

TikTok moderators struggling to assess Israel-Gaza content, Guardian told

Publication Date: 2023-12-20

Author: Hibaq Farah

Section: Technology

Tags: TikTok, Israel-Gaza war, Social media, news

Article URL: <https://www.theguardian.com/technology/2023/dec/20/tiktok-moderators-struggling-to-assess-israel-gaza-content-guardian-told>



TikTok moderators have struggled to assess content related to the Israel-Gaza conflict because the platform removed an internal tool for flagging videos in a foreign language, the Guardian has been told. The change has meant moderators in Europe cannot flag that they do not understand foreign-language videos, for example, in Arabic and Hebrew, which are understood to be appearing more frequently in video queues. The Guardian was told that moderators hired to work in English previously had access to a button to state that a video or post was not in their language. Internal documents seen by the Guardian show the button was called “not my language”, or “foreign language”. It appeared under a section that says “escalate” and once clicked, it is understood that moderators were not able to apply any policies to it. Moderators assumed that using the button would send the video to other colleagues who understood the language. However, about six to nine months ago the button was removed without explanation, according to a TikTok source. In recent examples seen by the Guardian, the button does not seem to appear in videos, including a foreign-language video. This has meant that moderators have often had to rely solely on what they see as at times they cannot understand the words they hear. In these instances, moderators can only apply policies to obvious, visible signs of a guideline breach. It has been alleged that when moderators raised these concerns, the advice given was to put the words through Google Translate, if possible. Moderators often have to hit targets related to how quickly they moderate – and using Google Translate is time-consuming. In one video seen by the Guardian that appeared in a queue for English-language moderators, a song plays in the background with what sounds like Arabic lyrics, in front of a Hezbollah flag representing the Lebanese militant group; the video has Arabic words rolling through the video. The video does not have the option to click a “not my language” or “foreign language” button. In a press release detailing its moderation strategy for the Israel-Hamas war, TikTok said it had launched a command centre, with “key members” who represent a “range of expertise and regional perspectives”. TikTok argues that it employs thousands of moderators globally who are provided with extensive training and a range of tools to keep the platform safe. It also points to an increase in moderation sources in Hebrew and Arabic since the conflict started and adds that it has removed more than 1.3m videos in the conflict region for violating content guidelines. The Guardian has been investigating TikTok amid ongoing concern about how it moderates its users, who number more than 1 billion users worldwide. The Guardian has seen internal communications that are likely to raise fresh questions about how the app is policed. The allegations of the language button disappearing come as TikTok is reportedly preparing to launch a number of initiatives to tackle complaints over coverage of the Israel-Gaza conflict on its platform. According to the tech website The Information, the plans under the “Swipe Out Hate”

campaign include establishing a taskforce to deal with discriminatory posts. Moderators were given an internal update on guidelines for the Israel-Hamas conflict in October in which they were told that certain Hamas and Hezbollah-related content was allowed, provided it did not incite violence or praise Hamas. The update said: "The following are allowed content about Hamas and Hezbollah: Content that shows Hamas or Hezbollah political leaders as they are discussing the broader Israel-Palestine conflict and historical context for ongoing events, that does not also (1) praise, incite or threaten violence and (2) that does [NOT] praise Hamas. This allowance does not apply to [Hamas military wing] Al Qassam militant leaders." Citing another example of permitted Hamas or Hezbollah-related content, the TikTok guidance allows "Praise of Hamas or Hezbollah only in the context of discussion that focuses on their administrative, political, and social welfare body without mention of their violent acts." But "praise of any other designated organization, including Al-Qassam, should be tagged with Promotion of Terrorism (International), regardless of the actions or focus discussed". Another example, the platform says, where the "Hamas or Hezbollah flag or insignia is not a violation" is in "depiction or discussion of Palestinian suffering with the Hamas or Hezbollah flag in the background, such as funerals, demolished buildings etc, as long as there is no praise of violent acts". There has allegedly been confusion among moderators on the conflation between Hamas and Hezbollah, and the Palestine flag. Moderators in English queues are seeing a sharp increase in the number of videos relating to the Israel-Gaza situation, and an increase in Chinese-language scam videos that seem to be promoting jobs. There have been no updates on how to moderate these. In recent months, it is understood that moderators hired to moderate in English have seen an increase in the number of videos relating to Israel and Gaza, in both Arabic and Hebrew, but particularly more videos in Arabic, and from Palestinians. It is understood that some videos are of Palestinians speaking in front of collapsed buildings and searching through rubble for people to rescue. In some cases, moderators have had to look at videos of Palestinians released from captivity and then try to guess if what is being said would need a policy applied, despite not understanding what is being spoken. In these instances, it is claimed that moderators have had to guess whether a video they have been asked to check appears to be praising Hamas, for instance, or discussing a totally unrelated issue. One moderator at TikTok said: "How am I meant to know who the bad guys and the good guys are if I don't speak the language?" "I wouldn't be surprised if the button disappearing could lead to moderators incorrectly interpreting video based on their own unconscious biases or anti-Islam rhetoric, for example." The platform was accused recently of pushing more pro-Palestinian content to US users than pro-Israel content. In a press release issued last month, TikTok said that more young people supported Palestine. "Attitudes among young people skewed toward Palestine long before TikTok existed," the platform said. "Support for Israel (as compared to sympathy for Palestine) has been lower among younger Americans for some time. This is evidenced by looking at Gallup polling data of millennials dating as far back as 2010, long before TikTok even existed." A TikTok spokesperson said: "These allegations about TikTok's policies are wrong or based on misunderstandings, while the Guardian has not given us enough information about their other claims to investigate. We have over 6,000 moderators in Europe who apply our community guidelines equally to all content on TikTok and we do not allow under-13s on our platform."