Hate speech regulation on social media: An intractable contemporary challenge

Research Objectives
The Bonavero Institute fosters world-class research in human rights law and promotes public engagement in and understanding of human rights issues.

References


Hate speech regulation on social media:
An intractable contemporary challenge

The Internet has allowed people across the world to connect instantaneously and has revolutionised the way we communicate and share information with one another. More than 4 billion people were Internet users in 2018, more than half of the global population. In many ways, the Internet has had a positive influence on society. For example, it helps us to communicate easily and to share knowledge on all kinds of important topics efficiently: from the treatment of disease to disaster relief. But the Internet has also broadened the potential for harm. Being able to communicate with a mass audience has meant that the way we engage with politics, public affairs and each other has also changed.

Hateful messages and incitements to violence are distributed and amplified on social media in ways that were not previously possible.

Through social media platforms (such as Facebook, Twitter, YouTube, Instagram and Snapchat), 3.19 billion users converse and interact with each other by generating and sharing content. The business model of most social media companies is built on drawing attention, and given that offensive speech often attracts attention, it can become more audible on social media than it might on traditional mass media. Given the growing problem of offensive and harmful speech online, many countries are asking themselves the challenging question whether they should regulate speech online and if so, how they should legislate to curb these excesses.

HATE SPEECH VS FREEDOM OF SPEECH

The regulation of harmful speech in online spaces requires drawing a line between legitimate freedom of speech and hate speech.

In countries committed to freedom of speech, it is necessary to develop a shared understanding of why freedom of speech is important. O’Regan and Thiel suggest that there are three main reasons why we value freedom of speech: because we think being able to speak our minds is part of what makes us free and autonomous human beings, for democratic reasons, because we need to be able to talk about politics and policy freely to enable us to decide as equals how to vote and to hold those in power to account and for truth-related reasons, to enable us to refute false claims.

Just as we need to understand why we value freedom of speech, we also need to understand why we should prohibit hate speech. There are two main reasons for outlawing hate speech: the first and most widely accepted reason is that hate speech is likely to result in actual harm to those who are being targeted (“the incitement to harm” principle): so speech that incites violence against, for instance, people of a particular race, sexual orientation or gender identity is outlawed in most countries, including the USA. Many countries also agree that hate speech that is degrading of groups of people should also be prohibited (“the degrading of groups” principle), because it undermines their status as free and equal members of society. Again, many countries, but notably not the USA, prohibit such forms of hate speech online and if so, how they should legislate to curb these excesses.
Digital media allows online speech and content to be shared anonymously and often without a second thought for the consequences.

Hateful messages and incitements to violence are distributed and amplified on social media in ways that were not previously possible.

The age of digital media has allowed online speech and content to be shared anonymously and often without a second thought for the consequences. While the act of publishing online is instantaneous, mechanisms designed to regulate speech are often cumbersome and slow. Moreover, in traditional forms of media, there is editorial oversight from a person other than the author prior to publishing. Historically, this has often provided an effective restraint on hate speech, a mechanism that plainly does not work on self-published social media platforms.

The speed and sheer amount of content, as well as the lack of editorial oversight make social media platforms a particular challenge for regulators. Increasingly, policymakers are suggesting that social media platforms should bear the brunt of the regulatory burden: for instance, through obligations to provide effective complaint mechanisms and remove unlawful speech. The risk with this approach is that lawful speech may be removed in error, or that the general environment will inhibit individuals from expressing themselves online.

The United States differs from other jurisdictions being assessed in some important respects. The First Amendment however does not prevent private actors, like social media platforms, from imposing their own restrictions on speech. Social media platforms are further protected from private litigation because they are not considered publishers of the content posted to their sites in terms of section 230 of the Communications Decency Act 1996.

The United Kingdom imposes a range of criminal prohibitions on hate speech, both online and in print. The Crime and Disorder Act, Public Order Act, Malicious Communications Act 1998 and Communications Act 2003 prohibit speech that is derogatory on grounds of race, ethnic origin and religious and sexual orientation. A recent White Paper contains sweeping proposals to regulate online media by imposing a duty of care upon social media platforms, and establishing a regulator to ensure that the duty of care is observed.

The broad range of companies covered and open-ended list of online harms identified for regulation in the White Paper are a particular concern: it risks overburdening the regulator and leading to highly selective enforcement. The European Union has adopted the e-Commerce Directive which prevents monitoring of content on websites before it is published, a provision which shapes and impacts the development of regulatory initiatives in Europe. The EU is exploring further options in regulating social media.

The German Network Enforcement Law introduced just over two years ago imposes obligations on social media platforms to establish complaints management mechanisms which must work quickly, transparently and effectively. Where unlawful content (as defined by the German Criminal Code) is identified it must be removed or blocked within a specified deadline. The specific deadline depends on whether content is manifestly illegal, or simply illegal, and whether the social media platform cooperates with a recognised body of industry self-regulation. Fines of up to 50 million Euros can be issued for systemic failings in the complaints management system, including not consistently meeting the required deletion deadlines, and ignoring reporting and transparency requirements.

FUTURE DIRECTIONS

Regulating hate speech online is a major policy challenge. Policymakers must ensure that any regulation of social media platforms does not unduly impair freedom of speech. Given the complexity of the problem, close monitoring of new legislative initiatives around the world is necessary to assess whether a good balance has been struck between the protection of freedom of speech and the prohibition of hate speech. In order for this monitoring to take place, social media companies need to be transparent about the content that they are removing and make their data available to researchers and the wider public for scrutiny.

Policymakers must ensure that any regulation of social media platforms does not unduly impair freedom of speech.