Balancing the scales. General Comment No. 27 and child online redress: The COR Sandbox approach

Dr Ioanna Noula

1. General Comment No. 27 as an opportunity to rethink digital safety by design

At its 101st session, which closed on 28 September 2026, the UN Committee on the Rights of the Child adopted General Comment No. 27 (GC27) on "children's rights to access to justice and to an effective remedy" (United Nations Office at Geneva, 2026). GC27 follows General Comment No. 25 (GC25) adopted in 2021 (UN Committee on the Rights of the Child, 2021), which set out how the UN Convention on the Rights of the Child (UNCRC) applies in the digital environment. GC27 reminds us that there is a follow-up question: what happens when children's rights are violated? Putting it in the language of GC25: what happens when children's rights are violated in digital environments? How can justice be delivered and what do fair remedies look like? The text of GC27 has not yet been published, but the announcement is a timely opportunity to reflect on what effective and fair redress for children looks like in the digital era. We argue that any answer to these questions must pass through Article 12 of the UNCRC, on children's right to be heard on matters that affect them (UN General Assembly, 1989). Our starting point is that with regard to child online redress, the scales are tipped towards adult perceptions of fairness based on adults' definitions of harm, fair remedy and the process that leads to it.

Digital regulation increasingly requires platforms to provide redress for children, but it does not adequately answer the question of what fair and effective redress should look like from the perspective of the children who use it. The COR Sandbox (COR) aims to address this gap by creating a participatory regulatory environment in which children, regulators and service providers co-design, test and evaluate redress mechanisms children's participation, and in which the weight given to their views can be verified. COR is a pioneering proof-of-concept project for youth-centred digital safety sandboxes that responds to the community's calls for safety by design. Focusing on the concept of child online redress COR brings the concept of regulatory sandboxes, already tested in the financial and AI services sectors, to the digital safety space, with a mission to operationalise safety by design and advance children's best interests in digital environments.

2. The state of child redress in digital environments. When everyone's problem becomes nobody's responsibility

The boundary between children's online and offline lives has all but dissolved. Harms that begin offline are amplified online, and harms that occur online have real consequences for children's physical and mental wellbeing. Meanwhile, technology evolves at a pace that remains overwhelming for those responsible for safeguarding human rights and societal wellbeing. Regulation, institutions and professional practice are trying to keep up, navigating inflexible bureaucracy, policy-making processes that are not fit for purpose and sectoral silos which delay the sharing of knowledge and synthesis of intelligence that would lead to efficient and timely responses.

The lines of responsibility for harm have blurred with it. The number of actors who could support a child, or be held responsible for harm, keeps growing. The EU Digital Services Act (DSA) (European Parliament & Council of the European Union, 2022) is a compelling case of the complexity of the redress workflow for online harms. In addition to traditional routes to responding to and redressing online harm including platforms' reporting tools, a child's family and school environment, helplines and social services, the DSA mandates internal complaint-handling mechanisms by digital services providers where children can appeal moderation decisions, trusted flaggers who expertly report illegal content, out-of-court dispute settlement bodies, and direct reporting to regulators. These actors are all part of the redress ecosystem, and all have a role in supporting children's access to justice and fair remedies. Often, their roles interface, their responsibilities overlap, the liability regime is, at best, opaque, at worst absent, and the remedies unspoken. And with the increase in the number of redress stakeholders comes the increase of the risk that, as Floridi puts it, "everybody's problem becomes nobody's responsibility" (2016: 11). In a fragmented multi-agent redress ecosystem, responsibility becomes harder to locate. Users, including children, can find it hard to understand where responsibility lies, what remedy they can seek, and how information about harm moves between the actors involved.

Floridi further argues that this diffusion of responsibility "is morally unacceptable and pragmatically too risky" (2016: 11). Likewise, redressing harm is both a moral imperative and a risk management asset. Redress is a way of reestablishing the moral order by acknowledging the act of harm, recognising and appreciating the inherent worth of the injured party and addressing their claim for moral and/or material remedies. The health of redress mechanisms is an indicator of systemic accountability, as their role is to rebalance power asymmetries and uphold rights as enshrined in normative and legal frameworks. The moral imperative of redress is more urgent for children who operate in digital environments that are not built with their needs and rights in mind. In the adult-oriented architectures of

digital environments, children are “vulnerabilised” while being deprived of the informational capital that would allow them to navigate redress mechanisms and assert their rights.

From a pragmatic perspective, effective redress becomes a source of systemic stability and capacity building. The responsiveness and effectiveness of redress mechanisms are a form of assurance for the harmed parties, and society more broadly, that governance systems are fair and trustworthy. Functioning redress mechanisms sustain the legitimacy of the socio-political order by communicating that societal and political governance is underpinned by the values of care and justice. The deterrent value of remedial and restorative dimensions of redress encompassing sanctions and corrective measures drives a collective sense of security. Furthermore, the reporting and evaluation processes of harm bear important diagnostic value that allows systems to detect existing or emerging governance flaws that can cause dysfunction and disruption.

In digital environments, children’s complaints constitute data points signalling emerging risks. Digital services providers’ and regulators’ early engagement with such signals expedites access to children’s lived experiences and makes harms visible. At the same time, the responsiveness of redress mechanisms drives a positive cycle of their use by children who feel their voice matters and that they are cared for. Opportunities to participate in the determination of remedies and the design of redress processes, in line with Article 12 of the UNCRC, have the potential to enhance how children perceive the legitimacy and trustworthiness of the governance of digital environments.

3. Child Online Redress: a right out of reach?

But what happens when redress is negotiated entirely by adults? In August, Meta agreed to pay about $18 billion to resolve a multistate lawsuit claiming that it intentionally designed addictive platforms that harmed young people's mental health (Ortutay et al., 2026). The settlement was hailed as historic and was celebrated as a victory over Meta. There is more to take stock of regarding what this “win” means about children’s right to access to justice and effective remedies. And although the US has not ratified the UNCRC, this settlement is food for thought for any jurisdiction asking what redress for children should look like.

The act of the settlement between the two sides was framed as Meta’s settlement and the way it has been celebrated by adults reveals how far we are from a redress process that treats children as rights-holders (Levine, 2026). The resolution the adult parties reached has little to do with remedying harms incurred by children. Although the rationale behind

the terms of the settlement is that new measures may prevent future harm, the outcome does not remedy the harm children have already suffered. Children and their families are neither compensated financially (states allocate the funds for specific purposes), nor do they benefit from moral reparations as the legal filing expressly maintains that Meta denies all allegations of wrongdoing, deceptive trade practices, or liability, even if some commentators read the settlement as an implicit acknowledgement that its voluntary safety tools fell short (Yamin, 2026). On the contrary, children incur additional consequences by having their access restricted, and their freedom of expression moderated through limits on social comparison features such as 'like' counts.

Considering the reasoning behind the settlement between Meta and the US states involved in the litigation and the wave of social media “bans” announced across borders, there is a rather clear pattern: when children are harmed, imposing restrictions on children is the (quick) fix. And there lies a deeply ingrained culture that shapes the terms of engagement with children in digital environments, and which is the source of regression for children’s rights. Even when it comes to redressing harms, children are viewed as inherently vulnerable and in need of protection, management, discipline and, effectively, what can be perceived as punishment. They are widely perceived as “adults in the making” (James, 2011) and therefore not owed an explanation, an acknowledgement and a say in what is fair to them. Challenging adult opinions and exercising their right to appeal adults’ decisions on what is good for them lies far outside the dominant imaginary of childhood.

Child online redress is no longer optional: the EU's DSA, the UK's Online Safety Act and Australia's Online Safety Act all require or expect platforms to offer complaint mechanisms usable by children, but in practice, online redress remains under-defined and operationally and jurisdictionally fragmented. From a business point of view, redress is often treated as a compliance cost and a reminder of customer dissatisfaction making it difficult to make a business case around it. Compounded with the “deficit mindset” which runs through much of how adults approach children (James & Prout, 1997) and describes adults’ perception of children as “citizens-in-waiting" rather than citizens in their own right (Osler & Starkey, 2005), child online redress does not invite either the attention of regulators or the interest of service providers. This culture pushes Article 12 of the UNCRC further away from the decision-making equation when it comes to what fair and effective child online redress mechanisms look like.

4. Sandboxing child online redress and safety by co-design

Children, as both active and vulnerable users of digital services, occupy a distinctive position within this evolving landscape. Research has compellingly evidenced important negative effects of existing governance models on their rights and wellbeing (van der Hof et al., 2020; Kirkorian, 2025; Opree et al., 2016). While efforts to regulate digital services to advance their best interests have made important progress, policy makers seem to be missing the mark. New regulatory frameworks are criticised for their limited ability to address the root causes of harm and emerging risks (Yu et al., 2025), their ineffectiveness in challenging the profit-driven, market-focused outlook of innovation (Calabrese & Virah-Sawmy, 2025) and, most recently, their failure to account for children's full suite of rights (Third, 2025; Wood, 2024). The compounded effect of these shortcomings is driving a vicious cycle of disempowerment and harm for children. In this context, we approach redress mechanisms as a disruptor of the current state of affairs, with the potential to serve as a conduit for children's voice, an amplifier of their rights, an intelligence instrument for existing and emerging harms and systemic risks, and a beacon for the navigation of shapeshifting risks. A streamlined, accessible reporting process that encourages children to report, combined with the systematic documentation of reports, can reveal trends around reported harmful content, contact or conduct. By creating accessible and positive reporting experiences, children and young people are more likely to use reporting tools. A privacy-preserving reporting process can become a forum where children and young people not only report harmful content and experiences after harm manifests, but also feel empowered to engage with the service provider to share concerns about early signs of harm and provide context about their experiences. Each step in the redress process can become a space for dialogue where young people can claim their rights through reporting, appealing decisions, seeking independent review of platforms' decisions, and having access to remedies.

For platforms and regulators, redress can also become a resource-efficient process to access the experiences of users that are hard to reach, such as socio-economically disadvantaged groups or users with mental health issues, allowing timely interventions and meaningful support when harm occurs, and improving response readiness at a systemic level by allocating resources in an informed manner.

In our response to the Committee's consultation on the draft GC27, we recommended that the General Comment encourage national and international regulatory sandboxes as a tool for the implementation of child online redress. The law already mandates redress for young people who experience harm online, without however making clear how this will look in practice and how it will catch up with new harms which result from the rapid evolution of technology. What the speed of change and unpredictability of technology make clear is that in order to regulate innovation, we must innovate regulation. New regulatory

instruments are needed with transformative capabilities (Attrey et al., 2020; Molnar, 2026). In rapidly changing sectors, markets and disciplines transformed by private actors' capitalisation of data points, and intelligence building, regulators acting in the public interest need access to fresh and authentic insights and competence to transform knowledge into measures fit for purpose.

The advent of AI and the associated risks of its unchecked adoption have led to a steady increase in sandboxes across sectors, as they appeal to regulators who initiate them to minimise possible negative impacts on people and institutions and businesses keen to ensure compliance and avoid reputational risks of technological backlash (Qiu et al., 2025). Sandboxes offer institutionalised quasi-experimental spaces that aim at reducing regulatory uncertainty for companies, give regulators access to observe if new regulatory regimes are fit for purpose and foster knowledge exchange that allows regulators to understand technological capabilities and companies to comprehend the rationale behind regulatory regimes and how these might impact their products and wider market (OECD,2025; Shaik, 2026).

Technological developments and the ensuing need for improved adaptive governance models addressing challenges emerging from the fast deployment of AI have resulted in the rapid evolution of the scope of sandboxing and the adoption of the methodology beyond strictly regulatory and compliance-related needs. Emerging typologies propose the categorisation of sandboxing into operational sandboxes focusing on technical issues such as interoperability or data flows, regulatory sandboxes characterised by the testing of new technology under regulatory supervision, hybrid sandboxes as spaces of mutually beneficial testing of technical and legal innovation or policy sandboxes focusing on prototyping rules and governance models prior to their formal adoption (Guio et al., 2025). As it becomes clear from this typology, sandboxes are initiated by and involve two stakeholders, i.e. regulators and regulated service providers.

Sandboxes involving or initiated by civil society like the one initiated by the Institute for the Future of Work (n.d.) are strict exceptions meaning that end-users of technologies or policy beneficiaries are not involved in the design, proceedings and exit processes of a sandbox. However, as sandboxes become a mainstream regulatory methodology in the context of the development of Digital Public Infrastructure (DPI) like digital identity and payment services the involvement of end-users/citizens whose daily lives are dependent on these infrastructures becomes both a moral imperative and an operational dependency with experts calling for the active involvement of citizens in the sandboxing process (Porciuncula et al., 2025). The Datasphere Initiative (2026) highlights the importance of the participation of civil society in DPI experimentation processes such as sandboxes, through

usability testing, co-design, feedback loops, and governance discussions, which allows developers to surface adoption barriers, concerns, and unintended impacts early, before systems are scaled.

5. Balancing the scales: COR Sandbox and youth-centred child online redress

The question that the COR Sandbox sets out to answer is how fair and effective redress processes can honour Article 12 of the UNCRC and how child participation can be achieved across the workflow of redressing harm. We argue that designing products and policies, including mechanisms for child online redress, with children's rights in mind does not go far enough, and it is not consistent with children's right to be heard in matters that affect them per Article 12's requirements. By treating redress as a systemic function linking harm detection, reporting and remedy (Bennett & Guio, 2026), COR aims to translate children's right to access to justice and to an effective remedy into the design of the mechanisms through which harm is prevented and children are empowered in digital environments.

We propose that sandboxes offer a transformative environment for the design of rights-respecting digital safety characterised by the active participation of children and their meaningful and verifiable input in the design process. To meet their full potential and ensure consistency with Article 12, sandbox environments should be co-designed with children to become a space for youth participation and safety by co-design. Co-designed participatory sandboxes can offer to participating regulators and businesses access to authentic insights about what happens on the ground in relation to new types and sources of harm, how these are experienced by children and the processes that might work to identify it and address it.

COR is prototyping the regulatory methodology of sandboxing in the field of digital safety. It puts Article 12 of the UNCRC into action through the active participation of youth across the lifecycle of the Sandbox spanning the design of the Sandbox environment to the Sandbox exit and re-planning of future sandbox iterations. The geographical scope of the sandbox spans the jurisdictions of Greece, Ireland and Australia stress-testing the potential of regulatory harmonisation by taking a principles-based approach to safety by design rooted in the UNCRC. The strategic focus of its operations is child online redress, which is approached as a lever for agile regulatory responses and a key element of self-improving organisational governance. The unique character of COR lies in the fact that it is a civil society-led sandbox that puts in focus the experiences and needs of the user. It is

being supported by helplines that hold the ground-level intelligence on systemic failure and can support the development of a common lexicon that bridges youth experience and adult tech policy. COR is developing a collaborative environment where regulators, civil society, industry and young people work together around co-defined outcomes and shared objectives. This approach draws on the state-of-the-art model of Outcome-Based Collaborative Regulation (OBCR) which "focuses on what we are trying to do – objectives, goals, purposes- but also on whether we are succeeding by achieving the right outcomes" (Hodges, 2026: 112).

Recent analyses of the sandboxing landscape recognise COR as an emerging model that departs from conventional sandboxes. Rather than facilitating market entry or regulatory exemptions, COR tests governance approaches, participation models and safeguards and responds to governance fragmentation in the field of digital safety and the exclusion of children from policy and service design processes (Bennett & Guio, 2026) and it treats children not only as subjects of protection but as knowledgeable users with distinct insight into how redress mechanisms work in practice (Datasphere Initiative, 2026).

GC27 invites us to rethink justice for children and new "digital" laws already mandate child online redress. Considering the breakthrough that the UNCRC and Article 12 were for the field of children's rights almost 40 years ago and in line with the historical turn of the Convention through GC25, the COR Sandbox proposes that child redress that is fit for purpose in digital environments should be designed with the children it is meant to serve. It is creating a rights-centred infrastructure with the potential to transform the currently tedious and ineffective process of reporting harm inside or outside platform environments into a proactive way of learning about real-life impact and capability to mitigate emerging risks as well as create new routes to redress and fair remedies, turning redress into a critical infrastructure for safety by co-design. It aspires to balance the scales, giving children a vote on what justice and remedies mean in the digital world.

References

Attrey, A., Lesher, M., & Lomax, C. (2020). The role of sandboxes in promoting flexibility and innovation in the digital age (Going Digital Toolkit Notes No. 2). https://doi.org/10.1787/cdf5ed45-en

Bennett, M., & Guio, A. (2026). A Sandbox Playbook: From design to implementation for Digital Public Infrastructure. Policy Innovation Lab, Stellenbosch University; Global

Network of Internet & Society Centers. https://policyinnovationlab.sun.ac.za/wp-content/uploads/2026/07/Sandbox-Playbook-FA3-digital.pdf

Calabrese, S., & Virah-Sawmy, R. (2025, March 26). Big Tech Is Avoiding Responsibility – Here Is What the EU Can Do About It. European Democracy Hub. https://europeandemocracyhub.epd.eu/big-tech-is-avoiding-responsibility/

Datasphere Initiative. (2026). *Sandboxes for DPI. Co-creating the blocks of digital trust*. https://www.thedatasphere.org/wp-content/uploads/2026/03/Report-Sandboxes-for-DPI-2026_v6.pdf

European Parliament & Council of the European Union. (2022, October 27). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng

Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), 20160112. https://doi.org/10.1098/rsta.2016.0112

Guio, A., Molnar, A., Siewert, M., & Gasser, U. (2025, July). AI Sandboxes: Global Insights for Regulatory Learning and Adaptive Governance. TUM Think Tank. https://tumthinktank.de/wp-content/uploads/PolicyBrief_Final.pdf

Hodges, C. (2026). Regulation That Works Transforming the Future of Regulatory Spaces. Bloomsbury Publishing Plc.

Hof, S. van der, Lievens, E., Milkaite, I., Verdoodt, V., Hannema, T., & Liefaard, T. (2020). The Child's Right to Protection against Economic Exploitation in the Digital World. The International Journal of Children's Rights, 28(4), 833–859. https://doi.org/10.1163/15718182-28040003

Institute for the Future of Work. (n.d.). Responsible AI Sandbox. Retrieved https://www.ifow.org/landing-page/sandbox

James, A. (2011). To Be (Come) or Not to Be (Come): Understanding Children's Citizenship. The ANNALS of the American Academy of Political and Social Science, 633(1), 167–179. https://doi.org/10.1177/0002716210383642

James, A., & Prout, A. (1997). Constructing and Reconstructing Childhood: Contemporary Issues in the Sociological Study of Childhood (2nd ed). Taylor and Francis.

Kirkorian, H. (2025). Introduction to the Section on Digital Media, Cognition, and Brain Development. In D. A. Christakis & L. Hale (Eds.), Handbook of Children and Screens: Digital Media, Development, and Well-Being from Birth Through Adolescence (pp. 7–11). Springer Nature Switzerland [Imprint].

Lansdown, G. (2022). Article 12: The right to be heard. In Z. Vaghri, J. Zermatten, G. Lansdown, & R. Ruggiero (Eds.), *Monitoring State Compliance with the UN Convention on the Rights of the Child: An Analysis of Attributes* (Vol. 25, pp. 41–50). Springer International Publishing. https://doi.org/10.1007/978-3-030-84647-3

Levine, I. (2026, September 4). Without a child rights framework, Meta's $17.1bn settlement will only go so far. Tech Policy Press. https://www.techpolicy.press/without-a-child-rights-framework-meta-settlement-will-only-go-so-far/

Molnar, A. (2026, May 11). Governing digital harm without borders. TUM Think Tank. https://tumthinktank.de/en/news/governing-digital-harm-without-borders/

Opree, S. J., Buijzen, M., & Van Reijmersdal, E. A. (2016). The impact of advertising on children's psychological wellbeing and life satisfaction. European Journal of Marketing, 50(11), 1975–1992. https://doi.org/10.1108/EJM-06-2015-0393

Ortutay, B., Chan, K., & Huamani, K. (2026, August 26). Meta reaches $18 billion settlement with states in landmark trial over teen social media addiction. AP News. https://apnews.com/article/meta-trial-instagram-settlement-97d342f2a33d835eda2356c5e1af9e37

OECD. (2025). Regulatory sandbox toolkit: A comprehensive guide for regulators to establish and manage regulatory sandboxes effectively. OECD Publishing. https://doi.org/10.1787/de36fa62-en

Osler, A., & Starkey, H. (2005). Changing Citizenship. Democracy and Inclusion in Education. Open University Press.

Porciuncula, L., Onyango, R., Amutorine, M., & Rozo-Paz, M. (2025, April 25). *Moving fast together: How sandboxes can help build trust in Digital Public Infrastructure*. https://www.thedatasphere.org/news/moving-fast-together-how-sandboxes-can-help-build-trust-in-digital-public-infrastructure/

Qiu, Y., Yao, H., Ren, P., Tian, X., & You, M. (2025). Regulatory sandbox expansion: Exploring the leap from fintech to medical artificial intelligence. Intelligent Oncology, 1(2), 120–127. https://doi.org/10.1016/j.intonc.2025.03.001

Shaik, A. S. (2026). *Sandboxes, Safe Harbors, and Sunset Clauses: Designing Adaptive Regulation for AI That Doesn't Become Obsolete on Arrival*. Zenodo. https://doi.org/10.5281/ZENODO.20893580

Third, A. (2025). On board the Australian social media 'ban wagon': Regulatory theatre, the public child, and the hyper-enthusiastic State. In C. Nelson, D. Buiten, & J. Death (Eds.), The Public Child (pp. 25–52). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-97267-6_2

UN Committee on the Rights of the Child. (2021). General comment No. 25 (2021) on children's rights in relation to the digital environment. https://undocs.org/en/CRC/C/GC/25

UN General Assembly. (1989). Convention on the Rights of the Child. Office of the United Nations High Commissioner for Human Rights. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

United Nations Office at Geneva. (2026, September 28). Committee on the Rights of the Child Closes One Hundred and First Session after Adopting Concluding Observations on Reports of Kazakhstan, Lesotho, Slovenia, Tajikistan, Timor-Leste, Uganda and Tanzania. https://www.ungeneva.org/en/news-media/meeting-summary/2026/09/le-comite-des-droits-de-lenfant-acheve-les-travaux-de-sa-101e

Wood, S. (2024). Impact of regulation on children's digital lives. Digital Futures for Children centre. https://researchonline.lse.ac.uk/id/eprint/123522/

Yamin, S. (2026, September 9). *Meta's $18 billion settlement: Reckoning or managed exit?* https://www.mishcon.com/news/metas-18-billion-settlement-reckoning-or-managed-exit

Yu, Y., Liu, Y., Zhang, J., Qian, Z., Yun, H., & Yang, W. (2025). Understanding Generative AI Risks for Youth: An Empirical Taxonomy for Safer Digital Futures. Proceedings of the Association for Information Science and Technology, 62(1), 1737–1739. https://doi.org/10.1002/pra2.1526