Child Online Safety Compliance

Praesidio Safeguarding

Autumn 2023

16 minute read time

Praesidio
safeguarding

Contents Page

Who we are

Praesidio is a consultancy firm specialising in child online safety. We are true subject matter experts, recognised globally as leaders in our field. We work with stakeholders across the regulatory and policy landscape to enhance children and young people's safety and wellbeing in the digital environment. Our clients include tech companies, governments, regulators, international governmental organisations and NGOs.

Central to our approach is our multidisciplinary hub of digital safety and wellbeing experts. The hub includes child psychiatrists, behavioural scientists, academics, technologists, researchers and public policy experts. Our experts have experience working in major tech companies, regulatory authorities, NGOs, law enforcement and law firms. Through our hub, we convene bespoke teams to meet the specific requirements of each project. Our leaders and experts are supported by our in-house team of project, policy and compliance managers.

We believe every child has the right to be safe and to thrive in the digital environment. We work with those who share these values and who have a genuine commitment to protecting and supporting young people online, including complying with regulation.

Praesidio is a child digital safety and wellbeing consultancy. We are subject matter experts not legal professionals. We do not give legal advice and this paper should not be construed as such.

We are independent and not affiliated with any commercial provider of online safety tech such as content moderation tools or age assurance products.

Incoming regulation: age-appropriate design

In Europe and worldwide, digital services are subject to rapidly increasing regulatory oversight. Legislation such as the EU's Digital Services Act (DSA), Ireland's Online Safety and Media Regulation Act 2022 and the UK's Online Safety Act 2023 create significant new responsibilities, especially in relation to child users.

Decisions about which services to regulate and the standards they must meet reflect local and regional priorities and, inevitably, there are differences in approaches and emphasis. However, regulation reflects the consensus that (a) children merit a higher standard of protection than adults, and (b) a child's access to services, content, features and functionalities should take account of their age and stage of development.

Some online safety legislation also requires regulated services to consider the children's rights and best interests as set out in the UN Convention on the Rights of the Child and General Comment 25.

Areas of compliance

Below, we offer our perspective on the key themes and areas of focus across new and emerging child online safety standards. Our analysis is not based on any single piece of legislation. Regulated services should take legal advice from local legal experts on their regulatory obligations.

Child sexual exploitation and abuse (CSEA)

Preventing and detecting CSEA is, rightly, a priority compliance obligation for regulators. Obligations go beyond CSAM and include grooming, re-victimisation and reporting to law enforcement and designated reporting lines. Anticipating risk created by the design of services is also a key concern.

Our expertise includes those with prior careers at the UK's National Crime Agency (NCA) managing complex inquiries involving technology assisted child sexual exploitation. As a result, we have a strong understanding of how sexual offenders exploit tech platforms to identify and target children. We can support organisations to understand technology assisted child sexual exploitation and abuse including how platform design can contribute to and exacerbate harmful sexual interactions between perpetrators and children.

Evidence shows how important design choices are in stopping children encountering those who want to exploit and harm. For example, design decisions such as age-appropriate default privacy settings, restrictions on recommender systems, and age limits on high-risk features such as livestreaming make children less visible to potential predators. Active moderation strategies using languages and behaviour models, and reporting systems that are easy to find and straightforward to use are also essential.

Regulation in Europe has brought significant changes to reporting obligations for regulated services with a stronger emphasis on localised reporting to law enforcement and hashing databases. The Praesidio team brings years of experience in working within law enforcement. We understand how to navigate reporting from industry (e.g. what to report and in what format to do this safely).

We have experience in developing reporting triage systems and tools, as well as training staff in how to use these.

We are also expert in prevention education – giving children information and tools to know how to report and how to stay safe. We can develop and provide education resources that are known to be effective and lead to behavioural change amongst children.

An area of CSEA policy that has received significant attention due to its complexity is the proliferation of self-generated images. Debate continues on how to respond including how to categorise and risk assess such images. In partnership with WeProtect Global Alliance, Praesidio recently completed a global study on the drivers of self-generated images with clear recommendations for technology companies on prevention, intervention and education. We are therefore well placed to advise regulated services on best practice solutions.

Harmful and age-inappropriate content

The categories of content that are considered harmful to all children or children in certain age groups are still emerging, but a consensus has formed around some types including for example, bullying or content that promotes or provides instructional detail on suicide, self-harm and disordered eating. For all but the most egregious content, standards of what children should and should not be exposed to are informed by the age and developmental stage of the child. Decisions on age-appropriate content should be evidence based and developed with input from, and validated by, external experts and end users including children and parents.

Our expert hub includes those with expertise on harmful content and classification standards. This, combined with our research capabilities, means we can work with regulated services and their legal advisors to develop age-appropriate content standards across all content categories or for a single theme (for example violence, imitable risky behaviour, polarised public discourse or adult and sexualised content).

Praesidio can also advise on 'cumulative harm' – where individual pieces of content may not be harmful in themselves, but the volume and frequency (dosage) may lead to harm. Cumulative harm is discussed further below in the section on recommender systems and AI.

Features and functionalities

In addition to creating effective detection and reporting systems, regulation places great importance on safety by design - that is to say, anticipating and minimising risk through design. Age-appropriate design is an aspect of safety by design that considers how the design of a service should be adapted for children or children in certain age groups.

Regulated services typically offer a range of features and functionalities. For example, they may offer content recommendation feeds, content creation tools, private and group messaging functions, gaming, e-commerce, user engagement tools (e.g. like, share, comment) and the ability to livestream. These features and functions have different risk profiles which vary depending on whether they are used in conjunction with other features and functions or indeed with other platforms and apps.

Praesidio can advise on potential risks and mitigation strategies for features and functionalities as part of a wider review or as a standalone review of a specific aspect of service design. Mitigation strategies typically include on-platform information and guidance (see on-platform safety information below); changes to default privacy and safety settings; use of moderation and recommendation systems to reduce access to content and accounts and to reduce the visibility of child users; age restrictions for highest risk features and functionalities; and preventing or disrupting user pathways that direct children off platform to potentially less safe services.

Age assurance

Whilst this is an area with a particularly high level of localisation when it comes to regulatory approaches, preventing children from accessing platforms underage is a high priority issue. The need to ensure children who are old enough to join a platform have an age-appropriate experience is also a primary concern and critical aspect of age-appropriate design. Developing an effective and proportionate assurance system is not straightforward. Methods of age assurance include identity verification using hard identifiers, parental authentication, self-verification, and AI estimation (typically biometrics or user behaviour signals). There is also an emerging sector of third-party age assurance providers that offer a range of age assurance methods. Governments and regulators often emphasise the importance of interoperability.

Research consistently shows that parents and teens believe platforms should do more to enforce their age restrictions, but they flag choice, efficacy, fairness, and privacy as factors where they would welcome further reassurance.

Praesidio has a comprehensive understanding of the age assurance industry and the public and policy debate that surrounds this high-priority issue. Critically, we are entirely independent and have no affiliations to any age assurance provider. We can support regulated services and their legal advisors to create an age assurance system that is user-friendly and meets regulatory expectations.

Recommender systems and AI

Whilst a teen may search for a video or creator they have heard about or may connect with friends they know offline, most of the ways they experience regulated services are determined by recommender systems. These systems suggest videos, images, accounts, search terms, hashtags, comments, products, livestreams and filters and effects. The personalisation of user experiences is central to the design of regulated services. It is therefore unsurprising this a priority area for regulators.

Whilst the scale and nature of the obligations differ, regulatory frameworks across Europe require regulated services to assess and mitigate risk of harm from algorithms. For children, this has several aspects including (for example) consideration of whether or not recommender systems could: facilitate inappropriate contact between predators and children; autosuggest content or creators that are harmful or age inappropriate; use sensitive personal data to profile children in a way that conflicts with their fundamental rights or best interests; or serve certain types of content or creators in a volume or frequency that creates a risk of cumulative harm. Consideration may also need to be given as to whether recommender systems have been built and/or tested using data sets that include children's data; whether the system has an inherent bias that perpetuates discriminatory views or behaviours; or whether they may surface a child's account or content in way that means the child has a level of visibility that is inappropriate given their age and developmental stage.

Praesidio can work with regulated services and their legal advisors to understand how recommender systems impact child users and the efficacy of disruption and dispersal safety models. Praesidio's expertise on children's rights means we are also well placed to assist regulated services and auditors in drawing up

assessment frameworks to enable a comprehensive evaluation of the impact of the design and operation of recommender systems on children's fundamental rights and freedoms as set out in the UNCRC and General Comment 25.

Dark patterns (persuasive design)

Perhaps one of the most interesting aspects of the DSA is the inclusion of obligations relating to Dark Patterns which Recital 67 describes as *"practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions. Those practices can be used to persuade the recipients of the service to engage in unwanted behaviours or into undesired decisions which have negative consequences for them."*

Whilst there are precedents for regulatory oversight of persuasive design and children including the UK's Age-Appropriate Design Code, the DSA sets a new standard for expectations on regulated services. The decisive inclusion of this issue reflects the level of concern consistently expressed by policy makers, child safety advocates, teachers, parents, and children themselves.

This is a fascinating area of academic research. There is a highly developed pedagogy which describes ways in which our understanding of human psychology and behaviour can be integrated into digital product design to influence decision-making. The tension between persuasive design and freedom of choice and the heightened vulnerability of adolescents to dark patterns due to their evolving capacities are both well documented. What is less well understood is the impact on young people's mental health and wellbeing.

Praesidio's hub of experts includes experts on persuasive design. We can work with regulated services and their legal advisors to identify and disrupt dark patterns and to develop wellbeing tools and resources grounded in credible academic research. We do not support solutions which enforce unvalidated screentime limits or which facilitate a level of parental oversight that is incompatible with teens' developmental capacity, rights and freedoms.

As with everything we do, our work on persuasive design is grounded in data and we place strong emphasis on co-creation with teens so that the solutions we develop are child led not adult assumed.

Reporting, flagging and appeals mechanisms

The ability to access user-friendly reporting tools is a common regulatory theme. The DSA includes a requirement to create Trusted Flagger systems where users can approach an expert third party (usually an NGO) to escalate an issue on their behalf through a priority reporting channel. Regulated services may also be required to ensure fairness by providing those impacted by moderation decisions with transparent information about why the decision has been taken and how to appeal.

These tools should be built with child users in mind. This means taking a child-centred approach to design so that reporting systems are intuitive and straightforward to use. For larger services, meeting this threshold is likely to require consultation with teens and parents/guardians through qualitative research and external advice from (for example) child psychiatrists, child safety NGOs and prevention and behavioural scientists.

Praesidio can work in partnership to develop such systems using best practice design processes. Our expertise in this area and our experience enables us to work quickly and efficiently to provide digestible and actionable insights.

On-platform safety information

Legislation across Europe reflects a level of frustration with safety strategies that rely too heavily on digital literacy. However, there is increasing acceptance that information and advice that is surfaced to children on platform (for example using 'just in time' notices) can meaningfully enhance safety. On-platform safety information is most effective when it is pro-actively pushed to users, when it is bitesize, user centred, written in accessible language, actionable and aligned with how the child uses the service.

Praesidio has significant expertise in developing evidence-based, on-platform safety education strategies. We work with behavioural scientists to ensure strategies are grounded in the pedagogy of prevention science. We consult with NGOs who have designed and delivered digital literacy projects for years and we place emphasis on local and regional norms so that interventions resonate.

We seize opportunities to deliver safety advice in collaboration with content creators. Most importantly, we work with teens to understand what they need to know, when they need to know it and how they want to receive the information. We think of teens as co-creators and we work with young people to generate ideas, to evaluate prototypes and to provide feedback on preferred solutions.

Our approach gives regulated services a high level of certainty that the time and money investment required to build effective safety information strategies is well spent whilst regulators can be confident that the intervention is credible and reflects best practice.

Transparency

Another key regulatory theme is heightened accountability through transparency. When thinking about transparency, regulated services must keep in mind whether their audience includes child users. If so, they should ensure the information is conveyed in a way that is age appropriate. This might include use of video, audio, images and animation. It may also require platforms to think about how to breakdown and prioritise information to make it easier to digest and understand. This is a potentially complex task especially when transparency obligations increase the amount of information that needs to be provided.

Praesidio can work with regulated services and their legal advisors to ensure information published on platform including Terms of Service and Community Rules is presented in a way that is accessible to children. Typically, this involves consulting with teens, parents and UX designers to develop and test resources.

Risk assessments

Regulatory strategies across Europe place significant emphasis on understanding potential risks and creating services that are safe and, in relation to child users, age appropriate by design. The DSA and OSA require platforms to carry out detailed risk assessments. It seems likely that the Online Safety Codes under Ireland's OSMR will also include consideration of risk.

In addition to input from safety experts working within regulated services, regulators are likely to expect services (especially larger services) to demonstrate that they have validated their approach using data and advice from independent sources. For example, they may ask whether the risk assessment reflects current academic research and, where proportionate, what steps have been taken to consult with experts and service users including children and parents.

Praesidio can collaborate with regulated services and their legal advisors on all aspects of the risk assessment process from ensuring the framework and process meets expectations to evaluating risk and offering expert advice on risk mitigation strategies. We have exceptional research capabilities and have a strong track record in designing and delivering in-depth, participatory research with children, parents and carers and experts in Europe and worldwide. Our in-house capabilities and close relationships with academic researchers across the globe mean we can undertake rapid literature reviews and ensures our advice reflects current thinking.

Audits

Risk assessments carried out by VLOPS or VLOSE under the DSA are subject to annual external audits. Auditors appointed to undertake this work are in uncharted territory. In relation to child-focused aspects of risk assessment, auditors need to have an understanding of: child and adolescent development and its impact on product design (age appropriate design); the impact of technology on children's mental health and wellbeing; dark patterns or persuasive design strategies; trust and safety systems and processes; age assurance mechanisms; and the application of international child rights frameworks to the digital environment.

These areas do not typically form part of auditors' expertise and skillset. The auditors' task is made more complex by the fact that Trust and Safety is quite nascent, approaches differ between platforms and there is an absence of agreed standards and universal metrics that would allow for industry benchmarking. It is additionally complex for child safety because the Commission's work on the EU Code of Conduct on Age-Appropriate Design is ongoing.

As TikTok highlighted in its response to the European Commission's call for evidence on the draft Delegated Act relating to Article 37(7) of the Digital Services Act (DSA) regarding audit (June 2023):

"It will be very difficult during the early years of the DSA for an auditor – who may well have general risk management and audit experience but lack deep experience acquired from spending years handling user content at scale – to make a proper critical assessment of the measures used by the VLOP/VLOSE. It will take time to build this capacity within the audit industry."

In addition to supporting on all aspects of compliance, Praesidio is well placed to advise audit firms as they navigate novel processes and/or to collaborate as part of an audit consortium or as an audit sub-contractor as envisaged by the Commission.

Contact us

Our compliance work is led by:

Helen King, Director and Co-Founder

helen.king@praesidiosafeguarding.co.uk

Dr Zoe Hilton, Director and Co-Founder

zoe.hilton@praesidiosafeguarding.co.uk

Praesidio safeguarding