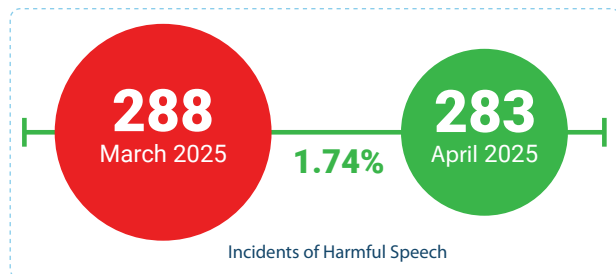


Harmful Speech in Sri Lanka

MONTHLY BULLETIN
A P R I L 2 0 2 5

Key Trends

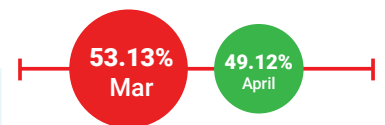


Compared to March, there was a 7.5% decrease in the number of anti-LGBTQI+ content and a 16.85% decrease in content targeting women this month. However, there was an 83.3% and 84.62% increase in content targeting children and diseased/disabled groups, respectively.

Key events in April that led to harmful speech



HATE SPEECH
INCIDENTS IN APRIL **139**



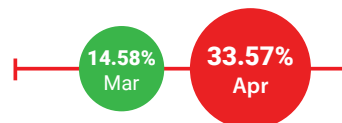
Anti-LGBTQI+ and Misogynistic Attacks on Prime Minister Harini Amarasuriya

Prime Minister Harini Amarasuriya continues to experience ongoing online hostility, including derogatory language related to gender and sexual identity. These comments are often tied to her appearance, political decisions, and participation in international forums such as the BIMSTEC Summit.

Caste-Based Hate Speech Targeting Tamil and Malayaga Tamil Communities

Disparaging caste-based terms such as "Vadakathayan" have been directed at political figures like Minister Chandrasekaran and broader Tamil and Malayaga Tamil communities. This rhetoric reinforces ethnic and caste hierarchies and contributes to social division.

DISINFORMATION
INCIDENTS IN APRIL **95**

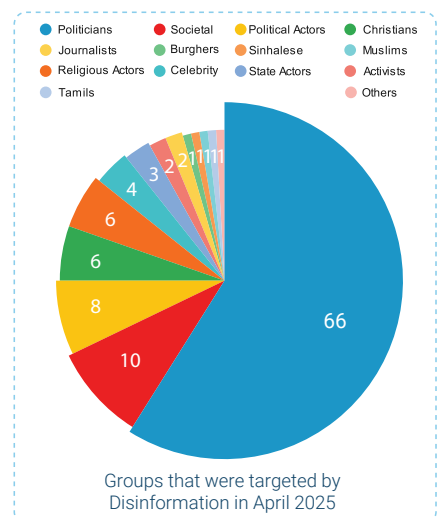


Disinformation Around the Special Exposition of the Sacred Tooth Relic (Dalada Exhibition)

A surge of misleading content related to the Dalada exhibition, including AI-generated images, fabricated quotes, and altered videos was seen in April. These materials appear designed to provoke unrest by spreading false narratives about communal violence and fabricated political statements.

Misinformation Around Local Government Election

Numerous false narratives involving the Local Government candidates were also observed. This included AI-generated media, manipulated images and videos, and misattributed quotes, all of which have the potential to polarise communities and influence public opinion along political, ethnic, and religious lines.



01. Harmful speech includes hate speech, dis/misinformation, harassment and incitement to or glorification of violence.
02. The UN Strategy and Plan of Action on Hate Speech defines hate speech as, "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor."
03. While there is no universally accepted definition of disinformation, it is commonly understood as information that is not only inaccurate, but is also intended to deceive and is spread in order to inflict harm.

RESPONSIVENESS OF SOCIAL MEDIA PLATFORMS

April 2025

Key Observations

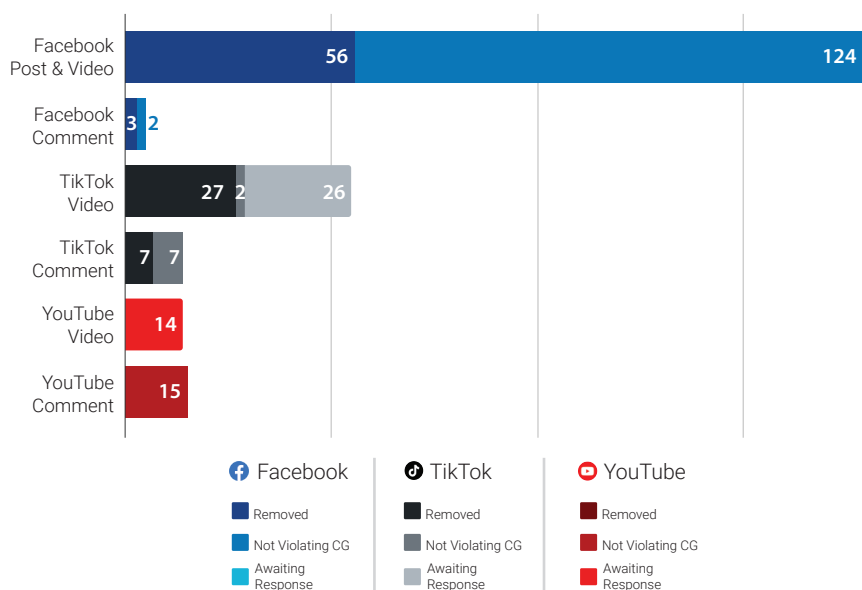
In April, 283 incidents of dangerous speech were reported to social media platforms. Of these, 185 were reported to Facebook, 69 to TikTok, and 29 to YouTube. Unlike other platforms, where both posts and comments can be reported, only videos can be reported on YouTube. This month, TikTok had the highest removal rate, at 49.28%, followed by Facebook at 31.35%. Notably, YouTube did not take down any of the reported content.

68.65% of content reported to Facebook was found not to violate its community guidelines.

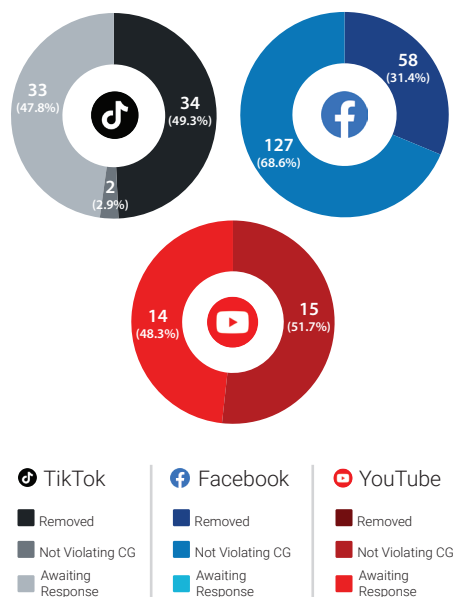
47.83% of content reported to TikTok awaits a response.

All platforms responded to 83.39% of the content within four days of escalation.

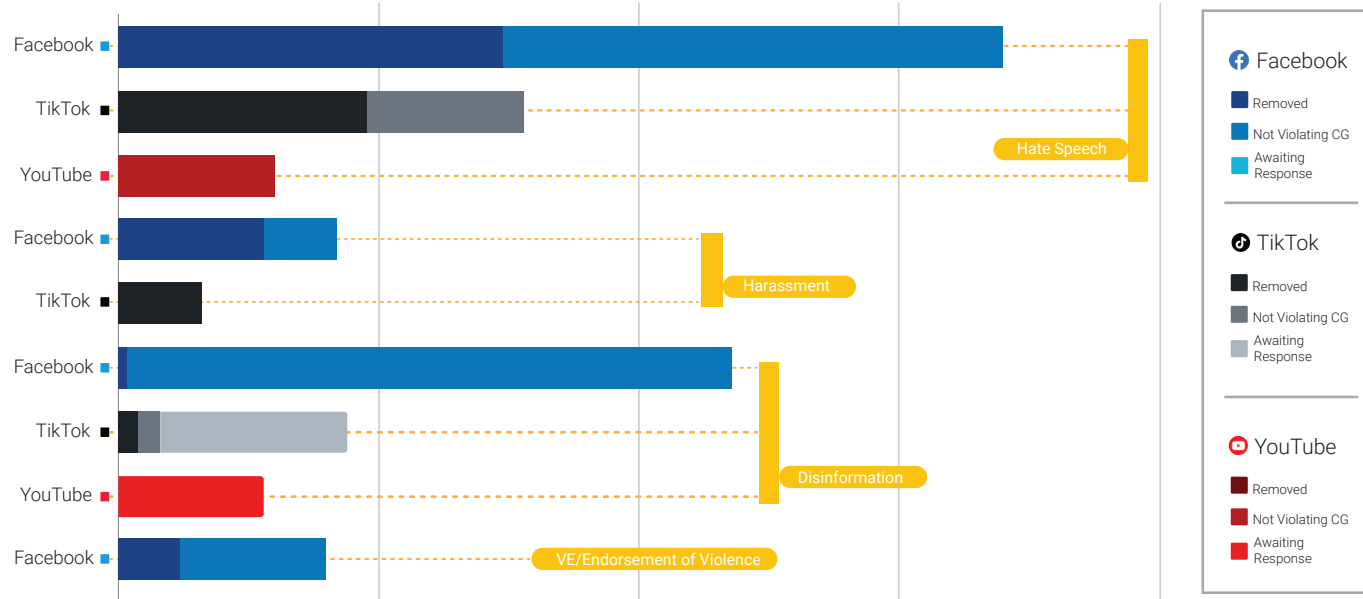
RESPONSE BY THE PLATFORM FOR EACH TYPE OF CONTENT



OVERALL RESPONSE BY PLATFORMS



RESPONSE BY CATEGORY



04. The ability to monitor Facebook was negatively impacted by the retirement of Meta's CrowdTangle tool on 14 August 2024.

05. The UN in Sri Lanka relies on trusted external fact-checkers to independently monitor dis- and misinformation. Actual levels of dis- and misinformation may be higher than reported due to monitoring capacity constraints.