



**Policy Briefing:**

# **Dealing With Hate Speech on Social Media**

Charlotte Spencer-Smith

TalkPolitics has an open access policy. Copyright remains with the copyright holder, but users may download, save and distribute this work in any format provided: (1) that TalkPolitics is cited; (2) that the web address [talkpolitics.org.uk](http://talkpolitics.org.uk) is published together with a prominent copy of this notice; (3) the text is used in full without amendment [extracts may be used for criticism or review]; (4) the work is not re-sold; (5) the link for any online use is sent to [executive@talkpolitics.org.uk](mailto:executive@talkpolitics.org.uk)

© TalkPolitics 2018 Published in the UK by TalkPolitics LTD.  
Some rights reserved  
Printed in England

## **TABLE OF CONTENTS**

<b>INTRODUCTION.....</b>	<b>4</b>
<b>SELF-REGULATION BY SOCIAL MEDIA COMPANIES.....</b>	<b>6</b>
<b>ALTERNATIVE APPROACHES IN EUROPE .....</b>	<b>8</b>
<b>WAYS FORWARD FOR THE UK .....</b>	<b>10</b>
<b>BIBLIOGRAPHY .....</b>	<b>12</b>

# INTRODUCTION

Hate speech on social media is damaging not just to the wellbeing of users, but also to an open, egalitarian society. The problem is well described by research at Amnesty International UK into abusive tweets against female Members of Parliament in the first half of 2017. The research found that abuse disproportionately affected Black, Asian and Minority Ethnic (BAME) subjects. For instance, Shadow Home Secretary Diane Abbott<sup>1</sup> infamously faced a slew of racist and sexist comments on social media. To intimidate individuals and groups who contribute to diversity and pluralism in British public life doesn't add to the national discourse and should thus not be considered through the lens of free speech.

Currently, in the United Kingdom the principle measure used to counter abuse on social media is criminal legislation. This notably includes Section 127 of the Communications Act 2003 which makes it an offence to send a message of "grossly offensive or of an indecent, obscene or menacing character" online. Although the Communications Act is typically used to prosecute online abuse, hate speech is also prohibited in several places in British law, including the Public Order Act 1986 (incitement to hatred based on colour, race, ethnic origin and nationality), the Criminal Justice and Public Order Act 1994 (amending the Public Order Act to include intent to cause alarm or distress), the Racial and Religious Hatred Act 2006 (amending the Public Order Act to include religious hatred) and the Criminal Justice and Immigration Act 2008 (sexual orientation). Freedom of expression is protected by the Human Rights Act 2008, which incorporates the European Convention on Human Rights into UK law. The Human Rights Act allows for a number of restrictions to freedom of expression, including, "for the prevention of disorder or crime, for the protection of health

---

<sup>1</sup> Amnesty International UK, n.d. *Black and Asian women MPs abused more online*. [Online] Available at: <https://www.amnesty.org.uk/online-violence-women-mps> [Accessed 25 January 2018].

or morals, for the protection of the reputation or rights of others". Freedom of expression is also included in the European Charter of Fundamental Rights, but Parliament voted in January 2018 not to transpose the Charter into UK law after Brexit. This was used to prosecute two Twitter users who abused the campaigner Caroline Criado-Perez in the high-profile case of 2014.

UK law does not hold social media companies liable for abusive messages sent by users. The justification is that internet companies are service providers offering communications platforms and are not publishers with an editorial responsibility. However, recent revelations of online radicalisation and recruitment by terrorist and extremist groups, as well as the dissemination of fake news has challenged this premise. These revelations result in real-world consequences and are considered direct threats to democracy and security. The UK government is currently working to place the onus on social media companies to prevent terrorists from using their services to recruit and organise.

However, the challenges go well beyond terrorism and security, and reach into other areas of public life. The massive influence of social media, as well as the changing, dynamic nature of its impact on both public and private life, means the UK would benefit from revisiting its approach to social media providers and updating governance.

# SELF-REGULATION BY SOCIAL MEDIA COMPANIES

To a large extent, social media companies police what is and is not acceptable to post on their platforms. Twitter, Facebook and YouTube have global user policies on hate speech<sup>2</sup> and remove comments judged to be in breach of these policies. In cases of repeated or extreme abuse, social media providers typically suspend or close user accounts. However, it is important to note that removal decisions are based on company policy and not on UK national law. Furthermore, user policies tend to lack transparency about how guidelines are put into practice, since more detailed information is not available to the public. In 2017, leaked Facebook training documents provided an insight into the counterintuitive way the company implements its user policy on abuse. In the training slides, Facebook considers hate speech to be an attack against a protected category. However, if a protected category (eg. sex, ethnicity) is combined with a non-protected category (eg. age, social class), this produces a non-protected category. As an example, the slides name “white men” as a protected category, but “black children” as a non-protected category, meaning that an attack against black children would not be removed<sup>3</sup>.

Offensive content that does not violate user policy is also a controversial issue on other platforms, particularly where it commercially benefits the company or the user. In 2017, YouTube - owned by Google - refused to take down extremist content that did not violate its terms of service, but it did demonetise the videos and make them more difficult for users to find. This step

---

<sup>2</sup> YouTube, n.d. Policies, safety, and reporting > Policy Center > Hate Speech. [Online] Available at: <https://support.google.com/youtube/answer/2801939?hl=en> [Accessed 25 January 2018].

Facebook, n.d. Community Standards. [Online] Available at: <https://www.facebook.com/communitystandards#hate-speech> [Accessed 25 January 2018].

Twitter, n.d. Hateful Conduct Policy. [Online] Available at: <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy> [Accessed 25 January 2018].

<sup>3</sup> Grassegger, J. A. & H., 2017. Facebook’s Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children. [Online] Available at: [https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms?utm\\_campaign=sprout&utm\\_medium=social&utm\\_source=facebook&utm\\_content=1498662906](https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms?utm_campaign=sprout&utm_medium=social&utm_source=facebook&utm_content=1498662906) [Accessed 25 January 2018].

was taken in response to complaints from advertisers that their advertisements were being shown in the context of videos with extremist content. The UK Government was one advertiser affected. Advertisements alongside content strongly suggests that the uploaders were financially benefiting from the content. The measures put in place by YouTube mean that although creators cannot earn money from extremist videos, the platform continues to host them.

## ALTERNATIVE APPROACHES IN EUROPE

By comparison, in January 2018 Germany started to enforce the Network Enforcement Law (“NetzDG”)<sup>4</sup>, one of the most advanced pieces of legislation against online hate speech currently in existence. Social media companies with over 2 million users in Germany must delete or block “evidently unlawful content” within 24 hours, or within a week for more complicated cases. The term “evidently unlawful content” refers to hate speech and abuse already deemed as criminal offences by German law. If social media companies fail to comply, they may be subject to fines of up to 50 million Euros. With the new law, Germany is taking a different approach by making companies legally responsible in the fight against hate speech and online abuse.

Germany’s Network Enforcement Law is an early venture into legislation governing the role of social media companies. Its introduction has not been without unintended consequences, however. Keen to avoid prosecution and fines, in particular, Twitter has acted beyond the necessary requirements of the law to delete tweets that are not “evidently unlawful”. In January 2018, the company deleted tweets by Alternative für Deutschland politicians Beatrix von Storch and Alice Weidel that could comfortably be regarded as evidently unlawful content. However, Twitter also deleted a tweet by German satire magazine *Titanic* which made fun of the original tweets. When the magazine repeated its tweet, Twitter suspended the magazine’s account for 48 hours<sup>5</sup>.

This raises questions about how to secure protections for freedom of expression, a particularly relevant topic for social media companies given their

---

<sup>4</sup> Government of Germany, 2017. Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz - NetzDG). [Online]

Available at: <https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html> [Accessed 25 January 2018].

<sup>5</sup> The Economist, 2018. Germany is silencing “hate speech”, but cannot define it. [Online] Available at: <https://www.economist.com/news/europe/21734410-new-social-media-law-causing-disquiet-germany-silencing-hate-speech-cannot-define-it> [Accessed 25 January 2018].

subjection to censorship and attempts by governments and groups to limit freedom of speech. For instance, Facebook is banned from operating in China due to the state's censorship and tight control of media. Furthermore, as was seen when Facebook suspended groups of Egyptian activists<sup>6</sup>, groups of users can coordinate to abuse the reporting functions on platforms by flagging content that does not violate user policy or legally constitute hate speech, with the aim of achieving a political goal or shutting down rival groups.

However, the German Network Enforcement Law is not the only example of governmental response to hate speech on social media. The European Commission has had some success with a softer governance approach. In 2016, the Commission agreed a Code of Conduct on Countering Illegal Hate Speech Online with Facebook, Microsoft, Twitter and YouTube, including a pledge “to review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary”. In January 2018, the Commission released results suggesting that the companies have significantly improved their responsiveness to user reports of hate speech. While social media companies deleted only 28% of reported hate speech in 2016, it has since increased to 70%. Furthermore, the percentage of reports reviewed within 24 hours increased from 40% in 2016 to 81% in 2018<sup>8</sup>.

---

<sup>6</sup> Akkad, D., 2018. Revealed: Seven years later, how Facebook shuts down free speech in Egypt. [Online] Available at: <http://www.middleeasteye.net/news/how-facebook-bans-free-speech-egypt-activist-social-media-april-6th-mubarak-1685366161>

[Accessed 31 January 2018]

<sup>8</sup> European Commission, 2018. Countering illegal hate speech online #NoPlace4Hate. [Online] Available at: [http://ec.europa.eu/newsroom/just/item-detail.cfm?item\\_id=54300](http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=54300)

[Accessed 25 January 2018].

<sup>8</sup> European Commission, 2018. Code of Conduct on countering illegal hate speech online: Results of the 3rd monitoring exercise. [Online] Available at: [http://ec.europa.eu/newsroom/just/document.cfm?doc\\_id=49286](http://ec.europa.eu/newsroom/just/document.cfm?doc_id=49286)

[Accessed 31 January 2018].

## WAYS FORWARD FOR THE UK

The UK now has an opportunity to review its approach to tackling hate speech online. The Committee on Standards in Public Life released a report at the end of 2017 which advocated making social media companies liable in certain cases<sup>6</sup>. Although the Network Enforcement Law is a German law (and therefore subject to EU law), the Committee cited the European E-Commerce Directive as the reason social media companies are “exempt from criminal or civil liability when their services are used to commit an offence – for example, publishing or transmitting illegal content” in the UK. According to the Committee, Brexit could provide an opportunity for British lawmakers to reshape legal liability for social media companies, as the E-Commerce Directive would no longer apply.

Unintended results of the Network Enforcement Law in Germany make such a strict legislative approach unappealing. However, the UK could consider implementing company liability in some cases with a more detailed definition beyond “evidently unlawful content”. The European Commission’s Code of Conduct serves as a good example of how to make companies responsible within a clear structure without making them legally liable. The UK should work to monitor the responsiveness of social media companies to user reports of hate speech, especially after it leaves the EU. This would give the government a measurable indicator upon which it can discuss the issue of hate speech with companies. It would also prevent companies from assuming they will be held to a lower standard in the UK than in the EU after Brexit.

---

<sup>6</sup>Government of the United Kingdom, 2017. Intimidation in Public Life: A Review by the Committee on Standards in Public Life. [Online] Available at: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/666927/6.3637\\_CO\\_v6\\_061217\\_Web3.1\\_\\_2\\_.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/666927/6.3637_CO_v6_061217_Web3.1__2_.pdf) [Accessed 25 January 2018].

The issue of online hate speech is only part of the much wider issue of the nature of the relationship between social media companies, governments and users/citizens. Challenges include data protection and privacy, as well as the proliferation of fake news, criminal and terrorist activity and hate speech and abuse. Governments need to catch up and be more responsive to the capabilities and consequences of developing technology. It will also require greater awareness by citizens about their rights and responsibilities online. For this reason, governments and civil society should invest in education about the digital citizen. Digital rights and digital responsibility should be taught to children and young people in schools.

# BIBLIOGRAPHY

- Akkad, D., 2018. *Revealed: Seven years later, how Facebook shuts down free speech in Egypt*. [Online] Available at: <http://www.middleeasteye.net/news/how-facebook-bans-free-speech-egypt-activist-social-media-april-6th-mubarak-1685366161> [Accessed 31 January 2018].
- Amnesty International UK, n.d. *Black and Asian women MPs abused more online*. [Online] Available at: <https://www.amnesty.org.uk/online-violence-women-mps> [Accessed 25 January 2018].
- European Commission, 2018. *Code of Conduct on countering illegal hate speech online: Results of the 3rd monitoring exercise*. [Online] Available at: [http://ec.europa.eu/newsroom/just/document.cfm?doc\\_id=49286](http://ec.europa.eu/newsroom/just/document.cfm?doc_id=49286) [Accessed 31 January 2018].
- European Commission, 2018. *Countering illegal hate speech online #NoPlace4Hate*. [Online] Available at: [http://ec.europa.eu/newsroom/just/item-detail.cfm?item\\_id=54300](http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=54300) [Accessed 25 January 2018].
- Facebook, n.d. *Community Standards*. [Online] Available at: <https://www.facebook.com/communitystandards#hate-speech> [Accessed 25 January 2018].
- Government of Germany, 2017. *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz - NetzDG)*. [Online] Available at: <https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html> [Accessed 25 January 2018].
- Government of the United Kingdom, 2017. *Intimidation in Public Life: A Review by the Committee on Standards in Public Life*. [Online] Available at: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/666927/6.3637\\_CO\\_v6\\_061217\\_Web3.1\\_2.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/666927/6.3637_CO_v6_061217_Web3.1_2.pdf) [Accessed 25 January 2018].
- Grassegger, J. A. & H., 2017. *Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children*. [Online] Available at: [https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms?utm\\_campaign=sprout&utm\\_medium=social&utm\\_source=facebook&utm\\_content=1498662906](https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms?utm_campaign=sprout&utm_medium=social&utm_source=facebook&utm_content=1498662906) [Accessed 25 January 2018].
- The Economist, 2018. *Germany is silencing “hate speech”, but cannot define it*. [Online] Available at: <https://www.economist.com/news/europe/21734410-new-social-media-law-causing-disquiet-germany-silencing-hate-speech-cannot-define-it> [Accessed 25 January 2018].
- Twitter, n.d. *Hateful Conduct Policy*. [Online] Available at: <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy> [Accessed 25 January 2018].
- YouTube, n.d. *Policies, safety, and reporting > Policy Center > Hate Speech*. [Online] Available at: <https://support.google.com/youtube/answer/2801939?hl=en> [Accessed 25 January 2018].