France: ARTICLE 19 comments on interim report for social media regulation

In May 2019, the French government published an interim report on social media regulation. It is one of a range of regulatory proposals in this area, including in the United Kingdom, Australia, Ireland and Canada. ARTICLE 19 welcomes the Report's positive framing and declared commitment to ensure protection of human rights and freedom of expression online. The focus on accountability by design and recognition of fundamental rights as the basis of any co-regulatory mechanisms are particularly laudable. As the French government fully develops this proposal, it should ensure that any proposals to regulate social media do not entrench private censorship, mandatory filters and content take-downs within unreasonable timeframes and does not consolidate the dominance of the major tech companies. ARTICLE 19 also suggests closer investigation of contractual and technical unbundling to reduce market entry barriers and increase the number of markets where diversity of service and products is offered.

In May 2019, the French government published an interim report (the Report) setting out a broad framework for holding social media platforms accountable for the content they host online. The Report reflects the provisional conclusions of an inter-ministerial working group after a few months spent at Facebook and several meetings with a range of stakeholders, from independent public authorities to private actors and NGOs. The working group also met with its British and German counterparts to better understand their respective regulatory models and initiatives. The French government is expected to take these proposals forward and flesh them out in a final report that will be published at the end of June. The Report also comes on the heels of a French Private Members Bill on Hate Speech Online,[1] which ARTICLE 19 analyses in a separate document.

The proposals

The Report focuses on the regulation of social networking services defined as '*an online service allowing its users to publish content of their choice and thereby make them accessible to all or some of the other users of that service*'. Its stated purpose is to force the 'biggest players' to be more responsible and take a more protective attitude to social cohesion. In doing so, the Report sets out a general framework based on five pillars with a special focus on algorithm transparency:

1. **Civil liberties:** the protection of civil liberties and the freedom to conduct business would sit at the top of the new framework's objectives. It explicitly mentions both individuals' freedom of expression and communication and the freedom of users to be protected in their physical and moral integrity (i.e. the

[1] See so called Proposition de loi Avia, National Assembly, Legislative Proposal to fight against hatred on the internet, Referred to the Committee on Constitutional Law, Legislation and General Administration of the Republic, March 2019.

right to privacy) alongside the freedom for companies to innovate and the right to define and apply their terms of use.

2. **Accountability:** the overarching goal of any regulation would be to hold social networks accountable by means of an independent administrative authority. Such an authority would oversee the implementation of transparency obligations imposed on the biggest platforms in relation to the way in which they order content and how they implement their Terms of Service and the moderation of content (e.g. how community standards are elaborated, use of automated processing tools, trusted flagger programmes etc). In addition, the regulator would oversee how companies implement a duty of care towards their users. These obligations would be applicable to services with a number of monthly users rising above a fixed percentage of the population (10-20%). Services with monthly users falling below 5% of the population would not be subject to these obligations and would continue to be governed by Law no. 2004-575 of 21 June on confidence in the digital economy (LCEN). In case of 'identified and persistent' breaches, however, these services could also be made subject to the new more stringent regime by the regulator.

3. **Dialogue**: the government would play a central role in setting the threshold for triggering obligations and defining those obligations, including an obligation to defend the integrity of the social network and its members. It would also organize with the regulator a political dialogue between social media companies, the government, the legislature and civil society. In practice, this could take the form of 'voluntary' commitments made by companies to the government, which would then be verified and enforced by the regulator.

4. **Powers:** the regulator would not be tasked with policing individual pieces of content or defining the types of services that could be offered. Rather, it could require the publication of certain types of information consistent with transparency obligations. It would also assess the effectiveness of the resources deployed to comply with transparency obligations and the obligation to protect the integrity of the network and users. It would have access to sensitive information held by the platforms, including special access to their algorithms to verify the accuracy of information published by the platforms about them. Failure to comply could be meted out with fines (percentage of global turnover) following formal notice. The independent administrative authority would act in partnership with other branches of the state, and be open to civil society.

5. **European cooperation:** the proposal would be a stepping stone towards greater European cooperation so as to hold global platforms to account whilst reducing the 'political risks' related to implementation in each Member State.

The Report makes clear that to minimize any risks to civil liberties, any legal framework for platform regulation should be consistent with a number of precautions, including:

- Respect for platform diversity;
- Transparency obligations with systematic inclusion of civil society;
- Minimum regulatory intervention consistent with the principles of necessity and proportionality; and
- Judicial characterization of lawfulness of individual items of content.

Finally, the Report highlights that the implementation of any "*ex-ante*" regulatory framework should:

- Adopt a compliance-based approach, whereby the regulator would only supervise the correct implementation of preventive or corrective measures rather than focus on the materialisation of risks or try to regulate the service provided;
- Focus on 'systemic' actors capable of creating significant damage to democratic societies, without creating barriers to entry for new European operators; and
- Be flexible in order to address future challenges as they arise. In this sense, the purpose of any laws in this area should be to create an institutional capacity to regulate rather than directly regulating current problems.

ARTICLE 19's comments and recommendations

ARTICLE 19 generally welcomes the constructive tone of the Report. Its framing rightly focuses on transparency and accountability as the core objectives of any framework for social media regulation. Placing the protection of individual freedoms at the heart of any regulatory framework of social networks is equally welcome. We also appreciate that the proposals do not ostensibly seek to regulate 'harmful' content online but seek to focus instead on transparency obligations. In our view, its focus on 'dominant' (biggest) social media actors is a more practical approach to the challenges raised by content moderation rather than seeking to regulate a wider range of operators in the Internet ecosystem, such as public discussion forums, online product review sites or cloud hosting providers. Recognising the need for inclusive dialogue between companies, government, regulators and civil society is also a positive step forward.

In the absence of further detail, however, ARTICLE 19 recommends further elaboration of the proposal to ensure that the French model avoids raising the same substantive issues as its German,[2] British[3] and EU[4] counterparts. We therefore make a number of suggestions in order to ensure full protection of human rights, including the right to freedom of expression.

- **Duty of care/due diligence**: at present, the Report posits that social media companies – particularly the dominant ones – should owe a 'duty of care' to their users. At this stage, the Report does not flesh out what the duty of care would entail. ARTICLE 19 believes that reliance on the concept of 'duty of care' is problematic: it is often based on analogies with duties of care owed in the physical world that do not tally well with the way in which people interact online.[5]

  If the duty of care/due diligence concept is maintained, at the very least, any duty of care should be defined sufficiently narrowly in order to meet the three-part test for restrictions on the right to freedom of expression under international law.[6] In

[2] ARTICLE 19, Germany: Act to Improve Enforcement of the Law on Social Networks undermines free expression, September 2017.
[3] ARTICLE 19, UK: ARTICLE 19 response to leaked reports on online harms white paper, April 2019.
[4] ARTICLE 19, Joint letter on European Commission regulation on online terrorist content, December 2018.
[5] The duty of care often relates to the foreseeability of physical injury in the context of health of safety but it is considerably broader when applied to 'online harms' that often have to do with emotional distress in response to something that was said.
[6] Under international law, any restriction on freedom of expression must (i) be provided by law; (ii) pursue a legitimate aim as exhaustively listed under Article 19 (3) of the International Covenant on Civil and Political Rights 1966 or Article 10 (2) of the European Convention on Human Rights; and (iii) be necessary and proportionate in a democratic society.

particular, in order to meet the legality test, any such duty of care should clearly define which operators are subject to the duty, to whom it is owed, under what circumstances operators are expected to act to protect users, what type of harms users must be protected against, what steps operators are expected to take to prevent or mitigate the risks of harm identified etc.[7]

Although the Report points out that it is not concerned with individual decisions to remove or not to remove content, any systemic failure to deal with 'categories' of content is bound to be linked to a series of individual decisions not to remove content (e.g. the Nancy Pelosi fake video) or a failure to remove an aggregate number of items of content. In this regard, the video of the Christchurch terrorist attack could be an interesting test case. Facebook said that it removed 1.5 million videos within 24 hours[8] of the attack yet hundreds of thousands remained on the platform. Facebook had a policy in place to deal with terrorist content. It also provided a live streaming feature. This then raises the following questions:

- Should live streaming features be entirely disabled to prevent 'harm'?
- Leaving aside prior censorship issues, how is it possible to identify the intent of those who use live streaming prior to upload?
- What if the intent was not to incite violence but rather to document it or raise the alarm?
- What if illegal content occurs after 1 hour of watching a video as opposed to the first few minutes?
- How long should it take for platforms to remove that kind of content?

At present, it is unclear what circumstances would trigger a duty of care, what measures the companies would be expected to take and the point at which a regulator might decide that a failure is sufficiently systemic to warrant severe fines. As noted above, ARTICLE 19 recommends that the final proposal provides clarity on these points. Failure to do so would raise serious legal certainty issues that would ultimately be harmful for freedom of expression.

- ***Ex ante* regulation.** We further note that the Report currently makes reference to *ex ante* regulation and the implementation of 'preventive' or 'corrective' measures without defining them. Based on ARTICLE 19's experience, 'preventive' in the field of content moderation, tends to mean monitoring obligations – whether general or 'specific' - that are both overly broad and inconsistent with Article 15 of the E-Commerce Directive (prohibition on general monitoring). It also usually entails the prevention of content uploads, at least for certain categories of content, which are generally loosely defined such as 'terrorist' or 'extremist' content. These kinds of preventive measures are akin to prior censorship, particularly when mandated by the state, whether directly or indirectly; they are also inconsistent with the protection of the right to freedom of expression. Meanwhile, 'corrective' measures tend to imply content take-downs within incredibly short timeframes and making it easier to report 'problematic' content without regard to either due process safeguards or freedom of expression.

  In our view, the government should refrain from mandating or recommending the use of upload filters to fulfill companies' 'duty of care' to their users. Equally, it should refrain from imposing unduly short time-frames within which to remove

[7] For more details about what a duty of care should entail, see Graham Smith, A Ten Point Rule of Law Test for a Social Media Duty of Care, 16 March 2019

[8] See , e.g. C.Lecher, YouTube took down an 'unprecedented volume' of videos after New Zealand shooting, 18 March 2019.

'hate speech' or 'terrorist' content or other types of illegal content. Instead, it should focus on improving due process safeguards in the enforcement of take-down decisions and ensure that users are able to challenge decisions made by companies that affect them. One way of ensuring adequate safeguards in content moderation processes would be to rely on notice-and-notice systems for certain types of illegal content. Under this model, intermediaries forward the notices they receive about allegedly unlawful content to content providers. Equally, they pass on counter-notices if any. In the absence of counter-notices, they lose immunity from liability. This preserves the intermediary's role of intermediation and ensures a degree of procedural fairness that is entirely lacking in notice-and-takedown processes.[9]

- **Delegation of censorship and rule of law**: ARTICLE 19 notes that the Report contains a number of positive features in this area: it does not suggest that a regulator should be developing codes of conduct in relation to 'harmful' but legal content. Nor does it suggest that immunity from liability should be repealed. Instead, the report seems to lay emphasis on 'improving' and legitimizing current self-regulatory initiatives developed by companies and greater transparency of their terms of service. Reference is also made to the courts' role in determining unlawful content.

  At the same time, failure by companies to live up to 'voluntary' commitments could lead to severe fines (up to a percentage of global turnover) if they are not upheld. Equally, in the absence of codes of conduct for 'harmful' content, the default position is likely to be the application of companies' terms of service, which tend to be far more restrictive of freedom of expression than the law.[10] In these circumstances, the new regime would likely entrench private censorship by companies as companies typically tend to err on the side of caution to avoid sanctions.

  In ARTICLE 19's view, a better way forward would seek to empower users to choose the type of experience they wish to have on platforms by letting them control the types of content filters of community standards they wish to apply.[11] We recommend the French government looks into "unbundling" options,[12] perhaps similar to those already enacted in European utilities markets or to such changes that were recently introduced for European payment services through PSD2.[13] Broadly speaking, the big platforms would open up a neutral version of their service (i.e. without their ranking or community standards). Competitors would then offer a service where users could find the same content but choose to apply different ranking and removal policies.[14] We believe that French lawmakers and others should further explore this ideas as it could also help solve some of the concerns arising from the dominance of US tech companies.

- **Sanctions:** As noted earlier, failure to comply with an undefined 'duty of care' could lead to heavy fines – a percentage of global turnover – for social media companies.

[9] For more details, please see ARTICLE 19's policy, Internet Intermediaries: Dilemma of Liability, 2013.
[10] See ARTICLE 19, Analysis of Facebook's Community Standards, 2018.
[11] See M. Masnick, Platforms, Speech and Truth: Policy, Policing and Impossible Choices, Techdirt, 9 August 2018.
[12] See D. Keller, Who Do You Sue? State and Platform Hybrid Power Over Online Speech, 2019, p. 26.
[13] M. Zachariadis and P. Ozcan, The API Economy and Digital Transformation in Financial Services: The Case of Open Banking, SWIFT Institute Working Paper No. 2016-001.
[14] Ibid.

Whilst we welcome the absence of harsher measures seen elsewhere,[15] such as blocking of platforms that fail to comply, in our experience, the threat of heavy fines is a strong incentive for companies to remove more rather than less content. In such circumstances, legitimate content inevitably gets caught. In our view, a regulator's power to fine social media companies could only be compatible with the right to freedom of expression if it relates to clear and narrowly defined obligations and any fines imposed are proportionate to the gravity of the conduct at issue.

Rather than pursuing a highly punitive approach, ARTICLE 19 urges French lawmakers to explore alternatives that would reward companies for embracing higher standards of conduct. This could include kite-marking or grading that would enable the public to recognize companies that abide by higher standards of conduct. We would also encourage policy-makers to consider multistakeholder models which also include independent oversight over the content moderation practices of social media companies.[16]

ARTICLE 19 understands that the French government plans to release a final report on social media regulation at the end of June. As it sets an ambitious agenda for online content regulation at home and abroad, ARTICLE 19 stands ready to provide further assistance to ensure that the final proposal achieves the highest standards of protection for freedom of expression.

[15] For instance in the UK White Paper on Online Harms or the recent amendment to the Australian Criminal Code.
[16] C.f. e.g. ARTICLE 19, Self-regulation and 'hate speech' on social media platforms, 2018.