



## EU Survey on measures to tackle illegal content online - ADDENDUM

### ARTICLE 19 written comments on the European Commission's Recommendation of March 2018

1. ARTICLE 19: Global Campaign for Free Expression (ARTICLE 19) is an independent human rights organisation that works around the world to protect and promote the rights to freedom of expression and freedom of information. ARTICLE 19 has extensive experience commenting on legislative proposals or other measures to regulate online content. We have responded to the European Commission's consultations on notice and action, the Audio-Visual Media Services Directive and the Copyright Enforcement Directive, to cite just a few.<sup>1</sup> We have also commented on the European Commission's EU Code of Conduct<sup>2</sup> and its Communication on tackling illegal content online.<sup>3</sup> Most recently, we have developed policy proposals to tackle the regulation of speech by contract in *Sidestepping Rights: Regulating Speech by Contract* (2018)<sup>4</sup> and made proposals for self-regulation of social media.<sup>5</sup> In these written comments, ARTICLE 19 makes a few general observations about the Survey on measures to tackle illegal content online and highlights the serious shortcomings in the Commission's Recommendation of March 2018 (C(2018) 1177 final).

#### General observations

2. At the outset, ARTICLE 19 regrets that many of the questions in the survey on measures to tackle illegal content have been framed in such a tendentious manner. For instance, stakeholders are asked "what features of notice systems ... do you consider to be most effective for enabling hosting service provider to make diligent decisions on the content notified?", the unargued assumption being that hosting ought to make diligent decisions on content removal rather than an independent court or tribunal. In another example, stakeholders are asked whether or not they agree with the statement "I disagree that my content should be banned by the community standards or terms of service". Stakeholders are not afforded any opportunity to challenge the underlying assumption that community standards are either an effective or desirable means of tackling illegal content online.
3. The questions asked in the survey also implicitly favour the Commission's efforts in encouraging the further uptake and deployment of automated tools to remove illegal content, regardless of how misconceived those efforts may be. The survey questions also suggest that the Commission would like to see the trusted flagger scheme potentially leading to automatic removal of content. These are worrying developments, though sadly consistent with the serious shortcomings in the March 2018 Recommendation ('the Recommendation').

---

<sup>1</sup> Our responses are available upon request.

<sup>2</sup> <https://www.article19.org/data/files/medialibrary/38430/EU-Code-of-conduct-analysis-FINAL.pdf>

<sup>3</sup> <https://www.article19.org/resources/eu-fails-to-protect-free-speech-online-again/>

<sup>4</sup> <https://www.article19.org/wp-content/uploads/2018/06/Regulating-speech-by-contract-WEB-v2.pdf>

<sup>5</sup> ARTICLE 19, *Self-regulation and hate speech on social media platforms* (2018): [https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-%E2%80%98hate-speech%E2%80%99-on-social-media-platforms\\_March2018.pdf](https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-%E2%80%98hate-speech%E2%80%99-on-social-media-platforms_March2018.pdf)

## Proactive measures

4. ARTICLE 19 is dismayed that the Recommendation actively promotes “proactive measures”, i.e. the deployment of filters to detect and remove illegal content (see paragraph 18). In our view, this is inconsistent with the general prohibition on general monitoring under Article 15 of the Electronic Commerce Directive and the decisions of the Court of Justice of the European Union (‘CJEU’) in the *SABAM* cases.<sup>6</sup> Although we note that the Recommendation is careful to caveat the use of filters to circumstances where it is “appropriate and proportionate” and subject to “effective and appropriate safeguards”, we believe that the Commission is being disingenuous. Although “where appropriate” may suggest that the use of filters would be limited to certain types of content, the reality is that filters are increasingly being used across the board: first they were used to tackle child abuse images, then rights holders demanded similar technology to deal with copyright infringement and now governments are putting pressure on intermediaries to use automated means to detect and remove terrorist or ‘hate speech’ material. In the case of terrorist material, intermediaries are encouraged to prevent terrorist videos from being uploaded in the first place (see para. 37). All this is despite the fact the CJEU clearly said that it would be disproportionate for intermediaries to monitor content with a view to identifying and removing copyright infringement.<sup>7</sup>
5. ARTICLE 19 considers the use of filters to detect and remove content to be problematic because, with the exception of child abuse images, the question whether content amounts to ‘hate speech’ or ‘incitement to terrorism’ is highly context-specific. No matter how sophisticated they may appear, filters are inherently incapable of making these kinds of assessments. In the case of copyright, filters are equally ill-suited to make evaluative determinations, e.g. whether the content at issue may fall under one of the limitations to copyright such as fair use or parody. For this reason, filters are likely to lead to many false positives and/or false negatives, resulting in the removal of wholly lawful and legitimate expression.

## Cooperation between hosting providers and Member States

6. ARTICLE 19 further notes that the Commission seeks to encourage greater cooperation between hosting providers and Member States. Whilst this is understandable to some degree, we are concerned that this cooperation merely seeks to speed up the removal of content on the basis of intermediaries’ terms of service rather than the law. In other words, it appears that the Commission seeks to encourage the removal of content that is not illegal but may fall foul of companies’ community guidelines. This is likely to be the case of ‘hate speech’, a poorly defined concept, which generally falls short of the legal threshold of ‘incitement to discrimination, hostility or violence’ under international law.<sup>8</sup> Legitimate content is highly likely to be removed as a result. Although in theory companies can refuse to comply with notices made under their terms of service, in practice, they are put under pressure to remove this type of content immediately, lest they face actual regulation.
7. ARTICLE 19 is also concerned that in focusing law enforcement efforts on the removal of online content, precious resources are being diverted from effective policing on the ground and, where appropriate, the prosecution of those responsible for the illegal content at issue.

---

<sup>6</sup> See [C-70/10](#), *SABAM v Scarlet Extended*, 24 November 2011 and [C-360/10](#), *SABAM v Netlog*, 16 February 2012.

<sup>7</sup> *Ibid.*

<sup>8</sup> See Article 20 of the International Covenant on Civil and Political Rights. See also ARTICLE 19, [Hate Speech Explained: A Toolkit](#), (2015).

8. ARTICLE 19 believes that it would be both unnecessary and counter-productive to require hosting providers to notify law enforcement of evidence of alleged serious criminality. This requirement does not exist offline across all the Member States of the European Union and there is no reason to create it online. It would impose an undue burden on Internet intermediaries.

### **Trusted flaggers**

9. ARTICLE 19 notes that the Recommendation seeks to rely on and further develop the trusted flagger scheme that already exists in several major Internet companies (e.g. YouTube). ARTICLE 19 remains very skeptical of this scheme, however. Trusted flaggers are not independent. Nor is the so-called 'expertise' of such flaggers a sufficient yardstick by which to assess the legality of content. Anti-discrimination advocates are unlikely to come to the same conclusion as free speech advocates in borderline 'hate speech' cases, for instance. We are especially concerned that the survey questions seem to suggest that notices issued by trusted flaggers could lead to immediate removal. In our view, this would be entirely outwith due process rights. The same rights that apply offline must also be protected online.

### **Specific recommendations related to terrorist content**

10. ARTICLE 19 is especially concerned that the Recommendation promotes the removal of terrorist content within an hour. In our view, this is a ridiculously short time-frame in which to make a determination as to the legality of content in borderline cases. In particular, we note that 'incitement to terrorism' is far from being a straightforward issue. Individuals' speech should not be removed when it falls short of inciting violent action. Given the extremely short timeframe within which companies are asked to remove that type of content, this can only mean that legitimate expression is bound to be caught in the process. It is easy to see, for example, how videos uploaded by human rights activists to denounce human rights abuses by terrorist organisations might be taken down under the rubric of countering extremism.
11. More generally, we are doubtful that the immediate removal content is a silver bullet in the fight against terrorism. In our view, immediate removal is likely to stigmatise legitimate expression by Arabic speaking and/or Muslim communities, which could make it more difficult for law enforcement agencies to build the trust necessary to engage in de-radicalisation efforts.