18 November 2019

Contribution to the examination of France's draft law aimed at combating hate content on the internet

European Digital Rights (EDRi) is an association representing 42 human rights organisations from across Europe that defend rights and freedoms in the digital environment. This submission has been developed with the contributions of our members ARTICLE 19 and Access Now.

Introduction

On 9 July 2019, the French National Assembly adopted a draft law on combatting hateful content online (hereafter the 'draft Avia law'). On 21 August, the French government notified the draft law to the European Commission in accordance with Directive (EU) 2015/1535. The French Senate is due to examine the draft law in December 2019.

The draft Avia law would impose several new obligations on a variety of 'online platform providers' in relation to an extensive range of illegal content under French law. EDRi's core submission is that the draft Avia law would seriously hinder the fundamental right to freedom of expression and opinion. It also risks fragmenting the Digital Single Market at a time where the Commission is looking to harmonise the rules that govern intermediary liability for user-generated content.

Should it be adopted, this national legislation would, in our view, pre-empt the necessary public debate that should take place at European level about common content moderation standards and processes that will apply to both transnational and national platforms.

It is also unclear how the draft Avia law relates to the implementation of the Audiovisual Media Services Directive and currently pending proposed Regulation combating the dissemination of online terrorist content. As such, it will almost certainly lead to greater legal uncertainty for both hosting providers and users in Europe.

Beyond concerns for the fragmentation of the Digital Single Market, the draft Avia Law would entrench the decision-making power of digital companies as to the legality of content with insufficient safeguards for the protection of users' rights to freedom of expression, due process and privacy, beyond France's territory. It is easy to see how the fear of high fines the draft law foresees will bring platforms to delete and block any content that appears to generate a risk of being punished under this new law, thereby **chilling freedom of speech online**. European Courts stressed that measures put in place to protect a public interest, including the protection of a fundamental right, must strike an appropriate balance with other fundamental rights[1]. **We do not see how a proposal that**

1 EctHR, Dupuis v France, App. No. 1914/02, 7 June 2007

undermines freedom of expression would pass that test. For this reason, we provide comments on specific elements of the draft Avia Law to which the European Commission should pay particular attention when developing its own proposal for a Digital Services Act.

With the upcoming reform of the rules governing intermediary liability for user-generated content, the EU has a unique opportunity to craft a model that will protect human rights and adequately address the challenges targeted by the French initiative. **Because of its role as the guardian of the Treaties, we suggest that the Commission responds to the French government with a detailed opinion expressing the incompatibility of this proposal with the current E-Commerce Directive and potential conflict with tthe upcoming proposal for a Digital Services Act**, announced by President-elect, Ursula von der Leyen, in her reform agenda for Europe[2].

A wrong balance of incentives

The draft Avia law requires hosting providers to remove hateful and other types of 'manifestly' illegal content within 24 hours of notice. It also imposes a series of transparency and procedural requirements, including the setting up of complaints mechanisms. In the event of non-compliance, the regulator may impose an administrative penalty of up to 4% of the overall turnover.

The 24-hour time-frame is plainly insufficient to allow the vast majority of companies falling within the scope of the draft law to assess a request properly, especially if they receive notices in high numbers. Financial sanctions that can potentially be imposed on companies for failure to remove content or comply with their other obligations under the draft law appear to be unduly high. Specially for those social networks that are not online giants (eg. Facebook, Google, Twitter, etc.), it will be easier to delete risky content on the basis of terms of service instead of spending hundreds of thousands of euros in training and hiring specialised staff to do that. **Overall, the foreseen provisions strongly encourage host providers to over-comply with notices and therefore, over-remove content, including legal content which is a risk for freedom of speech.**

The 24-hour time cap could have several other negative consequences. It strongly encourages providers to use automated moderation tools, which frequently carry with them significant risks of false positives and false negatives: it is well-established that they are not equipped to make complex judgments in cases, which are highly context-dependent.[3] They cannot tell infringement apart from legal uses like parody, counter-

ECtHR, Bayar v Turkey, App. Nos.39690/06, 40559/06,48815/06, 2512/07, 55197/07, 55199/07, 55201/07 and 55202/07, 25 March 2014
CJEU, Google v CNIL, C-507/17, 24 September 2019

2 Ursula Von der Leyen, *A Union that strives for more. My agenda for Europe*, 16 July 2019: https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

3 See CDT, *Mixed Messages*, November 2017: https://cdt.org/files/2017/11/Mixed-Messages-Paper.pdf

speech, or legitimate political dissent. Consequently, legal content will inevitably be taken down.

The 24-hour time frame is also likely to prevent hosting providers from processing certain types of requests as a matter of priority, for instance those involving content that is more likely to incite violence or discrimination or that significantly interferes with other users' rights or that may be restricted on national security, prevention of crime or public order grounds, and that is shared more widely than others. This is problematic when a specific piece of allegedly illegal content is suddenly massively shared and requires the full attention and resources of the host provider.

The current 2004 law on trust in the digital economy requires companies to remove illegal content 'promptly' upon notice. It therefore gives companies more flexibility to deal with complaints according to the urgency of the situation and their resources. In our view, this is a better approach and one that is in line with the requirements of international human rights law and the Charter of Fundamental Rights that any restriction on freedom of expression must be necessary and proportionate, i.e. the least restrictive measure to pursue one of the legitimated aims exhaustively listed under relevant human rights treaties.[4] **We also remind legislators at the national and European level that Member States have a positive obligation to protect fundamental rights of individuals both offline and online. Delegating this legal obligation to private actors is wholly inappropriate.**

Working with vague definitions of illegality

The original text proposed by Laetitia Avia MP only covered content that manifestly insults or incites to discrimination, hatred or violence on grounds of race, religion, ethnicity, sex, gender, sexual orientation or disability. The latest version of the proposal extends the scope of offences to other existing ones under French criminal law such as apology of acts constituting an offence against human dignity, war crimes, crimes against humanity, slavery, crimes of collaboration with an enemy, voluntary interference with life or physical integrity, sexual aggression and harassment, aggravated theft, extortion or destruction, voluntary degradation or deterioration which is dangerous to a person, human trafficking, pimping, incitement to or apology of acts of terrorism and child abuse content.[5]

For the most part, these types of content can legitimately be restricted under international and European human rights law. However, as pointed out by the UN Special Rapporteur on

4 Article 10 (2) of the European Convention on Human Rights and Article 52 of the European Charter of Fundamental Rights. See also La Quadrature du Net, PPL Avia, refusez les mesures inutiles et dangereuses, 27 June 2019: https://www.laquadrature.net/wp-content/uploads/sites/8/2019/06/lqdn_analyse_ppl_avia_27_juin.pdf

5 The following provisions are covered: 5th, 7th and 8th paragraphs of article 24 and 3rd and 4th paragraph of article 33 of the law on freedom of the press of 29 July 1881, articles 222-33, 225-4-1, 225-5, 225-6, 227-23, 227-24 and 421-2-5 of French criminal law.

the promotion and protection of the right to freedom of opinion and expression[6], provisions of the French criminal law relating the criminalisation and definition of acts constituting terrorist crimes, provocation and apology of terrorism are excessively broad and vague. They allow for an arbitrary and abusive interpretation of the law that could lead to violations of human rights once the law is implemented. Such provisions in France have been used not only to prosecute individuals making jokes in poor taste[7] but also young children who had made ill-timed, shocking statements – but that did not meet the criteria defining apology of terrorism[8]. It is essential that restrictions on fundamental rights are provided by law in clear and precise terms to provide people and intermediaries with legal certainty and enable them to regulate their behaviour accordingly. **Vague and overbroad definitions of criminal offences lead to the removal of legitimate content. They also amount to disproportionate restrictions on the right to freedom of expression and information.**

A wide range of host providers falling within the scope of the law

The personal scope of the draft Avia law is extremely broad: it applies to all platform operators that propose a communication service to the public or use algorithms to rank and reference third party-generated content and that meet certain thresholds of 'activity' on French territory as determined by decrees. There is precise way to define this threshold of activity in the draft law, though parliamentary debates make reference to a number of users rather than the annual turnover of companies. Although the original version of the draft law was conceived to be applicable to the dominant tech companies, the latest text gives great latitude to the French government to decide the scope of application of the rules, including to small local providers. Meanwhile, the draft law makes no provision for a tiered or proportionate approach depending on, e.g. the size or resources of the provider at issue.

In practice, this means that social media companies, search engines as well as not-for-profit platforms fall within the scope of the law. Providers such as Wikipedia would be required to have costly content moderation measures in place because they meet the thresholds set by decrees. This could also be the case for small providers. In our view, **this is deeply concerning for not-for-profit and small platforms as they are unlikely to have the resources to meet their obligations under the law.** In the case of Wikipedia, this would be both disproportionate and counterproductive as Wikipedia already has an efficient content moderation model based on the work of volunteers.

Re-upload filters and the risk of imposing a general monitoring obligation

6 Communication of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression of 3 February 2015 on the law no.2014-1353 on counter-terrorism: https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=15557

7 http://hudoc.echr.coe.int/eng?i=001-194114

8 Communication of the UN Special Rapporteur, see above.

In its latest version, the French draft Avia law includes a provision[9] imposing host providers to put in place "appropriate measures to prevent the re-distribution of content" that has already been deemed 'manifestly' illegal under the same law. The wording suggests the use of automated tools and the creation of hash databases that contain digital hash "fingerprints" of every piece of content that host providers have identified as 'manifestly' illegal and removed. In practice, these so-called "re-upload filters" mean that all user-generated content published on the intermediaries' services is monitored and compared with the material contained in those databases to be filtered out in case of a match. In our view, this **amounts to an obligation of general monitoring, which is prohibited under Article 15 of the E-Commerce Directive.** This is particularly problematic in circumstances where no judicial determination has been made that the content is indeed illegal.

We have long argued that filters are problematic because they are unable to understand the context in which content is published and shared, and they are therefore error-prone. Such algorithmic tools do not take proper account of the legitimate use of the content, for example for educational, artistic, journalistic or research purposes, for expressing polemic, controversial and dissident views in the context of public debates, or in the framework of awareness raising activities. As such, they risk suppressing legal speech accidentally.

Filters are also likely to have a disproportionate impact on already vulnerable individual users. The content of the databases is highly likely to reflect discriminatory societal biases. Certain types of content and speech are getting more reported than others and the decision by the intermediary to characterise them as illegal and to add them to databases often mirrors societal norms. It then influences the way algorithms work and screen content. As a result, content related to Islamic terrorism propaganda will be more likely targeted than white supremacist content[10] - even when the former may be documentation of human rights violations or serving an awareness-raising purpose against terrorist recruitment. As repeatedly stressed by EDRi[11], **databases, which are not accountable, transparent and democratically audited and controlled, will likely disadvantage certain users based on their ethnic background, gender, religion, language, or location.**

In addition, **the effectiveness of re-upload filters is questionable**: they are easy to circumvent on mainstream platforms: Facebook said that it had over 800 distinct edits of the Christchurch shooting video in its hash database because users constantly modified the original material in order to trick automatic identification.

Lastly, hash databases and related algorithms are being developed by dominant platforms, which have the resources to invest in such sophisticated tools. **Obliging all other actors on**

9 http://www.assemblee-nationale.fr/dyn/15/amendements/2062/AN/349
10 https://www.vice.com/en_us/article/a3xgq5/why-wont-twitter-treat-white-supremacy-like-isis-because-it-would-mean-banning-some-republican-politicians-too
11 https://edri.org/open-letter-on-the-terrorism-database/

the market to adopt such databases risks reinforcing their dominant position.

Restrictions on access to "mirrored" content

Article 6 of the draft law gives powers to an administrative authority to order internet access providers (IAP) to block access to websites or servers that make available mirrored content declared illegal by a court. In our view, **this measure is disproportionate.**

To begin with, mirrored content is unlikely to be illegal in all instances. The meaning of a post is likely to change significantly depending on the context and motivation of the user posting it. The list of possible legitimate re-uploads is countless, from journalism to satire, humour, and academic purposes. Even if the initial post was deemed illegal at first, this might not be the case of subsequent re-uploads. Automated measures are currently unable to assess context and will inevitably fail to grasp such nuance, no matter how much they improve. For this reason, digital companies should not be penalised for failing to remove content, in circumstances where it may well be lawful.

Secondly, even where companies fail to remove some instances of content deemed illegal by a court, any sanction applied against them should be proportionate. Blocking an entire platform because it may contain some pieces of illegal content is almost certainly never going to be a proportionate response. Moreover, it would significantly penalise other users who want to gain access to lawful information on the site.

EDRi hopes that our comments on the draft French law aimed at combating hateful content on the internet will make a useful contribution to the Commission's detailed opinion addressed to France in the context of the 2015/1535 notification procedure. We will keep contributing to the debate on intermediary liability and the Digital Services Act with a view to defending people's digital rights.

Fragmentation of the Digital Single Market

In recent years, regulatory proposals have multiplied at both national and European levels to address different types of content online, including terrorist content, copyright infringement, misinformation and illegal hate speech. Each of these initiatives usually involves different obligations in terms of content qualification, time frames, proactive and filtering measures, sanctions, and reporting duties. For example, initiatives to combat hate speech tend to mandate or encourage the removal of content within 24 hours whilst an earlier draft of the Regulation preventing the dissemination of terrorist content online required removal of terrorist content within an hour. Similarly, proactive filtering obligations feature, albeit indirectly, in the Copyright Directive but were removed from the latest version of the Terrorist Content Regulation.

In the wake of Germany's Network Enforcement Act, also known as NetzDG, France's

National Assembly adopted its own piece of legislation that would require online intermediaries to delete content that is "manifestly unlawful on grounds of race, religion, gender, sexual orientation or disability" within 24 hours of notice. The Senate is now poised to examine the draft Avia law, whose scope has been significantly extended to other types of user-generated content, such as apology of acts constituting an offence against human dignity, war crimes, crimes against humanity, slavery, crimes of collaboration with an enemy, voluntary interference with life or physical integrity, sexual aggression and harassment, aggravated theft, extortion or destruction, voluntary degradation or deterioration which is dangerous to a person, etc.

The French draft law against online hate speech thus adds yet another layer of complexity to an already confusing patchwork of national and European rules imposed on hosting providers that operate and offer services throughout the European Union. In our view, this patchy regulatory framework risks jeopardising the freedom to conduct a business. Furthermore, it contributes to a further fragmentation of the Digital Single Market.

The French proposal is also problematic given that the European Commission plans to put forward its own proposal for a Digital Services Act (DSA) that many assume will include a reform of the intermediary liability rules in the 2000 E-Commerce Directive. European Commission Head of Unit for "E-Commerce and Online Platforms" Werner Stengg has confirmed at the Mozilla Mornings event on the future of EU content regulation on 10 September 2019[12] that a core objective of the DSA will be to harmonise the rules governing moderation of illegal user-generated content in order to avoid diverging national obligations that may involve substantial investments and operating costs for service providers. Indeed, different and diverging legal regimes increase compliance costs while also being a source of legal uncertainty and unpredictability regarding the qualification of content as illegal and the scope of responsibilities and obligations of service providers. The new rules would also help prevent competition distortions and remove obstacles to the free movement of information society services.

It is likely that the DSA will take the legal form of a regulation. It would therefore apply automatically and uniformly to all EU countries as soon as it enters into force, without the need for it to be transposed into national law. Because the matters covered by the scope of the draft Avia law are intrinsically connected to the upcoming reform, there is a significant risk that the future European provisions could conflict with existing national ones.

The European Commission should recommend the withdrawal of the French draft Avia law in order to prevent the introduction of national measures that would compromise the adoption of the future Digital Services Act by the European Parliament and the Council in the same field.

12 https://blog.mozilla.org/netpolicy/2019/08/22/mozilla-mornings_content-regulation/

Further reading

ARTICLE 19, France: Analysis of draft hate speech bill
https://www.article19.org/resources/france-analysis-of-draft-hate-speech-bill/

La Quadrature du Net, PPL Avia, refusez les mesures inutiles et dangereuses
https://www.laquadrature.net/wp-content/uploads/sites/8/2019/06/lqdn_analyse_ppl_avia_27_juin.pdf

La Quadrature du Net, L'Assemblée national adopte et aggrave la loi "haine"
https://www.laquadrature.net/2019/07/09/lassemblee-nationale-adopte-et-aggrave-la-loi-haine/

EDRi and Access Now, Content regulation – what's the (online) harm?
https://edri.org/content-regulation-whats-the-online-harm/