyond the current focus on social media for the way in which this prejudice moves through communities.

The prejudice drives extreme socioeconomic and spatial discrimination and repeated atrocity, and that's the context into which social media comes. Hundreds of thousands of instances of malicious orchestrated misinformation, disinformation and hate against Muslims, Dalits, and Christians from links to speeches, memes, and gifts to long video posts or blogs circulate daily on social media platforms and peer-to-peer messaging apps, such as WhatsApp, Facebook, ShareChat, Instagram, Twitter, Telegram, and TikTok, to name but a few.

For our sins, some of us in my department have had a lot of time and also spent a very, very serious few months examining some of this content when it comes to India.

Cheap phones and data packages made available by corporate giant Reliance Jio since

2016, particularly courtesy of support, endless support from the Modi government, often preloaded with government apps, have made it easier than ever to promote hate and intolerance and to spread it deep within communities that previously one might not have thought to be connected.

At points of heightened tension, for instance, I refer to recent events which took place in Bangalore, and recent events earlier in the year in New Delhi, these rumors are triggered and become focal points for coordinated mob violence.

Since 2015, I think most of you will be aware that there have been more than a 120 instances of mob lynching, mainly against Dalits, Muslims, Christians, and Adivasis, and based on entirely false allegations of cow slaughter, cow trafficking, and cattle theft, but also of child theft and kidney snatching.

In a recent horrific incident, a Muslim man who had a sacred number tattooed on his arm had has arm severed by a group of Hindu men in one such incident of hate.

Many of these incidents are filmed by the perpetrators and circulated widely within social media groups, intentionally striking further fear into the hearts and minds of Muslim, Dalit, and Christian communities.

Those who dare to protest, the brothers, sisters, fathers and mothers of rape victims, for instance, or a lynched man, are harassed and intimidated, threatened or even killed.

Existing networks of disinformation lead to entrenched rumors that Muslims are intentionally, for instance, infecting Hindus with Covid-19. The context of Covid and the lockdowns in India have resulted then in further deaths caused by the turning away of Muslim citizens from hospitals, from educational institutions and residential settlements--a horrible byproduct of the already networked context of disinformation.

International human rights organizations, such as Amnesty International, which have been documenting the violence and oppression against Dalits, Muslims, Christians, and Adivasis in India,

have faced harassment and threats from the Indian state.

It won't be any secret to you that Amnesty International recently had its bank accounts frozen and have made the decision to quit India. There is a growing atmosphere of fear and intimidation not just amongst the civil society activists and journalists who are trying to document and protest the mob killings, but also amongst everyday communities of Muslims, Christians and Dalits who want to take up and fight for their own cause against the disinformation.

I think it's very important, commissioners, that I make the point that we can't just talk about social media separated from mainstream media.

Mainstream television and the social media context of anti-Muslim, anti-Christian, and anti-Dalit hate posts and violence are closely connected. Hate speech, misinformation and disinformation that circulates on social media in India is really linked to hate speech,

misinformation and disinformation that circulate and are produced by mainstream media outlets.

Both are linked to and contain malicious disinformation and hate speech by members of the ruling party and the government. This is something that many people fail to mention when they talk about social media hate. There's a clear continuum between the formats, the types of hate, of content of posts on mainstream and social media, in multiple vernacular languages, Hindi and English.

My colleagues and I have traced the use of fiction media formats in posts which are whipping up fear and anxiety about particular members of communities.

Everyday forms of hate speech and incitement against Indian Muslims, Christians, Dalits, and Rohingya refugees and Adivasis are normalized by mainstream media, such as Republic TV and Sudarshan News.

Hateful WhatsApp messages against Christians, Muslims and Dalits work in tandem with ideas which circulate in family and community

conversations outside the local cigarette shop or in the local mobile phone shop. A variant of any particular stereotype or hateful narrative containing misinformation will often appear at the same time in mainstream news media and in social media, a form which we call transmediality and which has really propagated many of the Covid conspiracy theories circulating at the moment.

Therefore, when some users call on their technical media literacy to go to multiple sources when in doubt, they often find only verification and repetition of false information and hate against Muslims, Dalits, and Christians.

The political ties of those who spread hateful misinformation are central to the allowance of attacks against Muslims, Dalits, and Christians in India. The same perpetrators of hateful speech against these communities with ties to the Indian BJP and RSS, for instance, the politician Kapil Mishra, repeatedly flout the regulations on incitement and get away with it.

During the recent Delhi violence of

February 2020 in which more than 50 people were murdered by mobs or shot to death by police, more than two-thirds of the victims were Muslims. The accounts of those who suffered body and financial harm are changed or refuted by the police.

WhatsApp groups, such as the Hindu Kattar Ekta, were allowed to organize pogroms with impunity despite some of their members being known incendiaries.

Colleagues, in this context, I really feel we need to ask how much we can call on the law and legal representatives in India to support us in the fight against hate speech. Media researchers and journalists with whom we are in touch for our work on social media hate have tracked the spread of and connection between hateful political speech, hateful postings online, and violence across multiple Facebook accounts and WhatsApp groups who are run sometimes by people with connections to the ruling party and to the police.

Attempts to combat misinformation and to instill media literacy are weakened by systemic and

the official nature of prejudice circulating throughout.

For instance, let's look at fact checkers, such as Factly, Boomlive and Alt News, which are overwhelmed by the volume and diversity of hateful misinformation against Muslims, Christians and Dalits, or against critical or dissident individuals in India.

Far right misinformation outlets have also set up their own fact checkers to discredit accurate information about hate speech and violence. Paid and unpaid trolls in the hundreds of thousands in India also delegitimize accurate reports and delegitimize anyone engaging in counterspeech. Platforms and corporations currently pursue sensationalism and profit over a commitment to all communities' quality of life and rights despite the fact that many of them avow a commitment to freedom of speech and quality of life.

Most mechanisms for reporting incitement in India on social media are merely technological,

and even where there are human subjects involved in looking at the misinformation, much further action is needed, and because there are many disappointed users who do report hate speech and get nowhere, many supposedly mild posts containing misinformation about particular communities and their leaders go under the radar because they are disguised as jokes or metaphors and never make it onto the list of what counts as hate speech.

So I want to conclude in the last minute by talking about a number of possible solutions to reduce such anti-minority hate speech and violence.

International bodies, including the governments of the United States and governing bodies in the EU and international corporate organizations, need to acknowledge and inform themselves about the links between various authoritarian regimes, government allied vigilantes, corporate platform executives and hateful disinformation.

There needs to be a meaningful social and economic incentive given to any government which

takes action against hate speech, including an early-warning system about impending anti-Muslim, anti-Christian and anti-Dalit violence.

There needs to be powerful business incentives to platforms and corporations which take swift action. Currently I think it's more encouraged than discouraged to ignore hate speech. Twitter, Alphabet and Facebook urgently need to join with local and international human rights organizations who know the on-the-ground context to ensure that their employees undergo rigorous human rights training on what constitutes hate speech.

There needs also to be a database of Islamophobic and anti-Dalit content in line with the same kind of databases around misogyny and pornography, which have been used very successfully, I think, in international content.

I will stop there and hope that my colleagues can ask some questions which push this issue further.

Thank you very much.

CHAIR MANCHIN: Absolutely, Dr. Banaji.

I'm sure there will be. And thank you so much for your input on this subject.

And last, but certainly not least, Dr. Waris Husain is a human rights attorney specializing in digital rights, human rights defenders, and business and human rights.

He is the former South Asian Policy Analyst for the U.S. Commission on International Religious Freedom. So we welcome him this morning and eager to hear your comments, Dr. Waris.

DR. HUSAIN: Thank you, Chair Manchin, and thank you to the U.S. Commission on International Religious Freedom for hosting this important discussion regarding the convergence of religious freedom and the ever-expanding digital world in which we're living. This was a conversation that we started to have while I was working at USCIRF as an analyst. So I'm glad to see that it's culminated in this hearing, and I'm thankful to you all for carrying that forward.

I'm also very honored to be serving on a panel with one of my personal heroes, Professor

David Kaye, who is here, who is a continual inspiration for people who are trying to get involved in this digital rights space from a legal perspective.

So I've changed my talking points as everyone was talking so I don't repeat anything anyone else said. So we'll try and keep things fresh.

But I'll focus my comments on regional developments in South and Southeast Asia, which seems to be a running theme that a lot of the speakers and commissioners have mentioned. To give an overall assessment, I think one must understand, just as Professor Banaji pointed out, that the already existing issues relating to religious minorities continue to impact all countries in the region, as they have for generations.

While social, economic and political disenfranchisement of religious minorities persists, the speed and reach of hate speech and fake news has changed dramatically with the astronomical expansion of Internet access in Asia.

Think of religious bigotry as a preexisting condition, a cancer, and think of proliferated Internet access as metastasizing that cancer. It's speeding the growth of that disease; right?

I mean as of 2020, out of four billion people living in Asian countries, more than two billion now are connected to the Internet. That's twice as many users as there was just ten years ago. We've seen the democratization of the information sector where a TikTok video by a 16-year-old in a remote Pakistani village can go more viral than a hard-hitting story by the BBC on a similar topic.

This presents opportunities for both good and bad faith actors along with religious bigots to us social media to expand the reach of their message.

And with this extended or expanded reach, we have seen interrelated issues that allow for misinformation and disinformation along with hate speech to be proliferated and cause real world

harm, which is exactly what the commissioners were mentioning in their introductory remarks, particularly for a religious minority community.

As the Internet access has blossomed, in Asia, we have also seen, as we've seen from the comments up till now, mob violence unleashed on minority neighborhoods, based on fake news going viral; very little digital education for users, making it hard for them to distinguish between information, disinformation and misinformation; the persistence of inauthentic behavior relating to religious minority communities that can be connected to troll armies or troll farms; social media platforms taking a far too passive role in content moderation, which I think Professor Kaye has rightly criticized, by Twitter and by its engagement with the social media companies that they just aren't doing perhaps enough; and then traditional legal tools and methods are either unable or unwilling to keep apace with the technological advancements.

So we need innovation and we need new

ideas, not just analogizing what we have already in terms of law or legal paradigms to the digital space. We need new ideas for them.

Some of the solutions--let's go through some of the solutions that actually we've seen in South Asia, particularly for these issues, and then see where there might be a gap or where there might be something to speak on.

Some of the solutions that have been implemented to stop the spread of hate speech or fake news with real world mob violence are over-expansive to the point of violating other human rights like the rights to free speech or the access to information. One of these over-expansive solutions is the increased use of Internet shutdowns in countries like Pakistan, India, Sri Lanka, to deal with hate speech and fake news.

So, in some instances, like in Sri Lanka, the shutdown is put in place to counter a viral news story falsely accusing Sri Lankan Muslims of various wrongdoings. This could actually instigate mob violence and a violent attack on a community.

In one way, this kind of swift action by the government can actually save lives. It can actually stop a mob from forming and attacking religious minorities, but in another way, the overreliance, overuse and over-expansion of shutdowns can have a counterproductive effect of encumbering interfaith efforts by activists to counter misinformation with increased cooperation between majority and minority religious groups.

Related to shutdowns, in Pakistan, there is a history of closing access to particular websites using take-down requests or rather using take-down requests to silence certain users on websites like TikTok, Facebook, YouTube or Twitter.

As you may know, TikTok was banned by the Pakistani government ten days ago, and the ban was lifted randomly without any transparency on why that decision was taken. So that's something that is happening in terms of take-down of sites, and also in terms of YouTube, YouTube has remained or was at least inaccessible due to a government shutdown because of an alleged blasphemous video

for several years.

And the government telecommunications authorities have consistently played a role in silencing religious minorities, particularly Ahmadis, in posting content that the authorities unfairly deem as blasphemous, and they go to social media companies to take down content based on Pakistani law which criminalizes blasphemy.

This goes back to what Professor Kaye was saying as it relates to already existing laws that are being brought in the digital field and then almost they're over-applied because the digital space is so wide and vast.

Having laid out these difficult scenarios, I believe that the speed of communication has both good and bad repercussions, and the speed has gone into overdrive with the penetration of Internet access in Asia.

While religious bigots have become increasingly adept at using this increased speed to their advantage, governments, activists, and social media companies are lagging far behind without

producing effective and narrowly tailored solutions.

And the goal for all three of these parties--governments, activists, and social media companies--is to come together in good faith and create social, legal and technological solutions that not only protect religious minorities but also protect the rights to free speech and open dialogue.

While analogizing traditional legal solutions from non-digital forms of press and communication can be helpful, policymakers have to understand there are unique challenges posed in the digital space, and therefore there is a need for a wholly new legal paradigm and solution rather than retrofitting existing rules or traditions.

Also, the silo that exists between engineers and technical experts and human rights, sociologists or linguistic specialists have to be broken down in order to avoid the kinds of mistakes we've been continually making in creating narrowly tailored or rather siloed digital strategies.

The human rights community needs to understand the limits and capabilities of the technology, while the engineers need to understand the value of the input from human rights specialists or linguistic specialists to bake into the technology ways to ensure safety, dignity, and respect for religious minorities rather than trying to reverse engineer solutions once a problem has arisen with a technical issue.

So a few specific recommendations we can talk through, and I'll try and be very quick with these. I think the AI that was mentioned--the artificial intelligence--artificial intelligence perhaps needs to be bolstered by human intelligence, right. HUMINT is also an in important part of creating, of facilitating artificial intelligence that makes sense.

Along with that, deprioritizing or downtracking content, which Professor Benesch has been talking about, that straddles the line between hate speech and free speech could also be an alternative tool. So we have to have different

tools in our toolbox, not just taking down the content but also maybe deprioritizing it and prohibiting it from going viral.

In addition, I think one of the things that we can look at is there is already existing sort of paradigms that look to early warning mechanisms for genocide, early warning action that look at heat maps, essentially saying a post in XY and Z country, a post in Pakistan, could cause violence at a much faster rate than it might in a country like France, for example, or maybe not-- France is a not a great example considering what they're going through now. But different countries have different sort of contexts with which to analyze how likely is speech to cause harm.

We can't apply global standards necessarily. We do have to look at country-specific examples.

And then, finally, I think that the adversarial relationship that we have that exists between technology companies, social media countries and governments has to be broken down in

a way that becomes more collaborative and less adversarial.

It feels like oftentimes these social media companies are rushing to create policies so that they can avoid government regulation on those policies rather than thinking of a way to collaborate with government authorities to create policies that make more sense.

So I know that's a lot of information I've thrown at everyone. I'll stop talking there, and I'll transfer it over to Chair Manchin for the questions.

CHAIR MANCHIN: Well, thank you. Thank you so much to all of our speakers.

Unfortunately, and I think David Kaye, Dr. Kaye mentioned this early in our program, that he did have to leave at 12 for another commitment so he will not be available for questions. But certainly our other panelists are still with us, and I'm just going to begin--and I kind of throw this question out to one of you or all three of you--but, you know, when we look at how

comprehensive and sort of all engulfing this issue is, but when we actually know for a fact that hate speech or disinformation is actually being government sponsored in a country, that they are the ones leading it, then obviously you can't go to the government looking for help on how to solve it.

So, kind of broadly in terms of what we can do, not only as USCIRF, but certainly the U.S. government, in actually intervening with government-sponsored hate speech?

DR. BANAJI: Can I pick up on that, Commissioner Manchin?

I thought that was a very interesting question and a good segue from the previous speaker because Dr. Husain talked about cooperation between governments and platforms, but in India, we've seen the absolute opposite problem, which is there is cooperation to suppress actions against hate speech.

And the recent case of Ankhi Das, who was for several years actually continued her own agenda of anti-Muslim postings as well as suppressing

attempts to take down hateful material against the Muslim and Christian communities on behalf of the BJP government. So I think I would say alongside you that I think we need an initiative which is multi-platform. I don't think that it can come from the United States government.

I think it needs to be multi-stakeholder, and it needs to be multi-country, and that's the only way in which it would retain both its integrity and its ability to do its job.

CHAIR MANCHIN: Thank you.

Susan. You're on mute, Susan.

MS. BENESCH: First, I couldn't agree more with what Dr. Banaji has said in all points.

I'd also like to add just a note, that this is, as you said, Chair Manchin, it's a serious problem regarding many governments around the world. In fact, many of the governments where religious communities suffer the most hateful content and disinformation are exactly the places where the governments are either producing such content or paying large numbers of people to

produce it or strongly, tacitly encouraging that, or all of the above.

When advocates like me ask the tech companies why they're not more vigorous in taking down such content, they sometimes say, well, you know, we're operating in this country. As I believe it was Vice Chair Perkins mentioned, India is, he said, the country in which Facebook has the largest number of users. But Facebook calls it its biggest market notably.

So when, for example, just to take another government, when Turkey reports enormous amounts of content for takedown to Facebook, and the Facebook staff look at it and see that it's mostly content that is sympathetic to Kurds, just for an example, the Facebook staff say, well, it's very difficult, you know, we can't very well go into court on every single one of these cases, they're also afraid of being prosecuted in the various countries.

So it seems to me we must seek ways in which the companies can push back more strongly against such governments, and, as Dr. Banaji has

said, one way for them to do this is to do it not individually, but collectively.

There's just beginning to be talk of some kind of meta-organization among companies. We have an example in what's called GIFCT, which is a consortium of companies to identify and take down terrorist content. There are also other possibilities such as I'll now reference David Kaye, who sadly had to leave us, but David is one of the principal people who have been advocating for requiring companies to adhere to international human rights norms for content moderation.

If they do that, companies could say to countries that are demanding take-down that is not in keeping with international human rights law, companies could then say, sorry, we must abide by the law.

I will say very honestly that I suggested this to a colleague, coincidentally someone from India, who laughed and said don't be silly; those countries, those governments are not taking human rights law seriously in any respect already. What

makes you think that if the companies use a legal basis for pushback, that they'll take it seriously?

But those of us who have been working with human rights law, including all of the commissioners, for a long time, know that like, like many pursuits, it's an uphill battle. It isn't always successful, but these are, at least, possibilities.

And, of course, as has been mentioned, the companies are often making policy in order to fend off regulation by government. So if governments, not singly, but jointly, can assure the companies that to push back against overbroad and repressive regulation of speech demanded by certain governments is seen favorably, then the companies will have more incentive to push back harder. They must do that.

CHAIR MANCHIN: Thank you. Thank you so much.

COMMISSIONER BAUER: Madam Chair.

CHAIR MANCHIN: Yes.

COMMISSIONER BAUER: Yes. As I mentioned

a little earlier, I'm going to have to leave in a few minutes also for another USCIRF event, so if I could squeeze a question under the wire here, I'd appreciate it.

I'm, I must admit, I'm still sort of hung up on what was referred to by a number of our participants on how to balance the incredibly important public policy goal of trying to limit hate speech with the broader question of freedom of speech generally. And I'm not sure I still understand how to completely do that.

There's some examples that are so obvious, and somebody is saying that Muslims are intentionally spreading Covid, and that's resulting in Muslims not being able to access medical care, I think everybody would agree that that is hate speech that's having very real consequences.

And likewise, the blood libels that have been used against Jews for centuries, I think we all know that that's beyond the pale, and social media should not permit that sort of just ridiculous hate to spread.

But then I look at a situation like Western Europe where I think there are legitimate debates going on about whether the influx of large numbers of migrants from third world countries that don't have the same attitudes as Western Europe does about women's rights, about sexual minorities, about religious pluralism, I think there are legitimate concerns on both the right and the left in Europe that maybe that mass migration needs to be slowed down or severely restricted.

And how do we make sure legitimate debate like that takes place without those arguing for restrictions on the mass migration being labeled as haters or engaging in hate speech? Anybody or everybody?

DR. BANAJI: I'll defer to Susan, but I just wanted to say very briefly that actually I'm afraid you can't just get away with saying that some debates are legitimate completely and some debates are not because part of that legitimacy is also giving legitimacy to people who engage in violence.

I'm afraid there's a continuum, and I speak from Western Europe where the bodies have been washing up on the shores, and so there is a continuum between the debate. And, again, nobody here is saying the debate shouldn't be happening. It's more that it results in violence for one group of people and not the other. So it's never the group of people who are saying we have concerns about these people coming here who end up dead on the beaches or who get beaten to death in racist incidents on our streets.

So I just wanted to make that point before the others come in and answer the question, that there is a continuum between something that looks like legitimate debate and something that ends in death and blood.

COMMISSIONER BAUER: Well, I would beg to disagree. I mean there's a lot of evidence that there are things being taught in mosques in Western Europe that does, in fact, lead to Islamic extremists engaging in violence to other religious groups or seculars or sexual minorities.

MS. BENESCH: If I may, I think I've heard you both saying that there is some speech that's--my term is dangerous. In other words, that some speech leads to violence, it seems.

To answer the question, I would suggest that every society in lots of ways, most of them not written down, most of them not actually law, develops and enforces what would be called in academia "discourse norms." Certain--you're permitted to say some things and not permitted to say other things, in your family, around your kitchen table, in your religious community, in a house of worship, on a field where you're playing a particular sports game, et cetera. We all abide constantly by these--by these unwritten rules about what one may say and what one may not say, according to what the rest of the community thinks those are norms.

Online, the rules are written and enforced by a very small number of people who come mostly from one cultural background. It used to be that they were mostly Californians. Now that's not so

much the case. The social media companies do have employees and even policymakers from a variety of backgrounds.

However, the rest of us, outside the companies, really don't have a good sense of what they are, where they're drawing the lines in practice.

So the first thing we need is--in my view--a system of oversight so that we do understand where they're drawing lines. And the second thing I believe we need is, although it will be difficult, to change their claim that they're making one set of rules for the entire world. Facebook claims to have one set of what they call community standards for the whole world.

Think about that. Facebook famously bans nudity, and so that would suggest that they're using the same rule in Sweden and in Saudi Arabia, which as you can imagine is nuts since different communities of all kinds have different norms for speech and for behavior.

So the second thing I would propose is

that companies develop some form of a system for some kind of input so that people from at least a country, if not from a region--perhaps this should happen on a more localized level--but they should permit systematic input from people who are affected, who are governed by their rules, into those rules and in particular into the enforcement of those rules.

So that number one, the rest of us who live under these rules understand how they're actually being enforced and, number two, so that there is some mechanism for input. Facebook has announced a new what they call oversight board, which is a collection of mostly lawyers, many, but not only, Americans, who will have once this board gets going in a few more months some sort of input.

But the input--and this is a really key point--is only into the rules, not the enforcement, and we know from American criminal justice and all too many other examples that if all you know is the rules on paper, and you don't know how they're being enforced, you don't really know what's

happening out there.

COMMISSIONER BAUER: Thank you very much.

I apologize again for needing to leave, and I'd love to continue the conversation past today with each of you, and I want to thank you all again for your enlightenment that I've gotten on some of these issues and on this important topic. It's good to spend time with you this morning.

Again, sorry I have to take off.

CHAIR MANCHIN: Thank you, Commissioner Bauer.

And I'd like to turn over to Vice Chair Anurima Bhargava to see if she would like to ask a question before we go to other commissioners.

VICE CHAIR BHARGAVA: Sure. Thank you, Chair Manchin.

I do. I have a couple questions that I'm going to try to speak to a few of the things that just got mentioned and then that Dr. Banaji mentioned about incentives.

And so, Susan, I want to start where you just ended, right, which is the question of

enforcement because I've had those same conversations with social media companies where it's sort of like we could have this community set of norms globally, right, and then it's sort of on governments to enforce.

And then we have the problem laid out very, very succinctly and beautifully by Dr. Banaji about what happens when the governments are not enforcing and, in fact, not even allowing the space for others to try and identify or hold those accountable or even being able to report; right?

So in that context, you had mentioned, a number of you had mentioned sort of incentives, and I'm sort of wondering what those incentives would be because the social and economic incentives to governments in this context, and so this goes back, Susan, to your point, which is that if it's a market, right, and the market works on what we know it works on, which is for both hate and disinformation, clickbaits, and lots of other people showing up, and how do you actually think about what the social and economic incentives are

to enforce in a really different direction?

When it's not a rights-based enforcement, it seems like it's an economically-driven incentive that we're thinking about. So for all of you in different ways, I just want to ask both about how do we think about the governments enforcing, where there are socioeconomic incentives that you were talking about, Dr. Banaji, and then also for the market-based way in which, you know, as some who-- I'm thinking about, you know, Geron Manay [ph] and others who talk about like what it means to have, you know, a conversation between two people and the companies are here to sort of manipulate it. How do you actually change that to account for what we're seeing for religious--

DR. BANAJI: I'd love to jump in quickly because this was one of the main issues that we found in our work on WhatsApp. So I'll use WhatsApp as an example because I know that in the Indian market, they clearly are trying to become a mechanism for using payment, which is in competition with an Indian one called Paytm, so

they need to be secure, and an incentive for them, for example, and incentives cannot be across the board. They need to be tailored to particular companies and particular circumstances.

So for them an incentive would have been to ensure that their decisions around hate speech were not somehow making them unappointable or unemployable as an alternative payment transaction company. I'm not supporting them in their sort of capitalist aim for global supremacy in the payment market or the financial transaction market, but I think what you'd need to do is you need to ensure that making decisions in favor of human rights were not then being used for them disfavorably when it came to economic competition with local companies.

So where you might have the government awarding a contract for financial transactions to a sweet local company who is okay with hate speech being circulated in their other formats, and I think what you can see in India is the building up of particular global, Indian global corporations with the absolute agreement of the Indian

government and who then turn a blind eye to things that are going on with regard to hate speech when it favors the ruling party.

So, for example, Reliance Jio, which I talked about, who have somehow beaten off all competitor phones, we almost had a situation where, you know, Vodafone had to leave India because their debts were being called in. You've got these situations where at the moment, it's a very unvia-- you know, sort of--it's untenable to compete in that market with companies who will turn a blind eye to human rights abuses and will therefore be given contracts.

So that would be an incentive, for example, a level playing field, not that I think the playing field is level, but that's a possibility.

DR. HUSAIN: If I could build off of that. I think this is a really great question from Commissioner Bhargava in terms of incentivizing. I also think that this is sort of categorically we need to look or at least technology companies need

to look at what their role is in a society or is in culture; right?

I think that they're going off the capitalistic and sort of the market will lead, and that's what we'll sort of reward with the algorithm, et cetera, et cetera. But if there was a difference within the companies themselves who felt they have a social obligation to perhaps counter speech.

There's a big sort of area where influencers, right, young users or users who are very, very influential and can spread a message quickly are kind of chosen by the algorithm itself right now. But if companies themselves were to look for people who are doing exactly what we were talking about earlier, looking at building communities, building bridges, that they're good at that, they could be awarded with influencer status, not with fake followers, but that the algorithm could preference those users as well.

That's something that could at least have that conversation like you were talking about,

Commissioner Bhargava. How do we level the playing field; right? I think that's one way to level the playing field, but it requires these companies to look at themselves in a completely different light than they do currently; right?

The other thing that I would mention that--and I know this is pie in the sky, but I'm a pie in the sky kind of guy--right. We need some kind of a treaty. We need a multilateral treaty that speaks to some of these issues, and I think to award governments, right, based on sort of that treaty goes along with everything else in international law, that you sign on to treaties, you become part of the international community. You're given certain access and given certain points of technical assistance, et cetera, from UN bodies, et cetera, et cetera.

So if we had a treaty that did this, we could have multilateral ways in which the United Nations or other international bodies could actually reward or assist governments who are trying to do the right thing.

For example, like a human rights committee that has leaders, right, that has chosen and elected leaders who help formulate the implementation of that, of the treaties could be something that if you had a technology treaty or a digital treaty, you could reward the countries that are doing this content moderation the right way, that are doing the take-downs the right way.

And I guess going back to what Professor Banaji said, my video dropped out, but I think, I would distinguish between co-conspiring and collaborating, but I think in India what we have is co-conspiracy happening on a lot of things with the technology companies and the person that you mentioned at the company, whereas collaboration was something else that I was discussing of like good faith interaction between the companies and the governments themselves rather than trying to have it be sort of led by a political goal.

So I would say for both the companies themselves, they have to look at their role in a different way in order for them to award, implement

and push out the algorithm to certain users who are actually pushing good content out, and then the multilateral approach through treaty bodies would be something that we could try.

Of course, I think Susan mentioned, you know, that some practitioners would laugh at that. If that's not possible, it's pie in the sky, but I think we have to make some big asks and we have to think imaginatively about what can we do, maybe not now, but maybe in five or ten years, and that's something we have to think about doing.

CHAIR MANCHIN: Thank you.

Susan, did you wish to add to that?

MS. BENESCH: No, I've talked a lot and everyone else has been so eloquent. No need. Thank you.

CHAIR MANCHIN: Thank you.

Commissioner Davie, do you have a question you'd like to present as we kind of come down to our final, to almost the end?

COMMISSIONER DAVIE: Thank you, Chair Manchin, and I want to thank the panelists again

for this very enlightening, as Commissioner Bauer said, conversation. Very educational for me.

One of the things I'm curious about is this notion of trying to be less reactive in enforcement and oversight and more anticipatory. It seems to me that reaction is sort of inherent in oversight and enforcement. And so even in the best of worlds, let's say we had a set of international acceptable standards around what is legitimate content when it comes to speech and other things. Does the technology exist or could the technology exist that would filter out violators, if you will, prior to its presence, the presence of speech on a particular platform?

MS. BENESCH: Maybe I will just jump in quickly on this one. I'm so grateful to you, Commissioner, for asking that question, which is a vital one.

It would be so nice and such a relief if someone could build a classifier, as the techies call it, that would--let me say this. It would be marvelous if, first of all, somebody could build a

system to identify all the content that is bad and distinguish it clearly from the content that isn't in advance. That would be very tough. Think about how difficult it is simply to get people to agree on what is and isn't hate speech.

We have no consensus definition for hate speech. As the chair mentioned, it doesn't exist in international law. It doesn't exist--there are many definitions, but they're almost all different. And then if I gave all of you a set of ten examples, I'm willing to bet you something really nice that you would code them differently. The different ones among you would call specific examples of content hate speech and not.

So if you can't get humans to agree, it is terribly difficult to build software that can consistently agree on something, and of course it is such a tempting idea in part because people are expensive. That's one reason why the companies are trying very hard to build the software.

Facebook has increased by many thousands of people its content moderation armies over the

last few years under pressure from people like the panelists, but that, that job, as you can imagine, does terrible damage to people. There's a very good film that was made about it in which a young woman talks about having watched a great many beheading videos.

So for so many reasons, it would be marvelous to have software. However, it is a very scary idea for those of us interested in freedom of expression. That's all of us, I know, since we're all human rights people.

In particular, I worry terribly about prior censorship, about building and deploying software that would take content down just as soon as it's posted, especially if once again we continue to do all of this without any mechanism for oversight.

Facebook is run--I began a recent article with this line--Facebook is running the largest system of censorship that the world has ever known. Facebook by itself. Never mind the other company.

It's bigger than that, than the system of

any government, even including China's, and yet we don't know what they're doing. So they are using algorithms, using--algorithms is just in this context a term for software. They're using automated methods more than ever before, especially now because of the pandemic. They sent lots of their subcontracted content moderators home because of Covid.

So at this moment, more hate speech is being taken down automatically from online platforms than ever before. In fact, I've written with a wonderful colleague a piece pleading that this not quietly become the status quo after the pandemic finally ends, especially if we don't have any oversight. It should terrify us that all of this is being done in the dark with no, with no oversight.

And, of course, those of us who want to prevent violence also want to try to help them to get it right. It is this, as Vice Chair Perkins said, I think he said it's a very tightrope that the companies are walking. It's tremendously

important for them not to fall off either side, not on the squelching speech side or on the failing to take down awful content side.

And I'm sorry. I know I'm talking a lot. I just want to mention one last thing. We cannot due to a lack of oversight even begin to get answers to questions like if Indians post a particular type of content, and Pakistanis also post that kind of content, is it being taken down at the same rate?

If women post certain kinds of content and men--is that--if whites and African Americans, et cetera, et cetera--how is it possible? If Christians post a particular kind of content and Muslims, Hindus, Muslims--and so on? We who are all so I think wisely interested in equality and nondiscrimination and the enjoyment of human rights, how is it that we cannot seek answers to these critical questions about the means of communicating that have become so, so important and dominant all around the world?

DR. HUSAIN: Just to add on to that, I

think the mistake we can make here is by becoming static in our analysis. And I think that's why overreliance on technology gives us a static mentality that we just can't afford.

It has to be dynamic. Users are changing as they go. The companies are changing as they go. The situations on the ground are becoming hotter or colder as they go.

But I definitely understand where Commissioner Davie is coming from. Where can we have a little bit of reliance in terms of like what's happening here? How can we have technology be a solution? But I think perhaps up till now our overreliance on thinking of technology as a solution in a static sort of way may have been the reason that we've gotten here and why things have gotten out of control, and we should maybe make it a more dynamic approach, just like Professor Banaji was saying.

Having that mix of AI, having that mix of content moderators or human, I think we just can't look--I mean I love the idea, that we could have a

technical fix-all and that would take care of everything. I just think that we haven't found it yet. So I think that we should maybe stop relying so much on it and then looking ahead at more dynamic perspective.

I'm so sorry, Professor Banaji. I know you wanted to say something. I'll quiet down.

COMMISSIONER DAVIE: And let me just say for the record, I'm not advocating a particular position. It's more out of curiosity and having my own I guess intellect around this piqued by the conversations that we've just had over the last 90 minutes.

DR. BANAJI: I think it's a great question, Commissioner Davie, and I think there are ways in which we can introduce technology into supporting human action around these issues.

So, for instance, the building of databases, which I know have been done by multilateral agencies around child pornography and child trafficking, which is one of the most important steps forward in building a database of

things which look innocuous, but which actually aren't, which have led to actual harm.

And so if we start a database from the basis that if something has caused actual harm, let us say someone has lost a job or their employment has gone, or they've had their arm chopped off because of it, that thing is marked as hate speech, whatever you or I may debate in our academic setting about that, then we could actually assist those people both in corporations and in government who are looking at this material on a daily basis.

So we can, and databases can be technologized, they can be easily shared, and I start from the premise that if we did that around things like anti-Christian posting, anti-Muslim posting, or anti-Dalit content, which is very rife in the U.S. and in the UK as well, we would be moving forward considerably.

COMMISSIONER DAVIE: Thank you.

VICE CHAIR BHARGAVA: Can I just add on that one part, and I know we're past time? Dr. Banaji, I just wanted to say I feel like the ways

in which you're talking about early warning systems should require that we don't actually have to have an arm cut off before we realize that the consequences of what it is that's being said lead to an arm being cut off; right?

And so we know that. We've seen it. That's what the heat maps are telling us. That's what the early warning systems can really grasp that on to. And to recognize that it's not, you know, we don't need a consequence which is so difficult to demonstrate right now in so many different contexts for even someone to be able to come forward, that I think it's important, as you said, that when we see lives being lost and the consequences of it, that we account for the fact that we have a good idea of what's about to happen.

And we also realize that people are intending for that to happen. And that is the saddest part of this, this conversation. So thank you again.

CHAIR MANCHIN: Well, Vice Chair Bhargava, thank you, and Commissioner Davie, thank you.

What a great question to sort of bring our conversation back around together, and obviously if this were easy, we wouldn't have it as a problem. And so we continue to look, research, think about how we can be more proactive rather than so reactive in these situations.

But I again on behalf of all of the commissioners and USCIRF want to thank our distinguished panel today and the excellent information and expertise they've shared with us, the recommendations that they have made that hopefully we can use as we move forward in trying to be part of the solution, trying to be part of the proactive base and hopefully some way as we continue to see because it's proliferating around the world. The use of hate speech and disinformation certainly is growing and expanding, not shrinking.

And so we will continue to be challenged, and we will continue to look to people like each of you, Dr. Banaji, Susan Benesch, Waris Husain, and David Kaye, who had to leave us. We will continue

to look to people like you as we try to find peaceful solutions moving forward.

Thank you so much for your participation to all of our guests out there. Thank you for joining us today. Again, the bios were on the chat line, the complete bios of our speakers, but thank you for being with us today, and until our next hearing, be safe and be healthy. Bye-bye.

[Whereupon, at 11:38 a.m., the hearing was adjourned.]