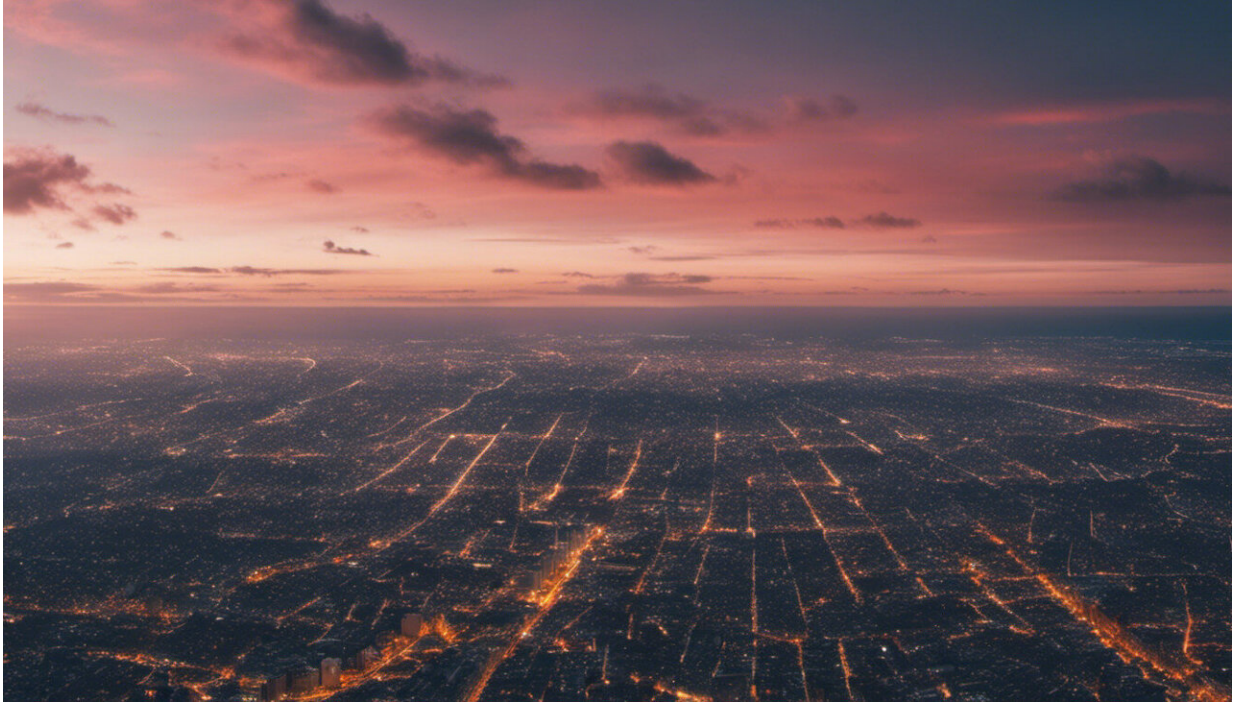# Mapping online hate speech

June 24 2016

Credit: AI-generated image ([disclaimer](#))

You might think from anecdotal evidence that hate speech on social media by individuals and groups appears quite a lot, but one of first academic studies to examine the empirical data concludes that these extreme forms of speech on Facebook are marginal as compared with total content.

Researchers from the University of Oxford and Addis Ababa University

examined thousands of comments made by Ethiopians on Facebook during four months around the time of Ethiopia's general election in 2015. Hate speech is defined as statements to incite others to discriminate or act against individuals or groups on grounds of their ethnicity, nationality, religion or gender. Using a representative sample of total online statements, they found that only a tiny percentage could be classed a such, just 0.7%. The paper says the findings may have wide implications for the many countries trying to address growing concerns about the role played by social media in promoting radicalisation or violence.

There have been increasing demands for research that can detect and monitor these types of online behaviours, says the report. Yet, until now, very little systematic research has been carried out into how people use social media to whip up hostility against others. The international research team used Ethiopian online conversations as a case study because of the country's distinct language, which meant they could target Ethiopians living in their home country and abroad. This made the task far more controlled and contained than trying to track English language speakers, for instance.

The researchers looked at statements made both before and after the election in May 2015. They found that fans or followers rather than people with any real influence online were mainly responsible for the violent or aggressive speech that appeared on Facebook pages in the sample studied. The study suggests that these individuals have little or no power and use Facebook to vent their anger against more powerful sections of society. Around 18% of total comments in the sample were written by fans or followers compared with 11% of comments made by highly influential speakers (the owners of web pages). One fifth (21.8%) of hostile comments were grounded in political differences, only slightly higher than the overall average of 21.4% of all conversations containing hostile comments. Religion and ethnicity provoked fewer hostile

comments (10% and 14% of overall comments in the sample respectively).

Study author Dr Iginio Gagliardone, Associate Research Fellow in the Programme in Comparative Media Law and Policy at Oxford University and Lecturer in Media Studies at the University of the Witwatersrand, says: 'In Ethiopia, as in many other countries – from the United States to Germany to China –there are growing claims that social media is used to spread hate speech and incite violence. Yet this study shows that the actual level of hate speech or dangerous speech is much lower than the anecdotal evidence suggests.We also find that social media in Ethiopia is emerging as a space for tolerance and acceptance. Hate speech and dangerous speech do occur on these platforms, but they are a marginal proportion of the total online conversations in Ethiopia.'

The study concludes that despite the polarisation that has characterised the media in Ethiopia, the social media channels appear to offer greater opportunities for engagement.

Co-author Dr Nicole Stremlau, from the Centre for Socio-Legal Studies at the University of Oxford, adds: 'We found many users went onto Facebook to communicate across divides, rather than to exacerbate existing tensions. This suggests we need to understand what resources are already available online that can help tackle the most aggressive forms of speech before resorting to more severe measures such as censorship.'

Researchers analysed more than 13,000 statements posted on 1,055 Facebook pages between February and June 2015. They mapped Facebook profiles, pages, and groups that had 100 or more followers or 'likes' or members, respectively. All content in the sample studied had to include an Ethiopian language and raise discussion topics about Ethiopia. The research team focused on popular spaces on Facebook, analysing pages daily to map ongoing trends. Their sample also included comments

on random pages, as well as pages capturing particular events, such as a protest or publicised speeches. Posts, status updates and comments were tracked over time. The researchers categorised statements that appeared on the pages sampled as 'offensive', 'hate speech' or 'dangerous speech'. They categorised comments that attacked particular individuals or groups, and whether they encouraged specific action. They also took into account comments that did not agree or disagree but appeared to engage with a particular topic and communicate opinion in a way that was not hostile.

More information: Gagliardone, Iginio and Pohjonen, Matti and Zerai, Abdissa and Beyene, Zenebe and Aynekulu, Gerawork and Gebrewolde, Tewodros and Seifu, Michael and Stremlau, Nicole and Bright, Jonathan and Bekalu, Mesfin and Moges, Mulatu Alemayehu, Mechachal: Online Debates and Elections in Ethiopia. Report One: A Preliminary Assessment of Online Debates in Ethiopia (October 2, 2015). Available at SSRN: ssrn.com/abstract=2782070

Provided by University of Oxford