LRC Legal Resources Centre

global witness

BRIEFING June 2023

"WE NEED TO KILL THEM": XENOPHOBIC HATE SPEECH APPROVED BY FACEBOOK, TIKTOK AND YOUTUBE

Violent and extreme hate targeting migrants and refugees in South Africa approved for publication by social media platforms

In the face of imminent danger, we'd do anything to protect ourselves and our families. This decision is a daily reality for thousands of refugees and migrants fleeing sub-Saharan countries who are forced to uproot their lives and flee poverty, starvation, civil unrest, rape, anti-LGBTQI+ laws and killing. With its progressive refugee protection policies, laws, and Constitution, many are led to believe that South Africa provides hope of safety. However, instead of finding the refuge they desperately seek, many refugees and migrants find themselves subjected to a new form of conflict.

UN experts have expressed concerns that South Africa is "on the precipice of explosive xenophobic violence". They cited campaigns like Operation Dudula which first emerged online and which became a catalysts for real-world outbreaks, fuelling violent protests, acts of vigilantism, arson, and murder. The African Centre for Migration & Society has created a Xenowatch Tracker where they estimate that since 2014 there have been 565 incidents, 15,093 displacements, 3,040 shops looted and 218 deaths. Furthermore, they note a steady increase in these incidents over the past few years.

February 2022: Members of Alexandra Dudula Movement during their operation to remove migrant street vendors in Johannesburg. Credit: PHILL MAGAKOE / AFP via Getty Images

South Africa is "on the precipice of explosive xenophobic violence" according to UN experts

With xenophobic violence on the rise, we wanted to understand if social media platforms were doing all they could to protect vulnerable communities by enforcing their own policies on hate speech and incitement to violence. Global Witness and the Legal Resources Centre carried out a joint investigation looking at Facebook, TikTok and YouTube's ability to detect and remove real-world examples of xenophobic hate speech targeting refugees and migrants in South Africa. This included calls to shoot and kill, highly offensive racist slurs, comparisons to cockroaches, and hashtags from widely publicised campaigns such as Operation Dudula.

Rather than share the examples organically, we submitted the hate speech to all three platforms in the form of adverts, so they could be scheduled in the future and, importantly, removed before going live. The test saw 10 adverts in English and translated into Afrikaans with nine of those translated into Xhosa and Zulu. Hate speech was taken from real-world examples, edited to clarify language and grammar. None of the examples were coded or difficult to interpret, and all violated the platforms' advertising policies.

Every single ad was approved for publication by all three platforms, apart from one ad that Facebook rejected in English and Afrikaans, although this was approved in Xhosa and Zulu. After the platforms stated whether the ads were approved or rejected for publication, Global Witness and LRC deleted them all before they were published.

This isn't the first time the platforms have failed to enforce their own policies on hate speech. Since 2021, we've conducted the same investigation more than 10 times in Brazil, Ethiopia, Ireland, Kenya, Myanmar, Norway, and the USA. The results uncovered stark differences in the platforms' abilities to detect content and large divergences in how users around the world are treated.

March 2022: Clashes as members of Operation Dudula protest migrants in Johannesburg. Credit: Bloomberg via Getty Images

So, what's preventing the social media platforms from enforcing their own policies? Ultimately, a crucial step forward necessitates more equitable investment in the countries where they operate. The safety of migrants, refugees, and all individuals will remain compromised in South Africa or any other country until these platforms prioritise content moderation, conduct comprehensive risk assessments into human rights and societal impact, establish protective measures during crucial election periods, and embrace greater transparency and external accountability.

The Legal Resource Centre and Global Witness are joining a new coalition that is driving this change. Recently launched, the Global Coalition for Tech Justice unites civil society groups worldwide, rallying for a global call-to-action pressuring social media companies to equitably invest resources toward safeguarding the integrity of the upcoming 2024 elections. With time running out, it is essential that we harness collective power and prioritise impacted communities in this moment.

In response to Global Witness' investigation, a Meta spokesperson said: "These ads violate our policies and have been removed. Despite our ongoing investments, we know that there will be examples of things we miss or we take down in error, as both machines and people make mistakes. That's why ads can be reviewed multiple times, including once they go live."

A TikTok spokesperson said that they were investigating our findings and that their content moderators speak Afrikaans, Xhosa and Zulu. They said that hate has no place on TikTok, that their policies prohibit hate speech and that ad content passes through multiple levels of verification before receiving approval. [1]

Google was approached for comment but did not respond.

[1] The full response from TikTok was: “Hate has no place on TikTok. While these ads never made it onto our platform, our advertising policies, alongside our Community Guidelines, prohibit ad content that contains hate speech or hateful behavior. Ad content passes through multiple levels of verification before receiving approval, and we remove violative content. We have local Trust & Safety teams responsible for regional content policy creation, safety product development and content policy enforcement, who have a clear understanding of local nuances across Africa. To support our global platform, our moderation experts speak more than 70 languages and dialects, including Zulu, Afrikaans, Sesotho and Xhosa. We are expanding our safety function in our African markets in line with the continued growth of the TikTok community on the continent and continue to learn how we can improve our systems and processes based on feedback from organizations like Global Witness. More broadly, our Community Guidelines prohibit any violent threats, incitement to violence, or promotion of criminal activities that may harm people, animals, or property. They also prohibit any hateful behavior, hate speech, or promotion of hateful ideologies. This includes content that attacks a person or group because of protected attributes, including: Caste, Ethnicity, National Origin, Race, Religion, Tribe, and Immigration.”