

Community Standards Enforcement Report

HATE SPEECH

Hate speech has no place on our platforms. We work to identify, find, and remove hate speech, all while being as transparent as possible to our community. Hate speech has never been allowed on our platforms and this timeline highlights recent milestones and steps we have taken to combat this issue, including policy updates and technology improvements.

Integrity-Related Announcement Highlights

2020

MAY

Advances in AI

We announced more details on the advances we made in using AI to combat hate speech—including developing new technology to help us detect more subtle and complex hate speech, and rolling out “language-agnostic” technology, to help us detect hate speech in more languages. [Read More](#)

Challenge

We publicly shared a data set designed specifically to help the wider industry develop new systems to help identify hate speech and launched a “Hateful Memes Challenge,” with a \$100,000 prize pool, to catalyze research in this area. [Read More](#)

JUNE

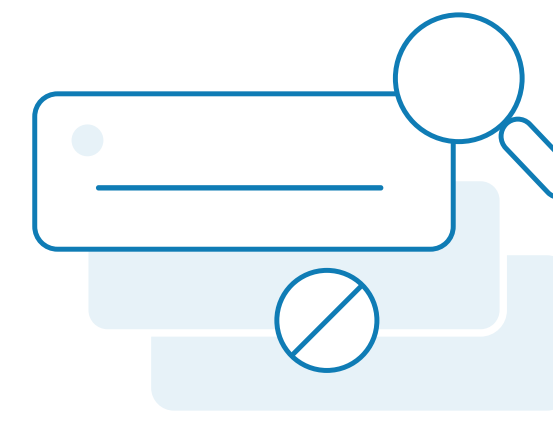
Inflammatory Ads Policy

We explained our work in the US elections, including how we created a higher standard for hateful content in paid ads. [Read More](#)

JULY

Civil Rights Audit

Final report in two-year independent audit that evaluated civil rights impact of Facebook. [Read More](#)



94.7%

Hate Speech Proactive Rate on Facebook in July – September 2020

AUGUST

Harmful Stereotypes List

We updated our policies to account for certain kinds of implicit hate speech, such as content depicting Blackface, or stereotypes about Jewish people controlling the world and major institutions. [Read More](#)

Technology Update

We explained how we use a combination of technology and content review teams to find and remove hate speech and other harmful content so we can get to the most harmful content quickly. [Read More](#)

OCTOBER

Holocaust Denial

We updated our hate speech policy to prohibit any content that denies or distorts the Holocaust. [Read More](#)

NOVEMBER

Global Hate Speech Prevalence

We released, for the first time, hate speech prevalence on our platform. [Read More](#)

2019

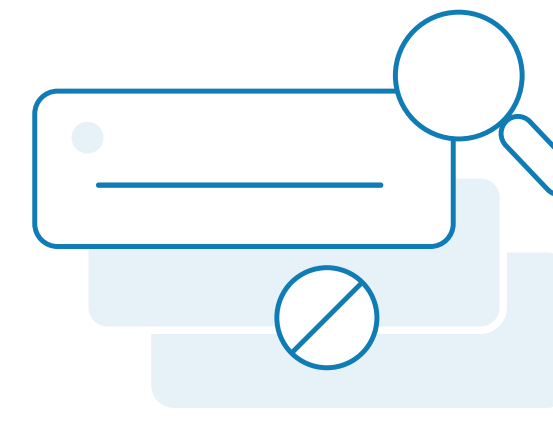
MARCH

Ban on White Nationalism and White Separatism

We banned praise, support and representation of white nationalism and white separatism. We started connecting people in the U.S. who search for terms associated with white supremacy on Facebook to resources by Life After Hate focused on helping people leave behind hate groups. We subsequently have expanded this program to Australia and Indonesia. [Read More](#)

Protections

We expanded protections against discrimination in housing, employment and credit advertising. [Read More](#)



80.6%

Hate Speech Proactive Rate on Facebook in July – September 2019

JUNE

Social Media and Countries in Conflict

We announced that we had created a dedicated, multi-disciplinary team to better understand and address the way social media is used in countries at risk of conflict—including how it can be used to spread misinformation, hate speech, and polarization. We shared details of the team’s work to address these challenges. [Read More](#)

Civil Rights Audit

Our auditors issued a second update on the Civil Rights Audit in the US. We announced new steps to combat white nationalism and white separatism, and new policies to ensure Facebook isn’t used to organize events that intimidate or harass people based on their race, religion, or other parts of their identity. [Read More](#)

OCTOBER

Investment

We announced that we have more than 35,000 people working on safety and security at Facebook. [Read More](#)

2018

MAY

AI Expansions

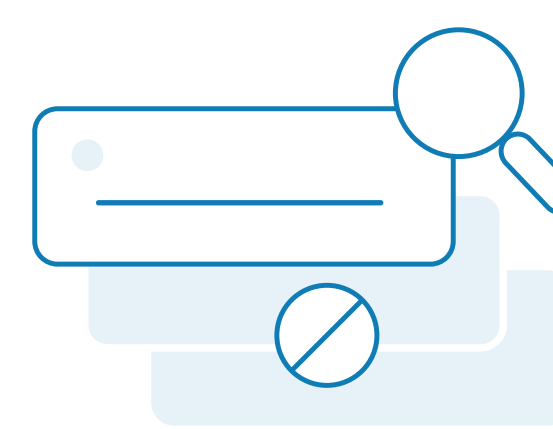
Facebook announced that it started using AI to proactively find and remove hate speech in English and Portuguese. [Read More](#)

Enforcement Numbers

We released our First Community Standards Enforcement Report. Between October and December we removed 1.6 million pieces of hate speech—23.6% of which was detected proactively by our technology before people reported it to us. [Read More](#)

Civil Rights Audit

We accepted the call to undertake a Civil Rights Audit in the U.S.



51.9%

Hate Speech Proactive Rate on Facebook in July – September 2018

DECEMBER

Civil Rights Audit

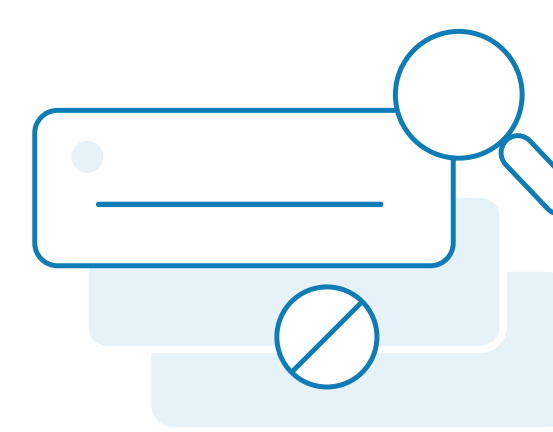
We shared our first update on the Civil Rights Audit in the U.S. [Read More](#)

2017

JUNE

Hate Speech Policy

We published more details on how we define hate speech, our approach to removing, and some of the complexities that arise when it comes to doing this in multiple languages, across many cultures. [Read More](#)



23.6%

Hate Speech Proactive Rate on Facebook in October – December 2017

AUGUST

Hate Speech Policy

We refined our hate speech policy by introducing three different tiers or types of attack, from most severe to least, with the goal of enabling more effective and nuanced enforcement.

How We Measure Progress

We use several metrics to evaluate how we’re protecting our community. These metrics include the percentage of times people see violating content on our platform (prevalence) and how much content we found before someone reported it to us (proactive rate). For the first time, we’re including the prevalence of hate speech on Facebook. In Q3 2020, hate speech prevalence was 0.10% – 0.11% or 10 to 11 views of hate speech for every 10,000 views of content. Due to our investments in AI, we have been able to remove more hate speech, and find it proactively before users report it to us. On both Facebook and Instagram, we now proactively identify about 95% of the hate speech we remove. You can read more about our work on hate speech [here](#).

FACEBOOK APP
INSTAGRAM

