

# Diary of a TikTok moderator: ‘We are the people who sweep up the mess’

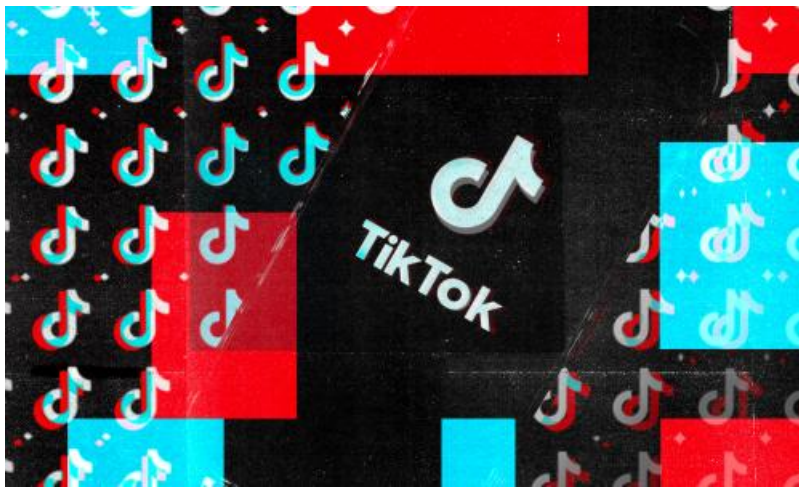
Publication Date: 2023-12-21

Author: Hibaq Farah

Section: Technology

Tags: TikTok, Social media, Digital media, features

Article URL: <https://www.theguardian.com/technology/2023/dec/21/diary-of-a-tiktok-moderator-we-are-the-people-who-sweep-up-the-mess>



TikTok says it has more than 40,000 professionals dedicated to keeping the platform safe. Moderators work alongside automated moderation systems, reviewing content in more than 70 languages. Earlier this year, TikTok invited journalists to its new “transparency and accountability” centre, a move aimed at showing the company wanted to be more open. It says moderators receive training that is thorough and under constant review. Yet little is really known about the working lives of these teams. One moderator, who asked to remain anonymous, explained to the Guardian how difficult the job could be. They said they were judged on how quickly they moderated and how many mistakes they made, with bonuses and pay rises dependent on hitting certain targets. They are also monitored. The moderator claimed that if they are inactive for five minutes, their computer shuts down, and after 15 minutes of doing nothing they have to answer to a team leader. “Our speed and accuracy is constantly analysed and compared to colleagues,” they said. “It is pretty soul-destroying. We are the people in the nightclub who sweep up the mess after a night out.” Here is a first-hand account from a moderator at TikTok: Training: ‘Everyone found it overwhelming’ When we joined we were given one month of intensive training that was so dense it was impossible to absorb. It was six to seven hours a day going over the policies. These are the rules that determine whether a video should be tagged or not, and watching example videos. Everyone found it overwhelming. At the end of the month, there was a test that the trainer walked us through, ensuring we all passed. This has happened in other mandatory training sessions after the probation period as well. Next, was two months of probation where we moderated on practice queues that consisted of hundreds of thousands of videos that had already been moderated. The policies we applied to these practice videos were compared with what had previously been applied to them by a more experienced moderator in order to find areas we needed to improve in. Everyone passed their probation. One trend that is particularly hated by moderators are the “recaps”. These consist of a 15- to 60-second barrage of pictures, sometimes hundreds, shown as a super fast slideshow often with three to four pictures a second. We have to view every one of these photos for infractions. If a video is 60 seconds long then the system will allocate us around 48 seconds to do this. We also have to check the video description, account bio and hashtags. Around the end of the school year or New Year’s Eve, when these sort of videos are popular, it becomes incredibly draining and also affects our stats. Going live: ‘Some of the training was already out of date’ After we passed probation, we were moved on to the real queues. We quickly realised that some of the training we had received in the past months was already outdated due to policies being updated. There are “live” queues where you moderate users streaming live. This is the most simple type of moderation with fewer policies ... but it’s also where we often encounter the worst stuff

and often there is little we can do except end the livestream or place restrictions on the user's ability to upload and go live. Then there are "uploaded" video queues where the length can vary from a few seconds up to an hour. In one queue you are presented with up to six videos from a user's account and have to decide if the owner of the account is over or under 13 years old. In other queues you are presented with a single video and you have to apply relevant policies to any infractions you notice. In another queue you moderate comments. If we have any doubts over what policies we should apply – a common problem due to near constant tweaks, additions and removals made to our policy guidelines – then we have a team of advisers. These are moderators who were promoted, have received extra training on policies and are made aware of forthcoming policy changes. They do a great job but we have seconds to apply these policies [and] it can take minutes, hours or days to get a response, particularly if it is a currently unfolding event such as a war or disaster. Everything we do is tracked by our laptop, which locks after five minutes of no input. We moderate videos up to one hour long, so we have to wiggle the mouse every few minutes to prevent this happening. If the moderation software we use receives no input for 15 minutes, your status is automatically changed to "idle". This can happen if your internet goes down or if you forget to change from moderation status to a meeting/lunch status. All idles are logged, investigated and count against your performance review. You must report the circumstances of your idle to your team leader as well as explaining it in a dialogue box in the software. We were hired to moderate in the English language and had to prove our proficiency as part of the recruitment process, but a huge amount of what we moderate isn't in English. When this happens we are told to moderate what we see. 'You have no control over what you receive' In the video queue you have no control over what you receive. We are given 10 videos at once to moderate before submitting them all. A typical selection of the videos we receive would look like this: Phishing and scam videos in a selection of foreign languages that promise guaranteed high-paying jobs at reputable companies and have instructions to send a CV to a Telegram account. Sex workers trying to direct you to their OnlyFans and so on, while not being able to mention OnlyFans. They use a variety of slang terms and emojis to indicate they have an account on OnlyFans as well as instructions to "check their Instagram for more", meaning that, while direct links to OnlyFans aren't allowed on TikTok, by using the in-app feature that lets you open the user's Instagram profile, the link is never more than a few clicks away. A 10- to 60-minute "get ready with me" uploaded by an underage user where they dress and get ready for school. A recap video featuring hundreds of photos and clips of an entire school year uploaded by someone who just finished their end-of-year exams. Footage of well-known YouTubers' and streamers' most controversial moments, or popular TV shows such as South Park or Family Guy in the top half of the video and Subway Surfers/Grand Theft Auto in the bottom half. A four-minute explicit video of hardcore pornography. Videos featuring what could be Islamist extremist militants but with little to no context because none of the text and spoken language is in a language you were hired to moderate or that you understand. A first-hand recording of young men/teenagers using power tools to steal a selection of motorbikes/scooters/cars, followed by clips of them either driving the vehicles dangerously, destroying the vehicles or listing them for sale. A recording of a livestream that happened on TikTok and has been reposted, probably because it contains controversial comments or behaviour. A list of a person's name, address, place of work and other personal information followed by harassing statements or requests for violence to be committed against the person. You moderate these videos, submit them and then are instantly presented with 10 more. You do this all day. After lunch you move to the comments queue as a backlog has developed. You spend the rest of the afternoon sorting through threats, harassment, racism and innuendo. TikTok declined to comment on the record. However, it insisted "moderator systems" do not shut down after five minutes, and it said it did not recognise the term "recaps". In response to other stories about how the app is policed, it 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."