Election Chaos: Platform Preparations for the US Election

OK, I'm going to get started and kick us off. My name is Oumou. I'm a fellow on the Berkman Klein Center's Assembly Disinformation Program. I am really excited to be joined today by two important voices at the intersection of law and technology. Evelyn Douek and Julie Owono Evelyn is a lecturer on law and an SJD candidate at the Harvard Law School, an affiliate of the Berkman Klein center, and she studies global regulation of online speech and private content moderation.

Julie is an attorney and the executive director of Paris based Internet Without Borders and a member of the Facebook oversight board. Thank you, both Julie and Evelyn for joining me today.

Thanks for having us.

Yes.

So the topic of our conversation today is lessons learned, particularly from the 2016 elections, which were a formative moment for platforms. The manipulation of social media by various actors in 2016 presented social media companies the opportunity to learn from and think through how to improve and refine approaches to content moderation, to develop functionality around assessing and addressing inauthentic behavior, and other automated abuses of the platform. And more generally, to develop a strategy instead of practices and policies to deal with disinformation.

There's also been a lot of fodder for lesson learning and adaptation since 2016, including the 2018 midterm elections, the COVID-19 pandemic, and other high interest world events. And so to begin the conversation, I'd like to have Evelyn set the stage with an overview of the major lessons learned and inflection points over the past couple of years.

At a high level, Evelyn, what would you say are the key lessons from the 2016 elections that the large social media platforms had in terms of content moderation and inauthentic behavior, in particular?

So shall I jump into my--

Yes.

Excellent. All right, let's do it. I'm going to share some slides because I am a slide addict.

OK.

OK, so can you say all that all right?

Yes.

Perfect. OK, so what have we learned-- what are the big lessons from 2016 to 2020? We've actually come quite a long way, I think, in many respects. I don't know if you remember, but here-- one week after the 2016 election, here is Mark Zuckerberg saying the idea that fake news on Facebook had an effect at all is a pretty crazy idea, whereas now, I can barely go anywhere on the internet or listen to any podcast without Facebook telling me how very seriously they are taking the 2020 election. And everything that they're doing to make sure that it's all going to be A-OK. They've made significant improvements. We're working hard to stop foreign interference. This is not-- it's no longer crazy. This is something that's absolutely front of mind.

And it's not just Facebook, of course. A bunch of other platforms, Twitter, Pinterest, YouTube, TikTok, all of them are telling us, reassuring us, they've got this, guys. So what does that mean?

Well, I'm trying to keep track of what platforms have been doing and all of their policy announcement, particularly in the last month, have sort of felt a bit like this, to be honest. It's sort of been like a DDoS attack of just constant policy updates. It's been pretty full on. And so I'm not going to be able to go through it in detail, obviously.

But I think we can probably say we can put it into four buckets, so connecting people to more information proactively. They've also rolled out a bunch of new content rules. We've seen some new thinking outside the takedown, leave out binary, and I'll talk about that a bit more later. And this, as Oumou mentioned, the coordinating of inauthentic behavior, the foreign influence campaigns, they're taking that more seriously.

Now these first two, I think, we can really sort of see in some ways as an outgrowth of what platforms did in the context of the pandemic. When the pandemic started, this was an opportunity where it was sort of like a break glass moment. It was obviously a state of emergency. They took some fairly heavy handed responses to that at connecting people to World Health Organization information and local resources and rolling out a bunch of new rules about misinformation, health misinformation, and aggressively taking it down, purely on the basis that it was false in a way that we haven't really seen platforms are willing to do before then. And I think there was by large a positive response to those efforts. And so it's sort of-- we've seen a more emboldened approach now in this context as well. So it's interesting to look at that trend.

What does that look like? Like I said, I can't really keep track of it all. And I'm not going to go through it now. But fortunately, the election integrity project, a partnership at Stanford, has been doing a bang up job. And especially Carly Miller, I'll give a shout out, of keeping track of all of these. They have been showing all of the new rules that platforms have been rolling out to deal with the election around false claims about procedural interference or participation interference or fraud in relation to the election. And you can see here a bunch of new policies. Also to do

with-- as it became clear that delegitimization of election results was going to be a major threat, a bunch of new policies around that.

If you're wondering what's keeping me up at night, it's this one here. YouTube does not have a policy about what it will do for false or premature claims of election victory. I don't think it's hard to see what the nightmare scenario is there. But as much as I deploy the awesome power of my Twitter account tweeting about this and saying, you need a policy on this, they haven't listened so I don't know what we're going to do about that, but we will see.

OK, so those are sort of the first two buckets of a bunch of new rules around election content and civic integrity content. Then we've seen a bunch of new things like labels and friction. So instead of saying, let's take down this bad stuff all the time, we're introducing a bunch of intermediate measures and thinking outside of that purely like, bad stuff comes down, good stuff stays up, sort of a false binary that I don't think we really need to be kept into when it comes to content moderation.

And so measure in there was these labels. And in fact, the very first labels that we saw were on-- in the context-- months and months back, back in May, Twitter put a label on a tweet from President Trump about voting in the context of the pandemic and mail in ballots. And that was sort of a big, explosive moment for content moderation. Mark Zuckerberg went on Fox News, I think, the next week or the next day, and took a swipe at Twitter and said, we would never do such a thing, we will not be arbiters of truth. How could you?

And then a couple of weeks later, reversed policy, and now they too, are deploying labels. And labels are the new hot thing in content moderation. We see them in all sorts of contexts. They are the new tool, because they seem to walk a nice line between being sensors and sort of adopting a counter speech approach, right? Like, the information is there. It's in the public interest. You might want to know that your public official is saying this, and you can take that into account when you vote. But also, we're going to make sure that we connect you to accurate information. So these are a tweet and a post from just yesterday, in fact. So those-- lots of labels in the context of voting.

And then a bunch of ideas around friction. And this stuff gets me quite excited. I hope that we see a lot more thinking in this vein as well. It's interesting. I just also want to note how quick this is. Like, these first labels were in May. That's only a couple of months ago, and here we are, and it's so-- like I think, in some ways, it feels like too little too late. And in some ways, it feel like this has been a really rapid development.

And again, so these friction ideas, I think we're already at the start of thinking about this. But Twitter has rolled out a bunch of things like when you attempt to retweet misleading information, you get a prompt about it. And then in particularly high profile accounts with 100,000 plus followers, or if something's gone viral, they will put a flag on it. And acknowledging that you can't catch everything, but if you catch the really high profile stuff, that can have a really much bigger impact.

The idea that when you go to retweet something, they're just going to add a little extra step in there prompting you to either add commentary or think about it before you retweet it to try and slow things down a little bit. And during the week of the election, they're also turning off some amplification functions.

I'm excited about this stuff. I think this is the kind of stuff that I want to see a lot more thinking about more generally, outside of the context of elections too. It's hilarious. Everyone loves fiction and loves the idea of friction in theory, but then when it gets rolled out, and it gets put into practice, everyone's like, Oh my God, what's Twitter doing that I have to like click an extra button to retweet something.

And then some of it's like, friction for thee and not for me, a lot of complaints about this. And it's prompted some sass from the Twitter comms account being like, literally just click one more time, and you can retweet it. It's not such a big deal. So I find that funny that maybe if they introduced a lot more friction, maybe we would all mutiny after all.

Troll hunting. So by contrast to Mark Zuckerberg saying, pretty crazy idea, fake news, not a big deal. Now, the phrase coordinating inauthentic behavior, which literally did not exist in 2016, it was made up by Facebook in 2017, now is everywhere. And he has a bunch of posts from Facebook about all of the coordinated inauthentic behavior that they are taking down and have taken down recently. And we get these updates constantly from Facebook and the other platforms about all the trolls and things that they have found.

Never-- not-- they're catching them now before they get significant engagement is the story. And so they're being much more successful on this front. And furthermore, there's a lot more cross industry collaboration and coordination with governments in finding this stuff. And we hear a lot of stories about how the intelligence community tipped off the platforms about certain influence campaigns, and that prompted them to take them down, and those kinds of avenues of communication that just weren't open in 2016 and have really sort of been-- have grown in the past couple of years.

We have absolutely no insight into this. We have no idea what they're talking about, what information they share, how effective it is, anything like that. This is something I am concerned about. I don't-- I think it's very problematic for accountability and transparency. But they release these press statements, and people tend to find them fairly comforting apparently, that they're all working together and taking it very seriously.

So it all sounds awesome. All this stuff sounds incredible. As usual though, with platforms, the thing is a policy can look great on paper, but the question is, will they and can they enforce it. And I think there's certainly historical and recent-- historical as in historically in the context of content moderation being like the last couple of years-- examples of ineffective enforcement of policies. Like you can have an excellent hate speech policy or an excellent incitement to violence policy, but if you just don't have the resources and aren't dedicating sufficient attention to it, does it really matter? And that's going to be the big question.

The other one with labels, for example we've seen, is you can have a policy that we will label something, but if it doesn't get applied for a couple of hours, and the tweet or the post has already gone viral or been seen by hundreds of thousands of people, it's almost not-- it's almost not much better than not enforcing it at all. We have seen improvement on that front. It's literally-- there's a few accounts, maybe one, that it would be great if they could keep an eye on and try to enforce against pretty quickly. And again, like yesterday, it was much, much quicker than the sort of three or four hours that we saw even a month ago. So, progress, I guess, but we will see-- we will see what happens in the next week or so.

Just acknowledging the elephant in the room, and I can't do this on a Berkman Klein panel without acknowledging the resources about this. I've just spent 10 minutes talking about content moderation, as if it's really important, and all these rules are really important. And I do think that they are. And I do think that platforms have a lot of responsibility here and need to do a lot better. But at the same time, content moderation is a fairly limited lever to pull in these circumstances when the president is the one tweeting out or posting disinformation and misinformation about ballots and voting processes.

And we certainly see that no matter what the platforms do in a sense, this is one of the defining articles of this era for me. Kevin Roose termed it, the president versus the mods. No matter what the platforms do, the president seems to be finding ways to push the boundaries, and find the gray areas and the ambiguities, and try and set up this dichotomy and this sort of conflict between them, this story about bias against about-- bias against conservatives and things as well.

And so building on important research released by our colleagues here, earlier this month, a couple of weeks ago, showing that the social media component of a lot of this is really secondary to the president. And the way that that gets picked up in the mass media and Fox News and that right wing ecosystem has a very large-- very large effect.

So, content moderation, important. They need to do much better. Platforms have a huge amount of responsibility here, but also is always going to be somewhat a limited lever to pool when there are massive other institutional failings. And keeping that-- with that sort of background in mind, I think the New York Post story of a couple weeks ago is a good example of this, where there was a story about a laptop with potentially hacked and leaked emails of Hunter Biden. I'm not going to go into the underlying story. But what was interesting was that the platforms reacted fairly quickly in this case, seeming to want to avoid the appearance of 2016 all over again.

Facebook came out and said, we have down-ranked this story across-- it's eligible for fact checking. We have down-ranked it across our platform. Twitter took a more nuclear option and said that it was blocking any URLs at all. It's still not entirely clear-- so, it's still not entirely clear what Facebook did at all on what basis it decided to down-rank it, how much it down-ranked it, why it decided that it was false. But it seemed to be an exceptional move that it was making, and so that's interesting. Twitter didn't explain its decision initially, then it appeared to be a fairly straightforward application of its rule about-- against posting personal information, which we can talk about more if we want to.

But then, when people got outraged about that, it sort of flip-flopped on that. And it created this thing where it's like, you have these policies, but you're not sticking to them, and you're not applying them, and you're moving away from them in certain circumstances. Now a lot of people have praised platforms quite a lot for their actions that they took here, quick responses and sort of prevented this from becoming a big story.

But on the other hand, I think that they also created the second meta-narrative that we're seeing play out against bias against conservatives and the fact that they departed from their policies in these particular cases. And so what I would really like to see over the coming weeks is really that platforms should tie themselves to a mast and say, here are our policies, and try and stick to them as much as possible and avoid the siren calls of public outrage or sort of-- or literal telephone calls from candidates or otherwise that pressure them into taking responses, because I think that otherwise run into a situation where you can win the battle of certain piece of information potentially blocking that, but you could lose the war of this idea of creating trust in the information ecosystem more generally, and the idea that here is the playing field. Here our rules. Here's what we're going to do. And we're going to apply them.

Now I do acknowledge that that's going to have some-- there's going to be some bad edge cases and some hard calls. And so for example, just this morning, we can see this. Former attorney General Holder tweeted out, it's too late to use the mails, in given the Supreme Court's decision yesterday about what votes will be counted. And Twitter flagged this as potentially misleading. Now taking literally, it's too late to use the mails, is in breach of Twitter's civic integrity policy. And I think this is a hardcore. Reasonable minds can differ about what Twitter should do here. What I'm saying is it's a fairly robust application of its policy, and that Eric Holder could retweet-- could tweet another clarification of what he meant. And that maybe erring on the side of rigid application of the rules is going to, in the long run, be better than platforms getting to steeped in subjective judgments about intention of tweeters and things like that. But that's a hard call, and I'm sure many people will disagree with me there.

And I'm just going to close by saying all of this is excellent. And we've seen a lot of mobilization around the US election. But I think it's summed up by this tweet here, in the election, we'll create a hashtag, the US election. The whole world can't use the retweet button. There's a real thing here. The rest of the world's watching this going, hold on, what about us? And when will we get similar kinds of measures as well? So that is basically high level, what we've seen in the last month or so.

Thank you so much for that. Thank you for setting the stage as comprehensively as you did. I want to pick up-- before moving on, I want to pick up on one of your key points in your opening, which is about labeling. It's sort of, as you mentioned, a good non-binary option within the tool cut-- tool kit of content moderation that sort of centers on a leave up, take down paradigm. How effectively in the past have you seen labeling work to bolster public confidence in the information ecosystem?

This is just such a great question, right? We are so at the-- like I said, this started in May. And we are just in the very earliest days of experimenting with these options. And we don't know whether any of them work. We need independent research to see, what are the effects of these

labeling options? And we've sort of seen platforms experimenting with like, well, what if we make it this color as opposed to this color? Like it used to be blue, and now it's going to be red? Does that make a significant difference? We have no idea.

But it all makes us feel a little bit better that, oh, they put a label on that, so we're good, right? I think we have these intuitions about what platforms should do and what makes a difference. But I could be completely wrong. So as-- to your question, we don't know. Have these had effects? Have they increased people's-- do people have more and better information? I don't know, and I hope that there's a lot more research about that in the future.

Thank you. OK. I want to now turn to Julie, who's a member of the Facebook oversight board. And so the perspectives that in question I have for you, Julie, is can you talk a little bit about to the extent to which we can extrapolate lessons learned in the US since 2016 to other institutional and Democratic contexts?

Yes. Thank you very much, Oumou And hello, everyone. And thanks to the Berkman Klein for the invitation. I will share my screen. I don't have a very thorough presentation. It was mostly to help me keep track of my speech, because I talk a lot and can derail.

It looks like Julie may have dropped.

Lose Julie?

I think maybe we did. So I am going to maybe move forward with a few of my additional-- Oh, it looks like Julia's back. OK.

Sorry about that.

No problem. OK. Glad you're back.

Yes. Are you seeing my screen now?

Yes, yes, I do.

Good morning again. Hello. I think, indeed, that it's really a good time to talk about the lessons learned, and Evelyn has truly explained, what we could take-- what we could have taken from the 2016 experience. I-- like you said, I am currently one of the members, the inaugural members of the Facebook oversight board, which was launched in May, and has recently announced starting taking cases.

So I think the oversight board is probably one of those solutions, or at least has catalyzed some of the conversations around this big realization after the 2016 election that platforms do have a huge impact on our expression, and particularly the political and even the electoral conversation and expression. So that-- but that said, despite the fact that it has been mentioned as-- or it has been part of discussions around how we could make things better, our country, to what probably

many expect, probably that the oversight board is definitely not here to be a judge, and not in any way of the US election.

But rather, I think what's interesting and what will be interesting in the future is that we've seen Evelyn present all these interesting policies. And there is debate. You were asking a few minutes ago whether or not labeling does make a difference or even-- but the question the board, for instance, could ask is, does labeling respect-- is proportionate to respect-- proportionate enough to respect freedom of expression? Because ultimately, that's what all these policies are about, about giving less visibility to speech that could create havoc and chaos outside of the platforms. But at the same time, not limiting freedom of expression in general.

So yes, I think that 2016 has shown us we needed more clarity. And hopefully the oversight board will do that in general, again, bringing more clarity to the discussions around content moderation speech and the boundaries set-- the boundaries, sorry, that we set to do our speech on social media platforms.

One thing, which is worth noting, is that, although, I'm sure you're aware, we have a process. I mean, we can be referred cases by Facebook and by users whose content has been taken down. And although there is a normal period of 90 days for us to make a decision, nevertheless, there are possibilities of expedited reviews, sorry, including one very expedited review, which has been worked on by our co-chairs at the oversight board, which would allow us to make a decision in seven days.

So theoretically, in a normal process, we wouldn't be able to make any decisions directly in the aftermath of the presidential election, of the US presidential election. But in practice, there is the existence of this expedited review process. So I don't know who would use it, but it is something we would have to discuss, depending on the criteria we set, and obviously, the imminence of the threat that the said content poses to the offline world will certainly be very determinant for us to use this type of expedited review process.

So yes, I think it was worth noting to kind of soothe the disappointments around many who have expressed that before we make any decision, yes, 90 days-- 90 days would not allow us to make a decision in the aftermath of the election, but there is this expedited review process, which is out there. So we'll see if it's used or not.

So back to the solutions that the recent policy developments that platforms have put out. I think there are plenty examples that Evelyn mentioned. I also read the excellent work done by Mozilla also, who has tried to make a research and to assess basically all the policies that have been rolled out by platforms prior to the US election. And Evelyn, I would like to question this narrow effort. There have been elections since 2016, and I'm sure there could have been also lessons learned even for the US election in those previous instances of elections outside of the US. And I think it's important because like a lot of things on platforms, in general, we have seen both bad and good uses being first tried, tested outside of the US, or even outside of the EU, if we are to talk about the global North.

There have been some case studies and practices that were rolled out before elsewhere, and we have seen the impact that they could have. So that's a little disappointment that I have with many of the policies that have been rolled out. Because obviously, like we've seen, there have been lots of efforts. The platforms have been really rolling out policies every day almost, or at least once a week since early this year. So-- but the deep devel is definitely into the details. And in many of these policies, the details will make the difference.

I will share an example-- and share two examples actually to show the importance of-- especially when we're talking about global platforms. As a reminder, a platform like Facebook has 70% of its users who are outside of the United States. That doesn't mean that the United States is not important. Of course it's super important. And as I said, at events in the US have input and influence on what will happen to the rest of the world.

But it's also interesting, again, to look at what's happening elsewhere in order to be better prepared. That's really the idea that I'm trying to share here. And an example-- and I hope you will also answer to one question that was sent to the panelists. Sorry. One of the examples is this issue of early victory claims or false victory claims before official results are out. This is a practice that has been widespread, especially on social media. I would even say especially since social media are available.

And in many parts of the world, and particularly in sub-Saharan Africa, it has become indeed a way for opposition candidates in environments where the electoral process is at the hands of a powerful president to kind of gain control of the narrative and highlight the potential irregularities that might have marred the process. But we see that, for instance, platforms like, I think, Facebook, Instagram, I can't remember the others, but some platforms have said that they wouldn't take down this type of speech, because it doesn't violate their community standards. And others have said they would take them down. I think Twitter has said it would take them down.

But it's interesting, for instance, could we have a sort of ratione temporis thought about this? Does it make a difference when the election results are out? When you make a misled claim before the official results are out. I mean, that could be in good faith, a mistake that has been made, or over-enthusiasm. Let's be naive about things, right? But obviously, when the elections are out, probably things change. Things are-- the situation becomes different when you continue to make claims despite the fact that official results haven't been put out. Should there be a different reaction to this? I'm not sure platforms have had this conversation. But I think it would be interesting, this ratione temporis kind of a thinking around this. That was one example.

The other example that I was hoping to mention here, although it's not directly related to the policy that have been recently rolled out, but I think it's an interesting-- very interesting example to show how important it is to pay attention to details, especially the ones that are not in our direct focus. It's the case of the Nancy Pelosi manipulated video. Suddenly, early 2019, we started having discussions in the West about-- not early '19, I think between 2019-- about what to do with when you have this type of content. And suddenly platforms woke up to the fact that, yes, people can share manipulated media on our platforms.

But had the platforms payed a little bit attention outside and a bit far from Washington or from the Silicone Valley, they would have seen that there was a country in which discussion around Deepfake has had huge consequences offline, and particularly has caused the second coup attempt in the history of the country. And that country is Gabon. It's located in West Africa, or Central Africa if you're a Francophone thinker like me. This country, yes, has witnessed a coup attempt because a video aired on Facebook Live. A video address of the president aired on Facebook live did not convince people that it was really not manipulated, and has led to people within the entry of the president staging a coup to protest against these images. So had the platforms, again, given a thought about or at least been open about this, probably they would have had this conversation even before the Nancy Pelosi case came out.

So I think it's really important that we pass-- Yeah, I think we're at a maturity stage, I hope, for the platforms to think that we have had this time where we are reactive, and we have to become very agitated when something big is going to happen in the US, because that's really the pattern, or in the EU, for that matter. But I think it's unfortunate, especially since many of these platforms are repeating all the time that they're global. Yes, you are global, and there has to be consequences to that globality. And one of these consequences is paying attention also to your user who are beyond the frontiers order of the United States, of Canada, and other places in the world. But I'm focusing today on the US.

So yes, these are some of the thought that I was hoping to share. I'm just checking if I haven't forgotten anything. No, I haven't. And I haven't derailed the conversation. I'm so happy about it. As a conclusion, I would say I completely agree with Evelyn. There has been huge improvements compared to if we had had this conversation in 2016. But the devil is definitely in the details and the actionability and the operationability of this new platform that are rolled out in-- yeah, in a kind of emergency mode could have-- could be better prepared. And we could have information as to does it work or not, if we were more innovative, if we were-- if we also tested the good things. The bad things are tested elsewhere in the world. We've seen Cambridge Analytica. It's very well. I mean, their test is one among the perfect ones. But it would be interesting, also, to test the good things, and see how they, yes, what kind of impact they can have. Thank you.

Thank you, Julie. I want to ask a question of both of you. You're both from outside of the US, and so you have a more global perspective on content moderation issues. I think after 2016, the platforms learned pretty quickly how to detect and deal with disinformation from foreign actors. But as we've seen over the past couple of weeks, and certainly over the last couple of years, that it's not necessarily the case when it comes to domestic disinformation, especially when the purveyor of that domestic disinformation is within the US government, including at the highest levels of the US government.

Do you think there are lessons learned when it comes to disinfo from elected officials that the US-- or that the US can learn from? Bearing in mind a significant sort of public pressure, platforms often come under for having to make difficult content moderation decisions. Maybe I'll start with you Julie.

Yes, well, it has always been a kind of criticism against the platforms, that they tend to have a heavier-- no, they tend to apply their policies with more severity when it comes to outside leaders, and particularly leaders from southern-- global southern countries, especially those that are not in very good terms with global north countries, including Iranian officials and Russian officials. And even on this issue of labeling of media control-- sorry, state controlled media, we've seen that being rolled out against many Russian state-owned comp-- sorry, media companies.

But honestly, the debate could also be had-- we could also have this debate over many other northern media, for instance. I know a lot about France 24, which is the French state-funded media. And that has-- that he's highly influential, especially in sub-Saharan Africa. Should they be labeled? That's a question, for instance, that we've received. So I think it's really also time that there is sort of more coherence and consistency in order to prevent these platforms to be seen only in some cases, only as an arm or influence of northern countries, especially at a time when we know that there is this splinternet around-- this idea that the internet is going to be split between those who will have a more controlled version of it, like a Chinese version, or those that have a less controlled. Although, I don't know, probably that frontier is narrowing given the TikTok case. We don't know.

But yes, I think it's important to keep this in mind, because the ultimate consequence is those platforms may end up being even blocked in these other countries based on this perceived inconsistency, or at least less severity towards those governments from Western-- Western nations, in particularly the US. So yes, I think it's important to have this in mind. I don't have a solution, particularly on that, but I just wanted to bring this perspective.

Thank you. Thank you. Same question to Evelyn. Do you think there are lessons to be learned from other countries when it comes to dealing with disinformation from elected officials?

So I think there's two issues here. There's the-- I mean, the official position of all platforms is basically, we don't treat foreign or domestic disinformation any differently. Quite clearly, that's not the case. There's sort of two types of buckets here. There's the disinformation of an elected official comes out and says something false. And that's sort of what I spent the most of the presentation talking about. I mean, I think that-- I don't know that any country-- I don't think any platform does this particularly well in any country. I think that my solution is just apply the rule.

But the second-- So, the second bucket is this sort of like-- the thing that I've flagged earlier about coordinated inauthentic behavior, and this kind of like influence campaign. That is not like someone comes out and says something that's disinformation, but the idea that some of the content may not even be particularly problematic, but it's being done in non-transparent ways to manipulate audiences and influence them.

And I think this was an area where I'd have a lot of problems with this coordinated inauthentic behavior. And I think that it was an area where foreign speech was scapegoated quite a lot. There was a lot of focus on foreign threats, and the Russians are coming kind of framing, which distracted from, I think, a more fundamental question of, what are appropriate standards of coordination online and activity? Like where do you draw the line between a legitimate political

activism grassroots campaigning, marketing behavior and something that just steps over the line and is suddenly trolls. And I think that we sort of fastened on the foreign part of that as a nice sort of simple way of doing that.

But then we're confronting more and more as domestic actors do a lot of things that sure look like what the IRA did in 2016. What do we do about that? And we don't have-- the platforms get nervous, understandably, because then it becomes like they are getting involved in domestic politics. But I think they need to, again, have much more transparent rules and clearer standards, because their platforms are the platforms that create the incentives to engage in this kind of behavior, and reward that kind of behavior as well with the amplification and the engagement that it gives it. So I think much more transparent standards.

And it's also on the rest of us. We need to have a much bigger societal conversation about where do we want to draw the lines around online activism. And what do we think is acceptable, and what is not. And I think that fastening on the foreign part of that is not going to help us advance that conversation.

Thanks so much. OK, so we have a few audience questions coming through. The first one looks like it's for Evelyn. Can you talk a little bit about the efficacy of fact-checking, and sort of the way you talked about the efficacy of labeling, does fact checking helped to bolster confidence in the online information space or to sort of bolst confidence in nonpartisan official sources when it comes to elections?

So again, fact-checking is another one of those ones that feels good and is a great intuitive answer to a lot of things, and everyone really likes it. Who can be against fact-checking? Well, apparently a lot of people. But in general and in theory, fact checking sounds great. And I am in favor of it. I think it's really good, and I think it's something where there should be a lot more resources. Fact-checkers are chronically under-resourced, especially in other parts of the world and in non-English speaking languages-- not English-speaking-- non-English languages. And so there should be a lot more resources and support for those. And I think that's very, very true.

It is also true, first that we need-- there is research that shows that it works. And it can correct certain kinds of belief. But there's also research that shows it can have some counter-intuitive results. So for example, some research showing that if you apply fact checks to certain stuff and leave other stuff un-fact checked, there's an implied truth effect, that people think that something that isn't flagged is more likely to be true just because it doesn't have a label attached, which might not be the case at all. It's just that the fact checkers haven't apply-- haven't checked that one.

And in a world where there's billions of claims and only so many fact checkers, that's a problem. So we need a lot more research about what works, what kind of fact checks work, and ways to sort of triage the most important claims and things like that. But also, I think it's another one of those things. It's not going to be a panacea, and it's not going to fix everything.

I was speaking to Maria Ressa recently, a journalist from the Philippines, who has been the subject and the target of a bunch of-- a lot of disinformation and trolling campaigns with the

most awful effects. But she runs a fact checking partner with Facebook in the Philippines. And I was asking her, why are you doing this? You know, Facebook-- And she's like, well, you know it's better than nothing, but it is-- and her phrase was, a thinking slow solution for a thinking fast world. Right? It's a very tiny piece of the puzzle. And it's a good one. But we need much more systemic change and thinking about things like friction and stuff like that if we're really going to tackle the bigger underlying problem.

So as with everything in this space, people-- we have so many problems. We're going to need so many solutions. There's not ever going to be one thing that fixes everything, and I think that applies to fact checking as well.

So in the vein of implementing interventions that are sort of more scalable globally, we have an audience question, and I'll pose this to Julie first with the context that just last week the opposition party in Guinea, a country in West Africa, claimed victory on Facebook and Twitter before official results had been announced. Do you think that Facebook rules should be global? Is that practical and is that desirable?

There should be principles. I think it's, again, a question of clarity. Freedom of expression doesn't change suddenly depending where you are in the world. The foundations, the principles are basically the same everywhere. But now what is clearly missing, and I think we haven't talked about that much is the granularity. What is the context. We refer that to that a lot. And particularly on the issue of early victory claim, especially in authoritarian regimes, it would have a huge consequence if suddenly the platform decided to delete the tweet from an opposition leader who's trying to bring in democracy, or at least with presenting himself or herself as a Democratic alternative.

Rather, I think what's important to do, and it's something that is really dear to my heart and that I've been working on a lot while at the Berkman Klein is, you have the best expertise ever, you will ever look for in those countries. [INAUDIBLE] the human rights defenders, the activists, the entire regime activists, who also have done investigations, including on potentially rigged elections. So it's really important and essential for these policies to be fully integrated into a big ensemble in which the context will be able-- would be at the disposal, basically, of these platforms.

And that's not necessarily always the case. Or even when there is a little channel of communication between these platforms and local organizations and in civil society experts. First of all, we don't really know how the platforms are using whatever information they're receiving from these partnerships. I think it would be essential. My organization, for instance, as [INAUDIBLE] chair, has had to do some work with platforms, all of them, not only Facebook. And we actually don't know how efficient our-- what we bring in is. We think it is because we see a difference on the field. But obviously, we don't have the data, which obviously, platforms have. It would be interesting to have that.

So yes, I think it's part of-- it should be part of the equation. And it's definitely one aspect of how efficient a policy is going to be, is how integrated it is to the reality in which you are deploying it.

Yeah. We've talked a lot about the limitations of interventions that we've already seen so far on premature claims of victory, on labeling, on fact checking. And one thing that's really emerged through conversations that I've had with both of you over the course of the last several months is that a whole of society approach is really needed to cut-- to tackle disinformation in a comprehensive way. What are the roles of some of the other components, the other pieces of the puzzle, civil society groups, governments. And I'll pose that question to you first, Julie.

Yes, governments have this, and I would even say a primary role, because this is-- the fight against disinformation is not just a product of platforms. It's really a question of our democracies, of our human rights, of the rule of law. So and that the primary responsible to make sure that we have all this is obviously the government. So it's the government's responsibility to make sure that the citizens have not only access to the information, but are able to read that information.

I remember very well a program that was rolled out in high schools in France, which was quite efficient. It was a program in which disinformation experts would go to these high schools and work with students for, I think, one trimester-- I think the word exists in English-- one trimester trying to share with them some of the methodologies that you can use to doubt things. Basically, it's a Cartesian country, so you have to doubt when you've received information. So yes, I think that the government has a role to play.

But obviously, a civil society organization are essential. First of all, because we wouldn't have all this conversation if we didn't have civil society organization researchers, who have been doing the work on alerting about what we are seeing now. Nobody listened to them. Now it's become a big thing, thank-- fortunately. But it's good to have them in the loop as well.

And I'll give another very good example of that related to another subject that's not directly related to disinformation, which is hate-speech. We've seen so many governments suddenly waking up to this realization that people use platforms also to share very offensive speech. So they just suddenly decide to find platforms and think that platforms have to over-censor. But in the case of France, where a similar legislation such as the one in Germany was adopted, the constitutional judge basically said that first of all, all these mercenaries were not proportionate, were not necessary.

There were other ways to deal with that, including working with judiciary authorities in the countries to define these boundaries around hate and doubted speech. But most importantly, criticizing also that dialogue only between platforms and governments that leads to censorship basically. And that's the risk. In my opinion, civil society is really-- brings the balance, basically, between all these two actors, these two set of actors, which to my opinion, don't necessarily have freedom of expression really at the core of their preoccupations all the time.

Brilliant answer. Thank you. Evelyn, same question to you.

Yeah, I mean, what can I add to that? I think that that's great. Julie has the expertise there, so I really don't have much to add. And I mean, the only thing I'm going to add, and it's very small, and I wanted to say that this is only a slight additional thing that should not take away from the importance of platforms doing far more and the importance of civil society and the absolute

fundamental importance of governments, just not purvey disinformation, that's like, if we could all just not lie, that'd be a really good start.

But I think there is something that we all could do as well. I think we all should play our part in this process as well. I'm always a little bit shocked and surprised by people who like a disparity about the state of the online ecosystem, and then smash the retweet button on spurious claims when they like them. And I do think I am guilty of it too, right? We see something, and we want to sort of endorse and signal with it. And I think if we all sort of try and muddle-- be the Twitter that you want to see in the world. That's a good way of sort of proceeding as well. To try and be good actors in particularly over the next week and couple of weeks here in America. To try and make sure you're checking things before you sort of spread them and amplify them is this is a small thing that we can all do as well.

Yeah. Yeah. Thank you. I have one final question for you both. And I want to just step back and think very, very big picture. We, at the time record-- recording this-- making this recording, we're about one week, exactly one week out from November 3. I don't know about you, but I am very scared. To both of you, pick one platform, and think about all of the worst case scenarios that have been running through your mind over the past couple of weeks and months. And think about what's as one change of perspective, one intervention, or one policy you want to see the platforms adopt to sort of mitigate what you've ascertained is the worst case scenario in your mind. And whoever would like to go first can jump in.

I'll jump in. So the first part of the question, the worst case scenario. I've been thinking about, sorry, artificially generated video showing voter rigging in a random, how do you say that, voting station. Yes, how do you treat such claims, especially before the official results have come out, and yeah. And what was the second part of the question? How-- what's the best response--

[INTERPOSING VOICES]

The best response to that, definitely, in my opinion, would be first of all, for platforms to be also connected with local-- I am not really familiar with the hierarchy, but the voting, I'm sure there are authorities gathering the votes for states. And I don't know what's the equivalent. It's kind of electoral committee--

At the state level.

Sorry?

There's state and local election officials election. Election and--

OK. Thank you. State and local election officials. I think it would be, yes, one of the best case scenarios to kind of be prepared to counter this as soon as it comes out, being in direct connection with those officials and getting real time results and information on how the counting is going.

Yeah. So maybe with like a link to the most authoritative non-partisan source.

Yes, diversely, not the website only, but internally directing between those two.

Thank you. OK, Evelyn, same question to you. One platform, one intervention that you'd like to see, and how it sort of allows us to avoid the worst case scenario.

OK, so I'm going to give a good answer twice, because I think I've already answered it the first time, which is like my simplest fix. The one thing I want is I want YouTube to have a policy about false early claims of election victory. I really hope I'm wrong about this, but the nightmare scenario that I'm seeing is that on November 3, some candidate, not naming any names, is going to just do a live stream of like, we won the election and everything's over, it's all great. And YouTube doesn't currently have a policy for doing anything about that. So that's my one, very simple fix. That's what I want. That's my one big ask. If I could rub a genie lamp and have a wish come out, that would be it.

So but my other wish then, is for all of them, my question is really around enforcement. And can they enforce these policies, and are they confident that they can enforce them. And if they're not confident that they can enforce them effectively enough or quickly enough, for example, labels on misleading information about voting and things like that, then what I would like to see them do is, someone mentioned this in one of the questions or something, about like introducing a pre-review of posts or tweets. I don't want that across all platforms for every post. I think that would be an infringement of freedom of expression. But if you have repeat offenders that repeatedly breach rules and that they have a history of posting certain kind of misinformation, and you're not confident that you can attach a label within minutes or half an hour of that tweet or post going up, then I would introduce some sort of trip wire or pre-review policy for repeat offenders so that those labels can be effectively applied before they're seen by tens of thousands of people.

Thank you. And for repeat offenders, those would ideally include elected officials like the President or other GOP officials?

Yep.

Great. Well, that concludes my set of questions. I want to thank you both for a really robust, comprehensive, and enlightening discussion. Thank you to our audience for joining. Thank you to those who submitted questions. Have a great day.

Thank you

Thanks.