

Nicole Kleeman
Executive Producer
Firecrest Films
Fairfield
1048 Govan Road
Glasgow
G51 4XS

10 July 2018

Dear Nicole,

Thank you for your letter and for bringing these very important issues to our attention. It is clear that some of the instructions given by trainers at Cpl, as well as comments they made, don't reflect Facebook's values and fell short of the high standards we expect.

Given these issues, we've taken immediate action to prevent this happening again, including:

- A review of training practices across our contractor teams, including Cpl;
- Refresher training by Facebook employees for all trainers working at Cpl (this has started this week);
- A review of staffing at Cpl to ensure that anyone who behaves in ways that are inconsistent with Facebook's values no longer works to review content on our platform; and
- Updated training materials for all reviewers — clarifying our policies in all the areas you've raised.

In addition, in relation to the content where mistakes were clearly made, we've gone back and taken the correct action.

Facebook's Approach to Enforcement of our Community Standards

We want Facebook to be a place where people can share freely and debate difficult, even controversial, issues. But people will only share if they feel safe. And that's why we have clear standards about what's acceptable on Facebook — standards which have been publicly available for many years. In April, for the first time, we also published the internal guidelines used by our review teams — you can read them here: <https://www.facebook.com/communitystandards/>.

Around 1.4 billion people use Facebook every day from all around the world. They post in dozens of different languages: everything from photos and status updates to live videos. Deciding what stays up and what comes down involves hard judgment calls on complex issues — from bullying and hate speech to terrorism and war crimes. It's why we developed our Community Standards with input from outside experts — including academics, NGOs and lawyers from around the world.

Cpl is one of several companies Facebook uses to help us manage and review content on our platform. We have worked with them since 2010, and they currently have 650 people working on these issues, including 14 trainers. Cpl employees take action on reports and escalate decisions where necessary to full-time Facebook staff with deep subject matter and country expertise. For specific, highly sensitive types of content such as child abuse, the final decisions are made by Facebook employees on our Community Operations teams.

As you point out in your letter, this work can involve looking at very disturbing content. It's why all Cpl employees working to review Facebook content are offered psychological help to ensure their wellbeing. The program at Cpl — which was put together with input from psychologists — includes:

- Wellbeing training and pre- and post- support for training modules;
- Peer supervision, and support for those reviewing certain content;
- Enhanced training on anxiety awareness, trauma, stress management, and personal resilience;
- Access to private healthcare from the first day of employment; and
- A 24/7 health support model, with practitioners on site.

In addition, Facebook's content policy team — which devises our Community Standards and internal guidelines — speaks regularly to the teams working at Cpl so that they can answer any questions people may have. We have a partner manager devoted to Cpl and every other company we work with — and our teams conduct regular inspections to review their performance. When a policy is updated, we retrain the reviewers so that they understand the specific changes.

1. TRANSPARENCY & ACCOUNTABILITY

While we strongly agree with some of your criticisms, we believe that Facebook is transparent about the details of our Community Standards, and how we enforce them. As I mentioned earlier, we've published these standards and the internal guidelines used by our review teams. We consulted more than one hundred organizations and experts about the substance and wording of the standards to ensure that they are comprehensive and readable. In May we also held a full day briefing session on our policies and enforcement with academics and legal experts in Oxford (as well as Paris and Berlin) to get their feedback and input. And, we have allowed journalists to sit in on our Content Standards Forum, the bi-weekly meeting where we discuss and adopt changes to our Community Standards. In addition, we have made several offers to the UK Home Affairs Select Committee over the last year to visit our site in Dublin. As with any organization that handles sensitive information, we have rules on how that should be managed and limits on how it may be discussed.

2. RESOURCING AND REVIEW BACKLOGS

The backlog your reporter experienced while filming occurred on select days in March and April. It was cleared entirely by 6 April because Cpl hired additional reviewers and sent reports to other locations with the additional capacity to handle them.

It is untrue to say that this backlog included suicidal content. Reports about suicide are considered "High Priority," and almost 100% of High Priority reports during those two months were reviewed

within the set timeframe (further details about how we deal with suicidal content are set out under point 6 below).

We understand that the timely review of reports is essential to keeping people safe on Facebook. This is why we're doubling the number of people working on our safety and security teams this year to 20,000. This includes over 7,500 content reviewers. We are also investing heavily in new technology to help deal with problematic content on Facebook more effectively.

For example, we now automatically route reports to reviewers with the right expertise and cut out duplicate reports, so that if 100 people report a piece of content, we don't have 100 people reviewing it. This technology can also help us detect and remove content such as terrorist propaganda and images of sexual abuse of children before it has even been reported. In the first quarter of 2018, for example, 99.5% of terrorist content that was removed from the site was identified by our technology without anyone needing to flag it. This also means that fewer reviewers have to see distressing content as it's removed automatically.

3. CONTENT CONTAINING VIOLENCE AND DISTURBING IMAGES

In terms of the video of minors fighting and assaulting a girl, this type of content can be very upsetting. It's why we always remove bullying videos unless they are shared to condemn the behavior, and even those shared to condemn are restricted to adults only and include a warning screen letting people know the content may be disturbing. We will always remove such content, regardless of the context in which it is shared, when the minor or their guardian has requested it. Since learning from you that the mother wanted this video removed, we have taken steps to ensure it has been deleted and to prevent further uploads. To be clear, the person sharing the video did so to condemn this behavior (which is not uncommon on Facebook).

4. CONTENT DEPICTING CHILD ABUSE

With regard to the video showing the terrible abuse of a Malaysian child, we removed it (in line with our policies) once we knew the child was safe, and have used image-matching software to prevent it being uploaded ever since.

The statement you recorded that “we never delete and we never ignore” child abuse content accurately reflects the way this type of content should be handled by the initial reviewers. They must mark it as disturbing and then escalate it to the Facebook Safety team, who can then decide whether the content should be deleted and if it should be reported to law enforcement.

We always remove non-sexual child abuse imagery except where the sharing of the image may lead to the rescue of a child in danger. This is on the advice of outside experts who have advised that leaving such content up can save children from abuse. In the relatively few number of cases where this disturbing content remains on the platform, we apply a warning screen and limit its distribution. As soon as we have confirmed that the child has been identified and rescued we remove it. If we're able to identify the country from which this abusive content came, we will report it to law enforcement.

We take a zero-tolerance approach to child sexual abuse imagery. Whether detected by our technology or reported to us, we remove it as soon as we find it. Any apparent child sexual

exploitation content is reported to the National Center for Missing and Exploited Children (“NCMEC”)¹. We also use technology to prevent this content being uploaded onto Facebook again.

5. MINORS

We do not allow people under 13 to have a Facebook account.

If a Facebook user is reported to us as being under 13, a reviewer will look at the content on their profile (text and photos) to try to ascertain their age. If they believe the person is under 13, the account will be put on a hold. This means they cannot use Facebook until they provide proof of their age. We are investigating why any reviewers or trainers at Cpl would have suggested otherwise.

In the example described in your letter, we would apply our policies regarding content depicting self-harm and the matter would be escalated to specialist Facebook staff (as described further below). It is right that the Cpl reviewer would not take action on the account, because it was not reported as being under 13. Our contractors do not proactively put a hold on accounts whom they suspect may be under 13.

6. SUICIDE AND SELF INJURY

We take the issue of self-harm extremely seriously and have developed our policies with significant input from experts, such as Save.org, National Suicide Prevention Lifeline, and Forefront.

Facebook does not allow content that promotes or encourages self-harm or self-injury, or in the case of Live video, a credible or successful attempt at suicide. However, we do allow content posted by people who contemplate or admit to engaging in self-harm or self-injury. This is because sharing these thoughts can be therapeutic as well as a “cry for help.” As friends and family are already connected on Facebook, our platform is uniquely positioned to help people in distress make contact with a loved one.

If anyone posts self-harm or self-injury-related content – regardless of whether the content is removed for violating our policies — we send them information about where to get help. And people watching a live video have the option to reach out to the person directly or report the video to us.

Finally, since November of 2017, we've been using artificial intelligence to identify people expressing thoughts of suicide. We can then reach out to offer help or work with first responders, which we've now done in over a thousand cases².

¹ <http://www.missingkids.com/home>

² <https://newsroom.fb.com/news/2017/11/getting-our-community-help-in-real-time/>

7. FACEBOOK'S BUSINESS MODEL

We strongly disagree with your suggestion that turning a blind eye to problematic content is in our interests. Creating a safe environment where people from all over the world can share and connect is core to our business model. If our services aren't safe, people won't share and over time would stop using them. Nor do advertisers want their brands associated with disturbing or problematic content — and advertising is Facebook's main source of revenue.

It's why we have very clear Community Standards and we've invested billions of dollars in working to enforce them effectively — using both people and improved technology like artificial intelligence, computer vision and machine learning.

8. LOCAL JURISDICTIONS

What is described in your letter is not an accurate description of our policies and we are investigating what happened. In terms of the facts, when something on Facebook is reported to us as violating local law, but which doesn't break our Community Standards, we review it carefully and may restrict it in the country where it is alleged to be illegal. You can see these content removal requests in our biannual Transparency Report³. For example, we restrict access to content that denies the Holocaust in countries like Austria, Belgium, France, Germany, Israel, Italy, Netherlands, Slovakia, and Spain where such content is illegal.

9, 10, 11 HATE SPEECH

The statement: *"If someone wants to post hateful jokes and none of their friends report them then it's fine. It's in the platform and it's fine"* is untrue and at odds with our values. It's why we're reviewing what happened here to see what went wrong and prevent it happening again.

Hate speech is never acceptable on Facebook, and we're increasingly using technology to proactively detect it — without anyone needing to report it. In fact, of the 2.5 million pieces of hate speech we removed from Facebook in Q1 of 2018, 38% was flagged by our technology.

The post comparing Muslims to sponges violates our hate speech policies because it de-humanises Muslims. The meme that includes the *"When your daughter's first crush is a little negro boy"* text also violates our hate speech policies as it is mocking a hate crime — in this case, violence based on racial bias. We have deleted this content and use image-matching software to prevent it being uploaded again.

You told us that you have other examples including one referring to *'stinking Muslim immigrants'*. We would ask that you send us the entire post as soon as possible so we can assess whether it breaks our standards. If the person posting it were condemning this abuse or raising awareness of what they had experienced, we would allow it. In other cases, however, we would remove it.

³ <https://transparency.facebook.com/>

12, 13. SHIELDED PAGES & BRITAIN FIRST

We removed Britain First from Facebook in March because their Pages repeatedly violated our Community Standards.

Britain First were a registered political party in the UK and until recently had stood in elections. It's why we took great care dealing with this issue; we are reluctant to censor political content or interfere in national debates.

We remove content from Facebook, no matter who posts it, when it breaks our standards. However, we do have a process to allow for a second look at certain Pages, Profiles, or pieces of content to make sure we've correctly applied our policies. While this process was previously referred to as "shield," or shielded review, we changed the name to "Cross Check" in May to more accurately reflect the process.

We use Cross Check when reviewing high profile, regularly visited Pages or pieces of content on Facebook to help prevent them from being erroneously removed. We may Cross Check posts of celebrities, governments, news organizations or Pages where we have made mistakes in the past based on the type of content posted. For example, we Cross Checked decisions on an American civil rights activist's account when content was reported to avoid deleting posts where he was highlighting hate speech he'd been subjected to. However, if someone posts something that breaks our Standard and their page was Cross-Checked, we would still remove that content from Facebook after double-checking this was the correct decision.

We would like to thank you again for bringing these incredibly important issues to our attention. Some of the behavior you have described is neither consistent with Facebook's policies or our values. It's why we're investigating what happened and have already taken action to prevent it happening again.

Yours sincerely,

Monika Bickert
Vice-President, Global Policy Management