February 1, 2021

Brenda Victoria Castillo
President & CEO
National Hispanic Media Coalition

Dear Brenda,

Thank you for your letter about how Facebook enforces against hate speech, content that may incite violence and misinformation, especially where content targets the Spanish speaking community.

I'd like to take this opportunity to provide details on our content policies and the way we enforce on those policies. We have policies against statements that may incite violence, and we understand how important it is to get enforcement right in all languages, especially when there's a risk that content has the potential to contribute to real world harm.

In enforcing these policies, we use proactive detection technology that looks specifically for content that might violate our violence and incitement policies. We also proactively search for Facebook Events that may violate our rules. Training artificial intelligence to identify content that has the potential to incite violence is difficult because context is so important. That's why we supplement our enforcement efforts with expert review teams who often serve as a check on content that is surfaced by our technology and tools.

And after the violent events of January 6, in which the Capitol building was stormed, we adopted additional emergency measures to more aggressively identify and remove content that violates our policies, including some forms of misinformation, that have the potential to contribute to incitement and violence. Specifically, please note the following:

- We are now removing content on Facebook and Instagram that contains the phrase "stop the steal" under our Coordinating Harm policy. We removed the original Stop the Steal group in November and have

FACEBOOK

continued to remove Pages, groups and events that violate any of our policies, including calls for violence.

- We've been allowing robust conversations related to the election outcome and that will continue. But with continued attempts to organize events against the outcome of the US presidential election that can lead to violence, and use of the term "stop the steal" by those involved in the violence in DC, we took this additional step in the lead up to the inauguration and removed a significant number of posts.
- We've increased the number of Groups that require admins to review and approve posts before they can go up, automatically disabled comments on posts in Groups that start to have a high rate of hate speech or content that incites violence, and continue to use AI to further demote content that likely violates our policies.

You asked specifically about a post published by "Directo y Sin Censura," a Facebook Page. At the time the post was published, it did not, on its own, violate our policies; however, as I've mentioned, the situation on the ground changed drastically following January 6, and our assessment of the post thereafter would have taken into account the totality of circumstances. We will continue to stay vigilant to additional threats and take further action if necessary to keep people safe and informed.

With regards to your concerns about misinformation, particularly Spanish-language misinformation, we take a multi-pronged approach. We remove accounts and content that are against our Community Standards:

- For example, we have cracked down on fake accounts which are often the ones spreading misinformation;
- We remove misinformation that has the potential to contribute to imminent violence or physical harm, including misinformation about COVID-19;
- We remove misrepresentations that would interfere with a person's ability to vote; and for content that doesn't violate our community standards, we work with independent third-party fact checkers who have discretion as to which specific pieces of content to review. When content is debunked, we significantly reduce its distribution and add informational labels to provide more context for those who do see it.

FACEBOOK

Three of our US fact-checking partners, AFP, AP and Reuters, review content in both English and Spanish. They've debunked many claims related to the elections and leading up to it including: claims that Biden is the candidate of the Chavismo, a photo showing Puerto Rican senator accompanied by 'militia' at pro-Trump event, an image showing outdated voting tallies purporting to prove voter fraud in Wisconsin, an image purporting to show a group of Native American individuals riding toward voting centers in the United States, a false account of the president of Cuba that shows a post in support of Biden, and a video falsely showing a Trump supporter confronting Biden. I can also confirm that Spanish language claims about election fraud and the Stop The Steal movement were fact-checked by our partners. As with all claims our partners debunk, we reduced the distribution of these posts in News Feed and applied a warning label on top of them on Facebook and Instagram so people have additional context.

WhatsApp also took a number of actions to prioritize connecting people with accurate information. In September WhatsApp partnered with IFCN to enable the launch of FactChat, a chatbot that unites 10 U.S. fact-checking organizations with two major Spanish-language broadcasters (Telemundo and Univision) aiming to combat misinformation and disinformation leading up to the election. According to IFCN, FactChat sent over 500,000 messages in the six weeks prior to the US elections and compiled more than 620 fact checked articles. Over 150 of these were published in Spanish. WhatsApp users are also empowered with a 'search the web' feature to quickly search for more information about highly-forwarded messages they may have received so they can reach authoritative sources.

WhatsApp systems work 24/7 to prevent bulk and automated messaging. WhatsApp bans over 2 million accounts per month for bulk messaging behavior; 75 percent of these happen without a recent user report, which means our automated systems stop abuse before users can report them. And of course, whenever a user receives a message from someone outside their address book, we surface the ability to report those messages directly inside the chat.

WhatsApp also has two different forward limits - the default is set to limit forwards to 5 chats at a time. This reduced the number of forwarded messages on WhatsApp by over 25%. For messages "forwarded many times," meaning they are at least five forwards away from the original sender, we limit forwarding to 1 chat at a time. This change reduced these kinds of messages by over 70%.

FACEBOOK

I want to thank you again for your advocacy and leadership on these issues. The concerns you raise are of utmost importance and we look forward to working together to ensure that we continue to make progress.

Kevin Martin
Vice President for U.S. Public Policy

Cc: Center for American Progress
       Color of Change
       Free Press
       Global Project Against Hate and Extremism
       Hispanic Association of Colleges and Universities (HACU)
       Hispanics in Philanthropy
       LatinoJusticePRLDEF
       League of United Latin American Citizens (LULAC)
       MediaJustice
       Mi Familia Vota
       National Association of Latino Arts and Cultures
       National Latinx Psychological Association
       The TransLatin@ Coalition
       UltraViolet
       UnidosUS
       United States Hispanic Leadership Institute, Inc.
       Voto Latino