Progress on Hate Speech

We’ve made significant progress combating hate speech. Prevalence of hate speech on Facebook has decreased for a third quarter in a row. We evaluate the effectiveness of our enforcement by trying to keep the prevalence of hate speech on our platform as low as possible, while minimizing mistakes in the content that we remove. This improvement continues to stem from changes we made to reduce problematic content in News Feed. On a quarterly basis, from Q1 2018 to Q1 2021, we’ve increased our content actioned by 10 times. We’ve also increased our proactive rate from 24% to 97% since we started reporting it.

Why
Hate speech has no place on Facebook and Instagram, and we’re committed to improving our policies, technology and enforcement against this type of abuse.

What’s new

In this report, we are expanding our metrics for IG and will now include prevalence for 1) adult nudity and sexual activity and 2) violent and graphic content. Prevalence is our most important metric because it shows us how big the problem is, and allows us to track our progress at removing harmful content while minimizing mistakes. In Q1 2021, prevalence of adult nudity and sexual activity was 0.03%-0.04% and violent and graphic content was 0.01%-0.02% on Instagram.

Why
We’re committed to matching Instagram metrics with what we share for Facebook.