UCD Centre for Digital Policy

Response to the Call for Evidence regarding Guidelines under Article 28 of the DSA: Ensuring a high Level of Privacy, Safety and Security for Minors

Authored by Dr Ioanna Noula and Dr Tijana Milosevic.

Contact: Ioanna Noula, ioannanoula@gmail.com

About the UCD Centre for Digital Policy

The UCD Centre for Digital Policy engages forward thinking research about the regulation of digital technologies that underpin all areas of society. The members of the UCD Centre for Digital Policy believe that policy making and evaluation must be deliberative, emergent, and iterative, with sociocultural values at their core. Such an ambitious agenda requires working with stakeholders and beneficiaries to develop effective and evidence-based formal and informal regulation and institutional digital policies, maintain such policies over time, and foreground urgent issues of sustainability, equity, and human rights. The members of the centre draw on interdisciplinary methods from computing, law, design, human rights, and social science to create policy, amplify positive effects on society (especially vulnerable citizens, who may include women, people of colour, the poor, migrants, children, and others), and study policymaking across technologies and sectors.

Overview

DSA's implementation coincides with important momentum built over the past five years around addressing online harms globally, advancing children's rights in the digital world while treading the last stretch of the path toward attaining the UN's Sustainable Development Goals by 2030.

Beyond the EU, online safety regulators and global organisations advancing human rights and fair economic development also advocate for children's access to redress mechanisms. The most prominent regulatory frameworks include UK's Online Safety Act and Australia's Online Safety Act. Children's redress is also emphasised by the Council of Europe through its Guidelines on Child-Friendly Justice and the United Nations Convention on the Rights of the Child with its most recent General Comment 27 focusing on Children's Access to Justice and Remedies whose draft was published in February 2024[1]. Although relevant policy guidelines and regulatory measures aim to ensure children's fair treatment and appropriate remedies in their online

---

[1] Draft General Comment No. 27 On Children's Rights To Access To Justice And Effective Remedies: https://www.ohchr.org/en/documents/general-comments-and-recommendations/draft-general-comment-no-27-childrens-rights-access

and offline environments, it is unclear how their implementation will match policy makers' ambition and what solutions will look like in a global, interconnected, diverse and everchanging online environment.

Our contribution builds on our most recent empirical research and conclusions from our review of the global policy landscape regarding the protection of minors online and considers the importance of children's rights to participation and freedom of expression as drivers for ensuring the highest levels of Privacy, Safety and Security.

Approach

In accordance with the priorities of:

- the UN Convention on the Rights of the Child (UNCRC) and Article 12 of the Convention on children's right "**to express their views, feelings and wishes in all matters affecting them, and to have their views considered and taken seriously**",
- the EU Strategy for the Rights of the Child and its emphasis on the role of children as agents of change, and the promotion of a model of rights-based participation, and
- the European Strategy for a Better Internet for Kids (BIK) and its priority pillar 3 on children's "**active participation, respecting children** by giving them a say in the digital environment",

and considering:

- UNCRC's General Comment 13 on the **right of the child to freedom from all forms of violence**,
- UNCRC's General Comment 25 on **children's rights in relation to their digital environment**,
- UNCRC's concept note for the drafting of General Comment 27 on children's **right to access to justice and fair remedies**,
- the Guidelines of the Committee of Ministers of the Council of Europe on **Child-friendly Justice**,
- The UN Digital Compact and its objective for **global co-operations that "must be forward-looking and capable of identifying, anticipating, assessing, monitoring and adapting to emerging technologies** so that we can seize opportunities and respond to new and emerging risks and challenges",

we argue that, in addition to the focus on children's right to privacy, safety and security, **the guidelines should give due weight to children's right to participation and freedom of expression**.

Recommendations

a) On young users' online redress.

In our response we consider the emphasis given by the Digital Services Act on the issue of user redress. The DSA includes nine articles intended to facilitate user redress and support and expand platforms' obligations to treat users fairly. These articles account for the full continuum of the redress process from the occurrence and reporting of harm to the reparation of those whose rights have been infringed upon. The following articles pertain to the process of redress:

- Art.12 on platforms' obligation to offer points of contact for users,
- Art.16 on notice and action mechanisms regarding moderation decisions,
- Art.17 on providing statement of reasons for moderation decisions,
- Art. 20 on providing internal complaints handling systems for processing users' appeals to moderation decisions,
- Art. 21 on users' right to seek redress using out-of-court dispute settlement services (ODS) paid for by platforms,
- Art. 22 on Trusted Flaggers (TFs),
- Art. 23 on measures and protection against misuse of reporting systems,
- Art. 25 on platforms' obligation to provide consistent design of reporting interfaces, and
- Art. 48 on the development of platform crisis protocols to respond to extraordinary circumstances affecting public security or public health.

Per recital 89 "*[p]roviders of very large online platforms and of very large online search engines […] should ensure that their services are organised in a way that allows minors to access easily mechanisms provided for in this Regulation, where applicable, including notice and action and complaint mechanisms*".

What the DSA is identifying from the outset in recital 89, which focuses on the needs of minors, is **the importance of building *coherence* around the process of redress.** The proposed coherence is important for building consistency of outcomes and child-user satisfaction from the process thereby enhancing the credibility of redress mechanisms, the confidence of users in their efficiency and encouraging their uptake and meaningful usage. **However, the legislative text lacks sufficient detail on how the proposed coherence can be built.** In fact, **provisions for new actors** and entities intended to support the process of redress **including Trusted Flaggers (Art. 22) and Out of Court Dispute Settlement bodies (Art. 21) add significant complexity to the task of achieving coherence and the harmonisation** of user experience across Member States.

To advance the coherence and efficiency of redress mechanisms for minors, the guidelines should account for the following:

1. A definition of redress that captures the meaning, purpose, stakeholders and elements of the process. A comprehensive and comprehensible definition will serve as a reference point for:

   a. **stakeholders responsible for delivering redress** (platforms, TFs, ODS bodies, regulators, judiciary) offering **clarity about their responsibilities** and the kind of information they can have access to, and supporting the development of benchmarks for their performance,

   b. **beneficiaries of redress mechanisms** including minors and their parents/guardians providing **clarity in terms of what they can expect from the process and how they can exercise their rights** and

   c. the **harmonisation of experiences and fair outcomes** for users across Member States.

2. Young people's right to appropriate remedies. **An important element of redress is the restitution of affected young users including remedies for moral injury** such as the acknowledgement of mistakes and system failures, offering adequate support and material restitution where appropriate i.e. when users have incurred undue financial loss. Our consultation with youth in the Irish and Greek contexts has shown that **young people enduring or witnessing unlawful and/or harmful practices related to the 5 Cs are primarily interested in the establishment of order in their communities in accordance with the Terms of Service of the platform rather than receiving financial or other material remedies** (such as access to premium features).

3. Clarity about the capabilities and responsibilities of new actors supporting redress. There is **insufficient detail** about the capabilities, resources and the role TFs and ODS bodies will play, how their independence will be ensured and how their efficiency will be measured. This vagueness harbours the **risk of fragmentation of redress and undermines the rights of children in Member States** where these provisions are not adequately implemented.

4. Systematic and harmonised monitoring and evaluation mechanisms to assess the efficiency of redress mechanisms. Research in other sectors shows that keeping track of and analysing complaints and reviews about services is an efficient way of

assessing and managing risk, identifying emerging harms and informing safety by design[2].

b) On content moderation systems and algorithmic design.

content moderation systems

Platforms are increasingly relying on Artificial Intelligence (AI) for content moderation practices. Proactive content moderation (leveraging AI to monitor content and behaviours for safety purposes even before content or behaviours are reported as potentially violative) is widely used, as reported in large companies' Transparency Reports. Nonetheless, there is insufficient transparency as to how AI models used for detection purposes are developed[3]; limited ability of independent researchers to assess the effectiveness of the models[4]; and insufficient information as to whether and how children and young people are involved in the process of designing AI-based moderation. **Our recent research in Ireland indicated that young people are largely not aware of AI-based content moderation and that they have concerns around privacy and freedom of expression implications of applying AI proactively for content moderation.**[5] We therefore find it important that an **effort be made on behalf of the companies to meaningfully involve children and young people in these processes and to include them in the evaluation process in a transparent manner.**

the design of any algorithmic system

Exposure of young people to online harms related to algorithmic curation of content has also been documented, and the same argument as per above regarding **involving children into the process of** designing AI-based interventions, applies to **algorithmic curation of content** as well[6].

c) On the implementation of the guidelines.

The implementation phase of the DSA creates an opportunity to turn stakeholders' intentions into transformative outcomes that bring together "the expected" (what civil society stakeholders expect from online platforms), "the required" (what regulation

---

[2] Deloitte. (2020). Unlocking the value of complaints: Improving the complaints management process as a strategic source for insights. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/regulatory/us-unlocking-the-value-of-complaints.pdf.

[3] Milosevic, T., Van Royen, K., & Davis, B. (2022). Artificial intelligence to address cyberbullying, harassment and abuse: new directions in the midst of complexity. International journal of bullying prevention, 4(1), 1-5.

[4] Verma, K., Milosevic, T., Davis, B., & Norman, J. O. H. (2023). DESIGNING ETHICAL ARTIFICIAL INTELLIGENCE (AI) SYSTEMS WITH MEANINGFUL YOUTH PARTICIPATION: IMPLICATIONS AND CONSIDERATIONS. AoIR Selected Papers of Internet Research; Verma, K., Davis, B., & Milosevic, T. (2022). Examining the Effectiveness of Artificial Intelligence-Based Cyberbullying Moderation on Online Platforms: Transparency Implications. AoIR Selected Papers of Internet Research.

[5] Milosevic, T., Verma, K., Carter, M., Vigil, S., Laffan, D., Davis, B., & O'Higgins Norman, J. (2023). Effectiveness of Artificial Intelligence–Based Cyberbullying Interventions From Youth Perspective. Social Media+ Society, 9(1), 20563051221147325.

[6] https://antibullyingcentre.ie/wp-content/uploads/2024/04/DCU-Recommending-Toxicity-Summary-Report.pdf.

requires of companies) and "the feasible" (what online platforms can do). Evidence suggests that **productive and effective regulation** should not be authoritative or drive a culture of deterrence from not complying[7]. Instead, it **is built on the cooperation of stakeholders, the identification of shared purpose and desired outcomes, the creation of incentives, and continuous learning from data and feedback**[8].

We propose that the use of regulatory sandboxes already encouraged in privacy and AI-related EU regulation[9] [10] **is going to be particularly advantageous for a rights-based, participatory implementation of the DSA including Article 28** and the development of solutions that will advance children's best interests in the digital world. Typically, Regulatory Sandboxes such as those deployed in the financial sector bring together stakeholders in a "safe space" to test innovative products and services without incurring regulatory sanctions and they are mainly used in the finance sector to test new services (e.g. digital wallets)[11].

We propose that **Regulatory Sandboxes on the Protection of Minors Online could serve the purpose of the effective and impactful implementation of the guidelines** in a collaborative space where rights-based solutions can be developed **with the active participation of youth and the non-adversarial engagement between online platforms and regulators**. An inclusive approach to the development of these Sandboxes creates **opportunities for the participation of "hard to reach" vulnerable youth** i.e. young people living in precarity, facing poverty, and victims of sexual and other abuse and exploitation.

Through **vulnerable young people's participation in the proposed Sandboxes**, which can be facilitated by organisations focusing on their care, these particularly vulnerable young people can have a new opportunity to have a say in the shaping of solutions that will minimise the risk of them being further victimised online. Sandboxes **can, therefore, result in more inclusive and high-impact approaches to safety by design**.

---

[7] Hodges, C. J. S., & Steinholtz, R. N. (2017). Ethical business practice and regulation: A behavioural and values-based approach to compliance and enforcement. Hart Publishing.

[8] Hodges, C. (2022). Outcome-based cooperation: In communities, business, regulation, and dispute resolution. Hart Publishing, an imprint of Bloomsbury Publishing.

[9] i.e. the Personal Data Sandbox by the French National Commission on Informatics and Liberty (CNIL), https://www.cnil.fr/en/sandbox-cnil-launches-call-projects-artificial-intelligence-public-services

[10] European Parliamentary Research Service, June, 2022). Artificial intelligence act and regulatory sandboxes (https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf.

[11] Goo, J. J., & Heo, J.-Y. (2020). The Impact of the Regulatory Sandbox on the Fintech Industry, with a Discussion on the Relation between Regulatory Sandboxes and Open Innovation. Journal of Open Innovation: Technology, Market, and Complexity, 6(2), 43. https://tinyurl.com/4errfpr5.