The Washington Post

A diminished Twitter faces a World Cup test for hate speech moderation

By Naomi Nix and Rachel Lerman
November 19, 2022

Advocates warn that chaos within Twitter could worsen the spread of hate speech during the tournament.

Last year, Twitter sprinted to remove a deluge of racist tweets targeting three Black players being blamed for England's loss in soccer's Euro 2020 championship, prompting world leaders and activists to blame social media for amplifying bigotry.

Now, hate speech advocates are warning that the spread of racism on Twitter could be even worse during this year's World Cup as the social media giant grapples with a mass exodus of employees and the tumultuous leadership style of new owner Elon Musk.

Even before Musk laid off half of Twitter's staff, his acquisition of Twitter had opened the door to a flood of racist and antisemitic trolls seeking to test Twitter's content moderation practices under a new owner who for months had pledged to support free speech. Twitter has fewer staffers to address an influx of rule-breaking tweets, bigoted hashtags that spread rapidly or misinformation targeting the soccer competition.

And many teams critical to keeping the site running smoothly are down to few or no engineers after a Musk ultimatum that spurred hundreds of employees to leave the company.

"This will be a moment when people around the world will be seeing the footballers they love that their whole nations are sort of pulling for suddenly being open to the most vituperous abuse on platforms that have a terrible track record of enforcing their rules on racism," said Imran Ahmed, the chief executive of the Center for Countering Digital Hate.

The World Cup, in which teams from 32 nations are scheduled to compete in Qatar, is likely to add to public scrutiny of Musk as he seeks to balance his support for free expression on Twitter with his interest in placating civil rights activists and advertisers concerned that the platform will become overwhelmed with offensive content. On Friday, Musk encouraged his 116 million followers to turn to Twitter for the "best coverage & real-time commentary" about the first match on Sunday.

Experts say Twitter, and other social media companies, will be forced to make tough calls about which posts to remove or leave up in real time as millions of soccer fans watch the games and turn to social media for commentary simultaneously. The sheer number of games over the next few weeks

coupled with the wide range of language and cultural expertise needed to make content moderation decisions may make the World Cup especially difficult for the platforms to police, experts said.

Other analysts point out that tensions around race and ethnicity may run higher around the games because Russia's invasion of Ukraine has led to a rise in anti-refugee rhetoric in Eastern Europe.

Twitter, whose communications team was largely cut during the layoffs, did not respond to a request for comment on its World Cup moderation plans. Twitter executive Ella Irwin, who was recently appointed the new head of the company's trust and safety team, tweeted this week that employees have been preparing for the World Cup for weeks.

"Ensuring a healthy platform continues to be our priority," she tweeted.

Musk weighed in on Friday, pledging to limit the distribution of racist tweets. "New Twitter policy is freedom of speech, but not freedom of reach," he tweeted. "Negative/hate tweets will be max deboosted & demonetized, so no ads or other revenue to Twitter. You won't find the tweet unless you specifically seek it out, which is no different from rest of internet."

Since Musk took over Twitter, civil society organizations and civil rights activists have increasingly been pushing the new CEO to be more aggressive in his commitment to fight misleading or hateful content on the platform as well as maintain the staffing levels required to implement the platform's rules.

Kick It Out, an anti-discrimination group that works with soccer organizations, published an open letter earlier this week to Musk and Meta chief executive Mark Zuckerberg, asking them to take bolder actions to curb online abuse. In an interview, Kick It Out Chair Sanjay Bhandari argued that the perception of Musk as a free-speech absolutist has been used as an excuse by some users to spread hate online, something that could be more problematic as Twitter's content moderation team shrinks.

"Some people have heard that as a dog whistle for racism and hate," he said. "When you combine those factors together that's a toxic cocktail and I fear what we will have in the World Cup."

On Tuesday, more than 40 civil society groups urged Musk in a letter to invest in "appropriate global resources" to curb the spread of hateful, misleading and violent content. The Global Alliance Against Digital Hate and Extremism coalition argued that Twitter has often not paid enough attention to protecting public discourse in countries outside of the United States or the European Union, "which has led to the spread of online disinformation and hate, and spurred violent extremism across the globe."

That letter followed a call by a coalition of more than 60 civil rights groups for top advertisers to suspend their marketing spending on Twitter in protest of Musk's decision to lay off thousands of employees, arguing that the company will be less equipped to fight problematic posts on its platform.

Musk has sought to address advertisers' concerns by reiterating the company hasn't made any changes to its content moderation policies, which bar users from promoting violence or attacking people on the basis of their race, sexual orientation, religion or other sensitive attributes.

But even before Musk's takeover, Twitter and other social media companies had a mixed track record addressing racism aimed at soccer players, especially Black athletes. After England lost to Italy in the Euro 2020 final last year, trolls and angry fans online spewed racist abuse at three Black athletes on the team including by posting monkey and banana emoji and other bigoted comments underneath photographs of the players on their personal Instagram accounts.

Then-Prime Minister Boris Johnson condemned the abuse, and social media companies announced new efforts to fight racism on their networks. Instagram said at the time it was working with U.K. law enforcement agencies and adding harsher punishment for accounts that sent discriminatory messages in private chats. Twitter said it removed nearly 2,000 tweets in the days following the final and was working on improving its detection of racist abuse online.

Activists say hate speech targeting soccer players on social media persists. FIFA, the international soccer governing body that oversees the World Cup, said in a report published earlier this year that more than 55 percent of players in the finals for the Euro in July 2021 and the Africa Cup of Nations this past February received some form of discriminatory abuse on Twitter or Instagram, which were most often homophobic slurs and racism. More than half of those posts were still live on the platform in April, the report found.

“I think it shows that sports can produce this kind of emotion and this kind of reaction,” said Rafal Pankowski, who works with the Polish-based anti-racism group Never Again and advises soccer groups. “In the past, it was probably similar, but people could just shout at their TV set. And now with Twitter, they have a tool to amplify …[racism] on a global scale, and there’s no mechanism of dealing with it.”

To fight the abuse, FIFA announced it would give players in Qatar access to a monitoring service that tries to filter hate speech targeting them.

The chaotic beginning to Musk’s tenure at Twitter, has some activists worried the problem could grow worse. In the hours of after Musk took over, a slew of anonymous trolls spewed racist slurs and Nazi memes onto Twitter.

“Elon now claims not to have changed the rules, but he certainly sent a bat signal up to every single racist out there that you know ‘we’re open for business,’” said Ahmed.

Meanwhile, Musk laid off roughly 50 percent of the company’s workers earlier this month including cuts to Twitter’s curation team — a group that is central to the company’s efforts to guide users to reliable news sources and tamp down on viral hoaxes and conspiracy theories. Tech publication Rest of World reported last week Twitter laid off significant parts of its international teams in India, Africa and Latin America. Last week, Yoel Roth, the company’s head of moderation and safety who had been reassuring users and advertisers about Twitter’s policies, resigned.

“With those teams changing, we don’t know if there will be enough continuity, if there will be enough measures for them to effectively step in and prevent the hateful content that comes out around the World Cup,” said Pinar Yildirim, a Wharton professor who studies media and technology.

https://www.washingtonpost.com/technology/2022/11/19/twitter-world-cup-hate-speech/