Joint submission on the draft Online Safety Code

Submitted to Coimisiún na Meán by more than sixty civil society organisations

31 January 2024

In this submission

1. This submission was prepared by more than sixty civil society organisations. Together, we represent a diverse cross-section of Irish society. Our submission is focused on two measures ("**the measures**") for recommender system safety, excerpted here:

> "…that recommender algorithms based on profiling are turned off by default;
> …that algorithms that engage explicitly or implicitly with special category data such as political views, sexuality, religion, ethnicity or health should have these aspects turned off by default;"[1]

2. This submission is presented in five parts:

- part 1 highlights widespread support for the measures;
- part 2 discusses the necessity, proportionality, and practicality of the measures;
- part 3 proposes strengthening the measures, without which they cannot be effective;
- part 4 discusses further measures for recommender system safety; and
- part 5 proposes enhancements for effective and efficient enforcement.

We also append a brief observation on age verification.

RECOMMENDER SYSTEMS

Part 1: support for the measures

3. Our organisations together join in commending Coimisiún na Meán for introducing the measures. If strengthened, they are an elegant means of providing the protections required by Section 139K(2) of the Online Safety and Media Regulation Act 2022, without intruding upon freedom of expression.

4. A national poll conducted by Ireland Thinks in January 2024 shows overwhelming popular support across all ages, education, income, and regions of the country for the measures: across Ireland 82% are in favour. We enclose further findings from this poll in Appendix 1.

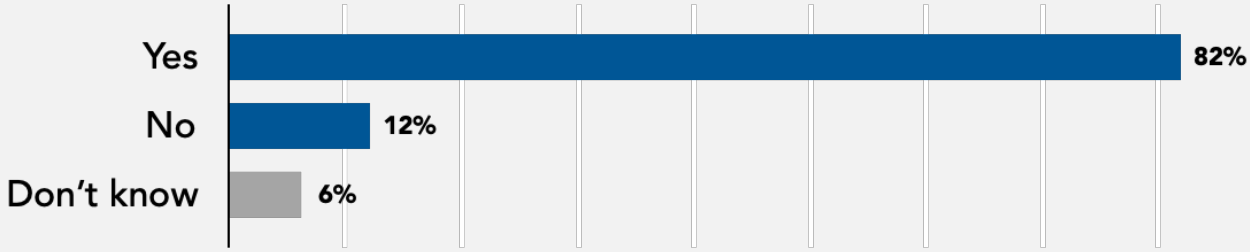
5. International reaction to the measures is also overwhelmingly positive. A cross-party group of Members of the European Parliament has formally written to the European Commission, urging it to learn from Coimisiún na Meán's example and to apply the measures across the Union under Article 35 of the Digital Services Act.[2]

6. The measures are also praised by United States Federal Trade Commissioner Alvaro Bedoya,[3] tech thought leader Cory Doctorow,[4] and famed Silicon Valley investor, Roger McNamee, together with pioneer of US Democratic Party digital campaigning, Professor Zephyr Teachout, who co-authored an opinion piece in *The Hill* about the measures:

> Coimisiún na Meán's bold move would ultimately make the Digital Services Act far more successful. Europe and the Irish government are stepping up at last to regulate harmful technology products. Social media may become social again.[5]

Very widespread support for the measures

Question: "Would you be in favour of social media companies being forced to stop building up specific data about you (your sexual desires, political and religious views, health conditions and or ethnicity) and using that data to pick what videos are shown to you (unless you have asked them to do this)?"



National poll conducted by Ireland Thinks. See detailed results in Appendix 1 of this submission.

Part 2: necessity, proportionality, and practicality of the measures

7. The measures are necessary and proportionate to the objective set by the Online Safety and Media Regulation Act 2022. Section 139K(2) of that Act requires that the Code protect children against harmful content. This includes (by reference to Article 28b of the Audiovisual Media Services Directive) that children must be protected against communications that may impair their physical, moral, or mental development. Section 139K also requires that the Code protect the general public from communications that incite violence or hatred (with reference to Article 21 of the Charter), and against provocation to criminal offenses including terrorism, racism, and xenophobia.

8. Providers' content recommender systems are known to create these harms. For example:

- 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.[6]

- Separate internal Facebook research concluded "64% of all extremist group joins are due to our recommendation tools… Our recommendation systems grow the problem".[7]

- Nearly three quarters of the problematic[8] content seen by 37,000+ test volunteers on YouTube was due to YouTube's recommender system amplifying it.[9]

- In August 2023, an Anti-Defamation League study found that Facebook, Instagram, and X recommended antisemitic and conspiracy content to 14-year-old test users.[10]

- The European Commission reports that Russian disinformation about Ukraine was achieved by pro-Kremlin actors and "algorithmic recommendation by the platforms".[11]

- U.N. investigators found that Meta played a "determining role" in Myanmar's 2017 genocide.[12] Amnesty International reported Meta's algorithms were key contributors.[13]

- Less than one hour after Amnesty created a TikTok account posing as a 13-year-old child interested in mental health content, videos encouraging suicide were recommended.[14]

- Researchers at the Institute for Strategic Dialogue found that YouTube's "shorts" video system routinely recommends extremely hateful misogynistic material to young boys.[15]

- The following two stories were shared by Uplift members:

    - "My beautiful, intelligent, accomplished niece was encouraged, incited to see suicide as a romantic way to end her life. She did end it. Earlier she had been encouraged to see more and more sites by people who espoused the idea that people suffering from mental health issues should stop their medications and force society to accept them as they were. This led her a dangerous downturn from which she never recovered, leaving her poor parents devastated and her family changed for the worse."

    - "My father has slowly been radicalised by the content pushed to his feed on Facebook. He watches the short videos and accepts all the information in the video without any verification on his part. If you ask him to verify it, he calls you a liar. The videos can directly state conflicting information, but he will accept it all as fact without thinking about it. This is fuelling his anti immigration thoughts and ideas. I fear he'll become homophobic too."

9. These harms are acute.

10. The Act requires that measures in the Code must be proportionate to the level of risk of exposure to the content and harms.[16] 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. Indeed, the question is whether the measures go far enough: should recommender systems that are based on special category personal data and profiling be prohibited entirely? The measures are the minimum intervention that Coimisiún na Meán can take, in view of the harms under consideration and the requirements of the Act.

11. There is an unarguable requirement for Coimisiún na Meán to implement the measures. Coimisiún na Meán can make no assumptions that the measures would be introduced if it did not act itself, despite the measures being necessary. There is no prospect that the providers will introduce the measures of their own volition. Indeed, they have already signalled to Coimisiún na Meán that they object to any provisions for the safety of their recommender systems being introduced in the Code.[17]

12. The necessity of the measures 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 a user's politics, sexuality, religion, ethnicity, or health necessarily process "special category" data, implicitly or explicitly. They continue to process special category data for their recommender systems at enormous scale despite the fundamental prohibition of any such processing of special categories of personal data established in Article 9(1) of the GDPR, in the absence of two-step explicit consent. Nor have the providers attempted to seek and confirm the giving of two-step explicit consent. We do not suggest that Coimisiún na Meán should enforce data protection law, but rather that providers have proven themselves unwilling to act even when required by law.

13. The Commissioner rightly notes the "move from an era of self-regulation to one of effective regulation".[18] Providers have a very poor record of self-improvement and responsible behaviour, even when lives are at stake as in Myanmar's genocide. As previous experience has shown,[19] even when a provider understands the harm its recommender system causes, it is unlikely to voluntarily act. Most recently, a senior Meta engineer, Arturo Béjar, reported sending Meta's top executives internal reporting that over 22% of surveyed 13–15-year-olds were bullied, and 13% had received unwanted sexual advances in just the previous week.[20] Despite this, no action was taken. Indeed, systems he had addressed to tackle these issues had been neglected since he left the company.

Practicality of the measures

14. Providers that have diligently brought their systems into compliance with existing legal requirements will already be effortlessly able to implement the measures. We highlight 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.[21] 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" actives. They must have also conducted a Data Protection Impact Assessment;[22] have established a lawful basis for the specific purposes for which they intend to conduct profiling;[23] be able to discontinue the profiling when requested to do so by a person being profiled;[24] and be able to delete the data concerned where necessary, too.[25] 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 a profile must be optional. 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 envisaged by Coimisiún na Meán operate as the default. This makes no practical difference to the technical burden on providers.

15. 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.

Part 3: strengthening the measures, without which they cannot be effective

16. The word "ensuring" in Section 139K(3) of the Act requires that the Code must be effective in achieving the objectives. Coimisiún na Meán also operates under a general principle of effectiveness, provided in Section 7(1) of the Act. We propose three modifications to ensure the measures are effective.

17. First, the measures on recommender systems in Section 1.3 of Appendix 3 should be relocated to Section 12 of the Code, where obligations upon providers are specified.

18. Second, the language should be amended to clarify that the measures are strict requirements.

i) The words "the choices that have been made about whether and" should be struck from the relevant paragraph on page 28, at section 6.4 of the Code, as follows:

> "Coimisiún na Meán therefore considers it appropriate that supplementary measures to the Code should require VSPS providers to prepare, publish and implement a recommender system safety plan that includes effective measures to mitigate the main risks and, at a minimum, explains ~~**the choices that have been made about whether and**~~ how they have implemented a number of specified measures."

ii) The words "consider the following measures and" and "whether and" should be removed from the text on page 77, at section 1.3 of Appendix 3, as follows:

> "In preparing a recommender system safety plan, a video-sharing platform service provider must at a minimum ~~**consider the following measures and**~~ explain ~~**whether and**~~ how it has given effect to them: [...]"

iii) The words "should have these aspects" should be replaced by "must be" on page 78, at section 1.3 of Appendix 3, in order to remove ambiguity and allow for efficiency of monitoring and enforcement. The amended text:

> "measures to ensure that algorithms that engage explicitly or implicitly with special category data such as political views, sexuality, religion, ethnicity or health ~~**should have these aspects**~~ **must be** turned off by default; and"

19. Third, providers are bound by EU law to request and confirm two-step "explicit consent" before commencing any processing of special category data.[26] However, to our knowledge, this consent has been neither sought nor obtained for the relevant recommender systems of the designated providers. The Code should specify that providers must introduce lawful consent requests and confirmation requests.

Part 4: further measures for recommender system safety

20. We highlight three further matters. **First**, the Code does not explicitly refer to digital addiction. We anticipate that Coimisiún na Meán will wish to examine addiction in detail, and establish further measures, too. This is a particular problem for children. We suggest the
21. that following minimum measures be added to the Code: notifications should be off by default, no infinite scroll, and no auto playing the next video.

22. **Second**, we applaud four further measures in Appendix 3.[27]

> …video-sharing platform service providers shall prepare, publish and implement a recommender system safety plan that includes effective measures to mitigate risks that their

recommender systems may cause harm by:

- exposing users to relevant content which, in aggregate, causes harm;
- amplifying relevant content which is harmful to children or to the general public;

…

- measures to ensure that a feed of content is not dominated by one type of content and contains a minimum amount of content that would be viewed positively by users;
- measures to allow a user to reset any profiling algorithm so that it functions as if the user was a new user;

23. **Third**, we suggest the Code should oblige providers to change the signals that their recommender systems use to rank content and measure performance. Instead of prioritising signals that place an overriding emphasis on engagement, which has proven disastrous in consequences, they should instead opt for signals that show the quality of content, such as providence and authorship, and whether the creator is well-regarded by other well-regarded creators. This is a practical measure: there are well established frameworks by which quality of content can be estimated in an automated way.[28]

Part 5: effective and efficient enforcement

24. We suggest three enhancements to ensure effective and efficient enforcement of the Code. First, procedures arising from complaints should involve all relevant parties. Section 14.7-8 provide that the provider will have the opportunity to make submissions. However, the provider is not the only party that should be heard. Section 139U of the Act requires Coimisiún na Meán to have regard for the rights of relevant persons involved in a complaint. Where complainants (per Chapter 4 of the Act) and other parties are involved they should have the opportunity to make submissions. The role of the parties and the procedure by which they are heard in the procedural "Scheme" developed pursuant to Section 139V(1) of the Act should observe the requirements of quasi-judicial bodies that administer justice, and be informed by the *Zalewski* decision of the Supreme Court.

25. Second, we suggest that the Code elaborate particulars of the "content limitation notice". Aside from a reference in Section 14.15 there is no further reference in the Code or supplementary measures. We suggest the relevant provisions in 139ZZD of the Act be articulated in Section 14, to inform the parties and the public.

26. Third, when Coimisiún na Meán deliberates over whether to issue an information notice it must, per Section 139ZZD(3) of the Act, consider the technical capacity of the provider to act on that notice. We strongly caution that expert opinion that is entirely independent of the provider should be obtained to do so. Otherwise, providers may evade their responsibilities by claiming spurious technical difficulties.

OTHER MATTERS

Note on age verification

27. Section 11 of the Code requires various "effective measures to detect under-age users". The guidance provided on pages 67-68 of Coimisiún na Meán's draft lists five purported measures to age verification that are presumably deemed to be effective. The listed measures are taken verbatim from the UK ICO Children's Code.[29] None are viable.

28. Self-declaration, a listed measure, objectively fails Coimisiún na Meán's test of effectiveness. The others are either unspecified or unworkable. Recent developments in Australian legislation,[30] and the reports of the French data protection authority[31] and of UK Ofcom,[32] all indicate that "age verification" measures are unreliable, circumventable, and legally fraught because of their disproportionate effects. Therefore, we urge utmost caution in accepting age verification measures proposed by providers. Furthermore, in the absence of effective and legally permissible age verification, Coimisiún na Meán may be obliged to apply the protections of Audio Visual Media Services Directive (AVMSD) Article 6a and Article 28b(1)(a) to all persons of unproven age.

Error in the Draft Code

29. We note that Section 4.10 of the Code incorrectly indicates that Article 6a of the AVMSD applies solely to commercial communications. This is inaccurate. The relevant point of Article 6a is not limited to commercial communication. Section 4.10 of the Code should be corrected.

Signed

Irish Council for Civil Liberties
Hope & Courage Collective
Uplift
People vs Big Tech
Community Work Ireland
Galway City Community Network
Cork Rebels for Peace
Irish Network Against Racism
Afri
Doras
Action for Choice
Social Rights Ireland
Helping Irish Hosts

Empower
Outhouse LGBTQ+ Centre
ShoutOut
Leitrim Volunteer Centre
European Anti-Poverty Network Ireland
Human Rights Sentinel
Donegal Intercultural Platform
Inishowen Together
Black and Irish
Dublin City Community Cooperative
Bridging The Gap Ireland
Bray for Love
Irish Traveller Movement
Clare Immigrant Support Centre
Mammies for Trans Rights
Together for Safety
Droichead FRC
Age Action
LGBT Ireland
Migrant Rights Centre Ireland
IDEN, Irish Doughnut Economics Network
Dublin LGBTQ+ Pride
National Women's Council
Irish Council for International Students
New Horizon Refugee Support
Pavee Point Traveller and Roma Centre
Belong To - LGBTQ+ Youth Ireland
Solas Project
National Traveller Womens Forum
Waterford Integration Services
Nasc, the migrant and refugee rights centre
Fermoy and Mallow Against Division
Women for Election
Circle VHA
Climate Action Wexford
International Community Dynamics CLG
Dublin Bay South Branch Social Democrats
Wicklow Volunteer Centre
Light Advisory
Women's Collective Ireland (WCI)
Good Day Cork
Parable Communications
Suas/STAND

Rialto Youth Project
Independent Living Movement Ireland (ILMI)
The Exchange Inishowen
NeuroPride Ireland
Friends of the Earth Ireland

Notes

[1] Section 1.3 of Appendix 3
[2] Big Tech's divisive 'personalization' attracts fresh call for profiling-based content feeds to be off by default in EU, TechCrunch, 20 December 2023 (URL: https://techcrunch.com/2023/12/20/dsa-recommender-systems/).
[3] "…an elegant proposal…". Alvaro Bedoya on Twitter (URL: https://x.com/BedoyaFTC/status/1744450499791695938?s=20).
[4] "…a fantastic regulatory proposal for recommendation systems…". Cory Doctorow (URL: https://doctorow.medium.com/https-pluralistic-net-2023-12-09-gallimaufry-marty-hench-rides-again-154871ffe4dc).
[5] "The EU should support Ireland's bold move to regulate Big Tech", The Hill, 31 December 2023 (URL: https://thehill.com/opinion/technology/4380369-the-eu-should-support-irelands-bold-move-to-regulate-big-tech/).
[6] "Carols 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 (URL: https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/).
[7] "Facebook Executives Shut Down Efforts to Make the Site Less Divisive", Wall St. Journal, 26 May 2020 (URL: 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.
[8] "YouTube Regrets: A crowdsourced investigation into YouTube's recommendation algorithm", Mozilla, July 2021 (URL: https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf), pp 9-13.
[9] ibid. p. 17.
[10] "From Bad To Worse: Amplification and Auto-Generation of Hate", ADL, 16 August 2023 (URL: https://www.adl.org/resources/report/bad-worse-amplification-and-auto-generation-hate)
[11] "Digital Services Act: Application of the Risk Management Framework to Russian disinformation campaigns", European Commission, 30 August 2023 (URL: https://op.europa.eu/en/publication-detail/-/publication/c1d645d0-42f5-11ee-a8b8-01aa75ed71a1/language-en), p. 64.
[12] U.N. investigators cite Facebook role in Myanmar crisis, Reuters, 12 March 2018 (URL: https://www.reuters.com/article/us-myanmar-rohingya-facebook/u-n-investigators-cite-facebook-role-in-myanmar-crisis-idUSKCN1GO2PN).
[13] "The social atrocity: Meta and the right to remedy for the Rohingya", Amnesty International, 2022 (URL: https://www.amnesty.org/en/documents/ASA16/5933/2022/en/), pp. 45-48, p. 71.
[14] https://www.amnesty.org/en/latest/news/2023/11/tiktok-risks-pushing-children-towards-harmful-content/.
[15] Algorithms as a weapon against women, Institute for Strategic Dialogue, April 2022 (URL: https://www.isdglobal.org/wp-content/uploads/2022/04/Algorithms-as-a-weapon-against-women-ISD-RESET.pdf).
[16] Requirements specified at sections 7(2), 139K(3), 139L, and 139M of the Act.
[17] "Consultation Document: Online Safety", Coimisiún na Meán, 8 December 2023, p. 28.
[18] "Consultation Document: Online Safety", Coimisiún na Meán, 8 December 2023, p. 5.
[19] Despite internal concern about amplifying hazardous content, from 2017 to 2020 Meta strongly amplified[19] posts that received "emoji" reactions from other people. Then, despite internal research in 2019 confirming that content receiving "angry emojis" was more likely to be misinformation, it persisted until late 2020.[19] "Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation", *Washington Post*, 26 October 2021 (URL: https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/).
[20] Written Testimony of Arturo Bejar, U.S. Senate Subcommittee on Privacy, Technology, and the Law, 7 November 2023 (URL: https://www.judiciary.senate.gov/imo/media/doc/2023-11-07_-_testimony_-_bejar.pdf).
[21] Article 9, GDPR.
[22] Article 35(3)(a), GDPR.
[23] Article 5(1)b and Article 7, GDPR.
[24] Article 21 and Article 22, GDPR.
[25] Article 17, GDPR.
[26] "Guidelines 05/2020 on consent under Regulation 2016/679", European Data Protection Board, 4 May 2020 (URL: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf), pp. 20-22.
[27]

[28] For example, the Trust Indicators, The Trust Project, 2020 (URL: https://thetrustproject.org/wp-content/uploads/2020/07/7.29.20The-Trust-Indicators-Handout.pdf).

[29] https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/how-to-use-our-guidance-for-standard-one-best-interests-of-the-child/best-interests-framework/age-assurance/

[30] In August 2023 the Australian Parliament concluded that it could not lawfully legislate for the age verification requested by the Australian e-Safety Commissioner. See https://www.theguardian.com/australia-news/2023/aug/31/roadmap-for-age-verification-online-pornographic-material-adult-websites-australia-law.

[31] CNIL, the French data protection authority, reported in 2022 that age verification is "circumventable and intrusive".https://www.cnil.fr/en/online-age-verification-balancing-privacy-and-protection-minors

[32] Ofcom's 2022 study of online user ages demonstrates the difficulty of achieving certainty of a person's age online. https://www.ofcom.org.uk/__data/assets/pdf_file/0015/245004/children-user-ages-chart-pack.pdf