



**THE CITIZENS**



**CENTRE  
FOR PEACE  
STUDIES**

**global witness**



**eticas**



**AI FORENSICS**



**jesuislà**

**Digital Action**



**Irish Council for  
Civil Liberties**

## **Joint-Submission on the Commission's Guidelines for Providers of Very Large Online Platforms and Very Large Online Search Engines on the Mitigation of Systemic Risks for Electoral Processes**

7th March 2024

This response was prepared by more than a dozen civil society organisations working across Europe. Together we commend the Commission for its efforts to protect elections.

We would like to draw attention to an area critical to achieving this aim: recommender system safety.

**This response is presented in three parts:**

- **Part 1 introduces how recommender systems contribute to systemic risks**
- **Part 2 responds to the Commission's proposals to moderate virality of content that threatens the integrity of the electoral process**
- **Part 3 introduces our proposals to moderate virality of content that threatens the integrity of the electoral process by strengthening recommender system safety**

## Part 1 - How recommender systems contribute to systemic risks

As Commissioner Breton remarked, *“the question of algorithmic amplification and the ranking of content based on interactions [...] is precisely at the heart of the Digital Services Act.”*<sup>1</sup>

Algorithmic recommender systems - currently in use - select emotive and extreme content and show it to people who they estimate are most likely to engage with it. These people then spend longer on the platform, which allows Big Tech corporations to sell ad space. Not only is this inherently prone to amplify harmful content but it also incentivises its production.

Earlier this month, Gartner published a ranking of the top risks identified by over 300 risk management leaders, risk professionals, auditors, and senior executives. Gartner ranked “Escalating Political Polarization” driven by “reinforcing social media algorithms”, number 2 in its top 5 risks, stating, “Algorithms that are designed to ensure engagement and retention on social media platforms by delivering tailored content to individual users, also lead to information silos that feed confirmation bias and can reinforce divisive political outlooks.”

Facebook’s own internal research found that Facebook’s recommender system was driving political recommendations to extremes: even if a person followed only verified conservative news, they were soon recommended extreme conspiracy content.<sup>2</sup>

Separate internal Facebook research concluded “64% of all extremist group joins are due to our recommendation tools... Our recommendation systems grow the problem”.<sup>3</sup>

Therefore, it is clear that behavioural profiling-based recommender systems are one of the major areas where platform design decisions contribute to “systemic risks”, as defined in Article 34 of the DSA, especially when it comes to “any actual or foreseeable negative effects” for the exercise of civic discourse and electoral processes.

Moreover, the necessity of measures tackling profiling-based recommender systems is all the clearer in view of the providers’ continued breaches of fundamental principles of EU law in how they operate their recommender systems. Recommender systems that engage with, and therefore process, sensitive data, such as users’ political views, sexual orientation, religion, ethnicity, or health information, may contravene the General Data Protection Regulation<sup>4</sup>(GDPR). Processing ‘special category’ data, in the absence of explicit consent, is prohibited under Article 9(1) of the GDPR, apart from under certain circumstances which are unlikely to apply in the case of mass data harvesting.

---

<sup>1</sup> Verbatim report of proceedings: The fight against hate speech and disinformation: responsibility of social platforms within the Digital Services Act (topical debate), 7 February 2024

[https://www.europarl.europa.eu/doceo/document/CRE-9-2024-02-07-ITM-012\\_EN.html](https://www.europarl.europa.eu/doceo/document/CRE-9-2024-02-07-ITM-012_EN.html)

<sup>2</sup> “Carol’s journey to QAnon”, Facebook internal research, 2019, cited in “Inside Facebook, Jan. 6 violence fueled anger, regret over missed warning signs”, Washington Post, 22 October 2021

<https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/>

<sup>3</sup> “Facebook Executives Shut Down Efforts to Make the Site Less Divisive”, Wall St. Journal, 26 May 2020

<https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499>). This internal research in 2016 was confirmed again in 2019.

<sup>4</sup> Regulation (EU) 2016/679, General Data Protection Regulation, <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

## **Part 2 - The Commission's proposed measures to moderate virality of content that threatens the integrity of the electoral process**

These guidelines include some useful measures to reduce the spread and algorithmic amplification of harmful content, which we welcome (please see below for details and suggestions for enhancement), however, these do not tackle the problem at a structural level.

We agree that platforms should introduce positive friction and circuit breakers aimed at slowing down posting and user interactions, giving users a chance to think before sharing. This includes 'think before you share' messages and limits on resharing.

Furthermore, we support the application of labels providing contextual information, for example, provided by independent fact checkers, and agree that the spread and algorithmic amplification of such identified content should be downranked. However, given the sheer scale of the problem, now exacerbated by the widespread accessibility of Gen AI, relying on this approach leaves safety and integrity teams and fact-checkers in a perpetual state of "catch-up".

The limitations of both human and algorithmic content moderation, prove why it is imperative that platforms redesign their systems away from models that inherently favour emotive and extreme content, and thereby, inadvertently content that threatens the integrity of the electoral process. The Center for Humane Technology made this call in their letter<sup>5</sup> to Mark Zuckerberg, following the release of the Facebook Files by whistleblower Frances Haugen in Haugen in October 2021<sup>6</sup>:

“No matter how many fact-checkers you hire, how much you invest in AI, how you tweak metrics like Meaningful Social Interaction (MSI), or how hard your Oversight Board works, Facebook and democracy will be incompatible until the underlying operating model changes.”

## **Part 3 - Our proposals to moderate virality of content that threatens the integrity of the electoral process by strengthening recommender system safety**

### **1) Profiling off by default**

As an immediate first step, we propose that platforms make their default recommender systems the versions that are not based on personal behavioural profiling. Switching defaults so that a person is now given the choice whether they wish to switch profiling-based recommender systems on rather than off is an elegant and restrained measure to address the acute harms created and amplified by such recommender systems.

---

<sup>5</sup> “To: Mark Zuckerberg”, Center For Humane Technology online, accessed February 19, 2024, <https://www.humanetech.com/oneclicksafer-letter>

<sup>6</sup> “The Facebook files” The Wall Street Journal online, accessed February 19, 2024 <https://www.wsj.com/articles/the-facebook-files-11631713039>

## **Practicality of this measure**

Providers that have diligently brought their systems into compliance with existing legal requirements will already be effortlessly able to implement the measures. Specifically, we refer to three existing legal requirements:

i) First, it is a well-established principle of EU Law that providers must carefully control, monitor, and account for their use of “special categories” of personal data, distinct from other personal data. Therefore, providers are required to have already implemented the necessary distinctions in how their systems handle different types of data. The measures add no new technical requirement.

ii) Second, providers are subject to several further legal requirements before they can commence any “profiling” activities. They must have also conducted a Data Protection Impact Assessment;<sup>7</sup> have established a lawful basis for the specific purposes for which they intend to conduct profiling;<sup>8</sup> be able to discontinue the profiling when requested to do so by a person being profiled;<sup>9</sup> and be able to delete the data concerned where necessary, too.<sup>10</sup> Thus, providers must under existing law already have created the necessary systems to switch off profiling. Again, the measures add no new technical requirement.

iii) Third, Article 38 of the Digital Services Act provides that recommender systems based on profiling must be optional with users given an alternative choice. Therefore, providers also have a separate and pre-existing requirement to be able to implement the measures. The sole difference is that the new measures operate as the default. This makes no practical difference to the technical burden on providers.

Providers should be able to implement the measures immediately, without any technical difficulty. Only providers who have previously failed to take the necessary steps under existing law will find the measures challenging. Any such difficulties will derive solely from the provider’s own unlawful conduct, rather than from the measures themselves.

### **Fairness of this measure:**

The Commission rightly notes that, “During electoral periods, it is of particular importance that providers of VLOPs and VLOSEs show that content moderation decisions do not affect the equality of candidates or disproportionately favour or promote voices representing certain (polarised) views.” Therefore, a mitigation measure that is content or narrative-blind such as disabling profiling-based recommender systems which equally affects everyone, does not risk leaving certain voices feeling disproportionately impacted.

## **2) Optimising for values other than engagement**

VLOPs should depart from signals and metrics that correlate with user engagement (especially short term engagement) and prioritise signals/features that correlate with (subjective) relevance and (objective) credibility of the recommended content. This includes: prioritising the signals provided by explicit user feedback and

---

<sup>7</sup> Article 35(3)(a), GDPR

<sup>8</sup> Article 5(1)(b) and Article 7, GDPR

<sup>9</sup> Article 21 and Article 22, GDPR

<sup>10</sup> Article 17, GDPR

preferences, bridging signals (e.g. the diversity of the users who engaged with a given piece of content and positive explicit feedback coming from users that are very different from one another), and signals that correlate with legitimacy, credibility and transparency of the source, especially when it comes to recommendations and search returns on sensitive topics.

Very Large Online Platforms (VLOPs) have shown how they can minimise the weight of the engagement factor and make their systems less harmful, for example, when Meta implemented "break glass" measures ahead of the US 2020 elections.

## Conclusion

Achieving safety by design is at the heart of tackling content that threatens the integrity of the electoral process. By disabling profiling-based recommender systems by default and optimising for values other than engagement, VLOPs can take significant steps toward mitigating the systemic risks that their recommender systems pose to election integrity. However, these must be done in tandem with other measures to ensure safety by default whilst maintaining a positive user experience.

Over the long-term, we urge the Commission to foster the development of rights-respecting alternative recommender systems, including **prompting conscious user choice, including opening up content curation to third party services, and removing predatory and addictive design features.**

For more detail on these measures, see [Safe by Default: Moving away from engagement-based rankings towards safe, rights-respecting, and human centric recommender systems.](#)

Signed,

AI Forensics  
Centre for Peace Studies from Croatia  
Defend Democracy  
Digital Action  
Ekō  
Eticas Foundation  
Foundation The London Story  
Global Witness  
Irish Council of Civil Liberties  
#jesuislà  
Real Facebook Oversight Board  
The Citizens  
Panoptykon Foundation  
People vs Big Tech  
Politiscope