Information Law and Policy Lab

Social media councils for online content moderation

Safeguarding social media users' freedom of expression through self-regulation

Amsterdam, 20 December 2019

This position paper is written by Lisanne Bruggeman of the Glushko & Samuelson Information Law and Policy Lab (ILP Lab) of the Institute for Information Law (IViR) of the University of Amsterdam. The ILP Lab is a student-run, IViR-led institution which develops and promotes research-based policy solutions that protect fundamental rights and freedoms in the field of European information law.

This position paper reflects the recommendations and conclusions of the author. The author is grateful for the input received from Ot van Daalen and Ronan Fahy.

Executive summary

This position paper provides recommendations on the structure of 'social media councils' (SMCs), taking into account the freedom of expression of social media users. An SMC is a self-regulative initiative to address online content moderation issues.

In short, an SMC should be structured as follows:
- It should contain an appeal mechanism which is in principle accessible to all users;
- It should apply a broad human rights framework, but be limited to assessing the application of platforms' terms and conditions;
- It should be established on a global level and should have national sub-departments;
- It should use the European Convention on Human Rights or the International Covenant on Civil and Political Rights as a framework;
- Major social media platforms such as Facebook, Instagram, YouTube and Twitter should take part in the creation and operation;
- The board of the SMC should consist of people from the judiciary, in addition to representatives of social media companies, the media, journalists, bloggers, digital rights organisations, the academic sector and minority and vulnerable groups; and
- The SMC could follow (inter)national rules regarding geographical jurisdiction and applicable law.

1. Introduction

People generate more content than ever before, especially through social media. The regulation of this user generated content is largely in the hands of platform providers. Their terms of service determine to a great extent which content may be uploaded, what should be removed and which user accounts must be terminated. As platforms have an incentive to prevent the sharing of undesirable content on their platforms, private censorship is a serious risk. As a consequence, the way in which platforms moderate content is increasingly being discussed and criticised.

Current content moderation by social media platforms has in particular the following shortcomings:
- limited application of human rights considerations;
- non-transparent processes, rules and decisions;
- inconsistent enforcement of terms and conditions;
- lack of accountability; and
- limited or non-existent appeals and remedies.[1]

Self-regulation is one method to remedy the above shortcomings.[2]

Examples of self-regulation in the (social) media sector are press councils and Facebook's proposal to create an independent Oversight Board. Press councils are bodies for complaints regarding journalistic practices, but focused on the traditional media. Facebook's Oversight Board would only apply to Facebook, has a questionable degree of independence and contains an appeal mechanism limited to selected cases.[3] In addition to these initiatives, the creation of a 'social media council' (SMC) as proposed by ARTICLE 19[4], is aimed at social media platforms in general and seems to be a good option to tackle the above shortcomings in online content moderation. Under this proposal, an SMC is defined as "a model for a voluntary accountability mechanism that would provide an open, transparent, accountable and participatory forum to address content moderation issues on social media platforms, on the basis of

---

[1] H. Tworek, *Social Media Councils*, Centre for International Governance Innovation 28 October 2019, https://www.cigionline.org/articles/social-media-councils; H. Tworek, N. van Eijk, R. Ó Fathaigh, L. Bruggeman & C. Tenove, *Dispute Resolution and Content Moderation: fair, accountable, independent, transparent and effective*, Transatlantic Working Group 2019 (draft version of 22 October 2019), p. 2 and *Report of the special Rapporteur on the promotion and protection of the right to freedom of opinion and expression,* United Nations Human Rights Council 6 April 2018, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement, in particular para. 26, 38 and 40.

[2] See M. Price & S. Verhulst, *The Concept of Self-Regulation and the Internet*, University of Pennsylvania, Annenberg School for Communication 2000, https://repository.upenn.edu/cgi/viewcontent.cgi?article=1143&context=asc_papers, p. 13-14 and *The Social Media Councils: Consultation Paper*, ARTICLE 19 2019 (version of June 2019), https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf, p. 5-6.

[3] See B. Harris, *Establishing Structure and Governance for an Independent Oversight Board*, Facebook 17 September 2019, https://about.fb.com/news/2019/09/oversight-board-structure/.

[4] ARTICLE 19 is a worldwide non-governmental organisation on freedom of expression and information, see for more information: https://www.article19.org.

international standards on freedom of expression and other human rights".[5] However, an SMC can still come in various models, depending on its functions, subject matter, scope and applicable standards. So how should an SMC be structured?

This position paper answers this question. The paper will first discuss the functions, thereafter the subject matter, then the scope and standards, and finally the specifications of an SMC.

2. Functions of an SMC

The main function of an SMC is to provide for an appeal mechanism which is in principle accessible to all users. Users currently have few effective options to challenge platforms' content decisions. Importantly, an appeal mechanism can empower users against the power of large platforms.

An SMC can have several other functions. Firstly, it can provide a forum to develop and share best practices and code of conducts for platforms.[6] An SMC can also promote cooperation between platforms. This would allow an SMC to address the issues related to online content in an integrated and consistent way across platforms. The functions of press councils can be taken as example,[7] although press councils deal with content written by journalists instead of content written by users.

3. Subject matter

The SMC should apply a broad human rights framework to protect social media users' rights, as for example not only the right to freedom of expression can be affected in content moderation decisions, but also the right to privacy. Besides, it should cover social media providers' decisions on deletion, as well as upholding of user generated content and user accounts. However, an SMC should be limited to assessing the application of platforms' terms and conditions, since the assessment of national criminal law must remain a matter of national courts.

4. Scope

The best option would be to establish an SMC with global jurisdiction, and with national sub-departments. A global SMC has the benefit of uniformity for platforms and their users. Furthermore, removals as part of platforms' terms and conditions generally apply globally.[8] However,

---

[5]   *The Social Media Councils*: *Consultation Paper*, ARTICLE 19 2019 (version of June 2019),
https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf, p. 7-8.

[6]   H. Tworek, *Social Media Councils*, Centre for International Governance Innovation 28 October 2019, https://www.cigionline.org/articles/social-media-councils.

[7]   See *The Media Self-Regulation Guidebook* from the OSCE Representative on Freedom of the Media, the Organization for Security and Co-operation in Europe 2008, https://www.osce.org/fom/31497?download=true, p. 46.

[8]   *Report of the special Rapporteur on the promotion and protection of the right to freedom of opinion and expression,* United Nations Human Rights Council 6 April 2018, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement, para. 19.

national departments would be better able to respond to the cultural and linguistic context of a case and the different freedom of expression standards.[9]

5. Standards

A difficult question to answer is whether an SMC should apply international standards on human rights or a privately developed code of human rights principles. While a private code can be tailor-made to the online environment, drafting it will be challenging and time-consuming.

The European Convention on Human Rights (ECHR) and the International Covenant on Civil and Political Rights (ICCPR) are, on the other hand, well established frameworks with a voluminous body of case law. The right to freedom of expression under the ECHR does not only include the obligation to prevent violations. It also provides for a duty to take protective measures. This also applies to relations between individuals.[10] Although the provisions of the ECHR and ICCPR are addressed to Member States, they could be applied in a similar way between social media platforms and its users. Content moderation restrictions should then, for example, (1) be provided by the platforms' (clear and accessible) terms and conditions, (2) pursue a legitimate aim and (3) be necessary and proportionate. It is still to be determined whether platforms should permit all legal speech.[11]

6. Specifications

Major social media platforms such as Facebook, Instagram, YouTube and Twitter should definitely take part in the creation and operation of an SMC as they have the largest number of users and therefore the greatest impact. In addition, the SMC should also be open to smaller platforms in order to ensure maximum accessibility.

The board of the SMC should include representatives of social media companies, the media, journalists, bloggers, digital rights organisations, the academic sector and minority and vulnerable groups, as well as people from the judiciary to guarantee a certain degree of independence. This is for example is also the case with the Dutch Council for Journalism.[12]

Regarding the geographical jurisdiction and applicable law, the SMC could follow applicable (inter)national rules. On EU level, the rules as laid

---

[9]   H. Tworek, N. van Eijk, R. Ó Fathaigh, L. Bruggeman & C. Tenove, *Dispute Resolution and Content Moderation: fair, accountable, independent, transparent and effective*, Transatlantic Working Group 2019 (draft version of 22 October 2019), p. 20.
[10]   ECHR 12 September 2011, application numbers 28955/06, 28957/06, 28959/06 and 28964/06 (*Sánchez and Others v. Spain*), para. 59.
[11]   D. Keller, *Who Do You Sue?*, Hoover Institution Stanford University 2019, https://www.hoover.org/sites/default/files/research/docs/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech_0.pdf, p. 24.
[12]   https://www.rvdj.nl/english (consulted on 9 December 2019).

down in the Brussels Ibis Regulation[13] and Rome I Regulation[14] could be followed.

In order to make an SMC as accessible as possible for users, an appeal mechanism should at least entail the following aspects:
- The entire process should be electronic to ensure full accessibility for social media users;
- Deadlines must be reasonable and not too long;
- Adjudicators should automatically add the human rights basis as many social media users are not legally skilled;
- There should be a focus on issuing a decision rather than reaching a settlement, as it is difficult to see how settlements within this specific context could be reached;
- Rules on expertise, independence, impartiality, transparency, effectiveness and fairness (in the EU included in i.a. the Directive on alternative dispute resolution for consumer disputes[15]) must be taken into account to guarantee a fair process;
- As far as the process is concerned, a look can be taken at the EU's online dispute resolution platform[16] and the EU's small claims procedure[17];
- It is important that the possibility of starting court proceedings is not excluded; and
- Special attention should be paid to the equality of arms between social media providers and users.

---

[13] Regulation (EU) 1215/2012, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32012R1215&from=NL.
[14] Regulation (EC) 593/2008, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008R0593&from=nl.
[15] Directive (EU) 2013/11, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32013L0011.
[16] Regulation (EU) 524/2013, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32013R0524.
[17] Regulation (EC) 861/2007, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32007R0861&from=NL and Regulation (EU) 2015/2421, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32015R2421.