Homo Digitalis
Protect your rights

CONTRIBUTION TO THE EUROPEAN COMMISSION’S PUBLIC CONSULTATION ON THE DRAFT GUIDELINES ON MEASURES TO ENSURE A HIGH LEVEL OF PRIVACY, SAFETY AND SECURITY FOR MINORS

Friday, 13 June 2025

By the Homo Digitalis team

Stavroula Chousou

Anastasios Arampatzis

Alkmini Gianni

Kallirroi Grammenou

Lamprini Gyftokosta

TABLE OF CONTENTS

Introduction

Homo Digitalis welcomes the Commission's proposed guidelines on measures *to ensure a high level of privacy, safety and security for minors online, pursuant to Article 28(4) of the Digital Services Act (DSA)*[1].

For the last five years, Homo Digitalis had the chance to speak to more than 6000 students of primary and secondary education in more than ten different regions in Greece, from rural areas to big cities. All children, regardless of their parents' financial or educational background, have access to online platforms and the internet, sometimes from the age of seven, using various devices, but mainly tablets, mobile phones and gaming tools.

Based on our discussions with them and their testimonies, Homo Digitalis understands that minors use the online platforms mostly to connect with others, play games, share and find information. The vast majority, regardless of age, admitted to spending a lot of time online, giving fake information when creating accounts or using their parents' accounts, especially when using their parents' old mobile phones.

What is more, a significant number of them came forward with information that confirms the findings of the Commission about the risks minors face online, including exposure to illegal and harmful content, unwanted contact, cyberbullying and extensive use or overuse of online platforms.

This is why Homo Digitalis welcomes the clarification about the scope of the guidelines[2] that will help to avoid grey zones and misinterpretations. In particular, Homo Digitalis agrees with the EC *that a provider of an online platform that simply declares in its terms and conditions that it is not accessible to minors but does not put any effective measure in place to avoid that minors access its service, cannot claim that its online platform falls outside the scope of Article 28(1) of Regulation.*

Similarly, Article 28(1) of the Regulation applies when an online platform provider already processes the personal data of those recipients revealing their age for other purposes, which also reveals that some of those recipients are minors.

In overall, Homo Digitalis strongly supports the EDPB statement[3] that *'Age assurance must respect the full complement of natural persons' fundamental rights and freedoms, and the best interests of the child should be a primary consideration for all parties involved in the process'***.**

With the present analysis Homo Digitalis wishes to contribute further to the Commission's efforts to protect minors online and provide comments on the proposed guidelines. This paper will also serve as an attachment to the survey that Homo Digitalis has also submitted [date].

[1] Regulation (EU)of the European Parliament and of the Council of 19 October on a Single Market For Digital Services (Digital Services Act)

[2] Points 39-126 of the Communication from the Commission guidelines on measures to ensure a high level of privacy, safety and security for minors online pursuant to Article 8(4) of Regulation (EU) 2022/2065

[3] https://www.edpb.europa.eu/our-work-tools/our-documents/other-guidance/statement-12025-age-assurance_en (last accessed 03/06/2025)

Risk review analysis

General comments

Homo Digitalis considers that the risk review section takes a responsible and child-centred approach in line with Article 28 of the Digital Services Act (DSA) and the UN Convention on the Rights of the Child (UNCRC).

Analysis

However, several key concerns emerge that demand serious consideration. The repeated use of "should" instead of "must" throughout the document significantly weakens the enforceability of these critical protections, creating a permissive rather than mandatory framework that may allow platforms to treat child safety measures as optional recommendations rather than binding obligations.

The guidelines' reliance on "appropriate and proportionate" measures presents another significant challenge. While proportionality is undoubtedly important in regulatory frameworks, the current language offers no clear benchmarks or thresholds for evaluating what actually counts as appropriate and proportionate in practice. This ambiguity could lead to inconsistent implementation across platforms, particularly among smaller operators who may lack the resources or expertise to develop robust internal standards.

The treatment of potential conflicts between safety measures and fundamental rights remains notably underdeveloped. Although the section briefly acknowledges that safety measures might impact rights like participation and expression under UNCRC Articles 13, 17, and 31, it fails to provide meaningful guidance on how platforms should balance these competing rights in real-world scenarios. This gap leaves platforms without clear direction when safety measures might inadvertently restrict children's legitimate participation in digital spaces.

The absence of children's voices from the assessment process is by itself a concern. While children rights impact assessments are encouraged, the guidelines do not emphasise the critical importance of child participation in these evaluations. This oversight contradicts UNCRC Article 12, which establishes children's right to be heard in matters that affect them directly. Without meaningful child participation, these assessments risk becoming adult-centric exercises that may miss crucial perspectives on how platform policies actually impact young users.

It should also be noted that the proposed guidelines provide insufficient direction on transparency and accountability mechanisms. The weak recommendation to merely "consider" publishing outcomes of risk reviews falls far short of what effective oversight requires. More troubling still is the complete absence of guidance on how children or their guardians can challenge harmful platform practices or gain understanding of how algorithmic and policy decisions affect them. This lack of redress mechanisms leaves children vulnerable to platform decisions made without meaningful oversight or appeal processes.

Finally, Homo Digitalis highlights that while the section introduces a solid foundation, it lacks a structured taxonomy to assess risks comprehensively. The 5C Typology of Risks must be viewed together with the CENCENELEC risk-based framework. Together they provide essential granularity and life cycle guidance for evaluating and mitigating harms to minors. It also omits child developmental factors and offers limited direction on how platforms should assess proportionality across different age groups. Furthermore, the role of children in shaping the review or providing feedback is not addressed, weakening its alignment with UNCRC Article 12 on participation.

Recommendations

1. **Strengthen language for accountability**: Replace “should” with “must” in key areas, particularly for identifying risks and assessing child rights impacts.
2. **Clarify “appropriate and proportionate”**: Provide examples or a decision framework that helps platforms gauge a measure's proportionality while upholding UNCRC principles.
3. **Require meaningful child participation**: Add a line recommending that children’s views be incorporated into the risk review process, especially via age-appropriate consultations (UNCRC Article 12).
4. **Balance safety with expression**: Provide more robust guidance on mitigating risks without unduly restricting children's rights to expression, freedom of thought, association, and information (Articles 13, 14, 15, 16, 17 of UNCRC).
5. **Mandate transparency and accountability mechanisms**: Require platforms to publish a summary of their risk reviews in accessible language and explain how they mitigate risks without violating children’s rights.
6. **Comply with data protection principles as stated by EDPB** in its statement,including data minimisation (General Data Protection Regulation (GDPR) Art. 5(1)(c)) and UNCRC Article 16 (right to privacy).
7. **Include algorithmic risk evaluation**: Require platforms to evaluate how algorithmic systems (e.g. recommender systems, content ranking) might expose minors to harmful content or exploitation and whether mitigation techniques are in place (e.g. opting out of profiling).
8. **Emphasise intersectionality and vulnerable subgroups**: Encourage platforms to disaggregate risks and consider how they impact vulnerable or marginalized groups of minors, aligning with UNCRC General Comment No. 25 (2021) on children’s rights in the digital environment.
9. **Include data governance and third-party risk review**: Require platforms to map and assess data-sharing practices with third parties that could impact minors' privacy or security, similar to the GDPR obligations.
10. **Establish a risk-based framework** that takes into account, the 5C Typology of risks, the CENELEC framework and how to determine the best interests of the children in the digital environment.

Age Assurance

General comments

Homo Digitalis considers that this section is principled and aligned with current technological realities. It embraces privacy-first verification, balances risk and proportionality, and rejects weak methods like self-declaration.

Analysis

Homo Digitalis advocates for a layered, risk-based implementation of age estimation (AE) and age verification (AV) systems, with mandatory use of privacy-preserving age verification specifically for services restricted to adults aged 18 and above. This approach represents a carefully balanced framework that aligns with established legal principles while addressing the complex challenges of child protection in digital environments.

The foundation for this risk-based approach lies in its alignment with core principles established under the DSA and GDPR, particularly the fundamental concepts of proportionality, necessity, and data minimisation. Age verification becomes essential for high-risk services such as pornography, gambling, and adult-only platforms, where robust protection mechanisms are required to meet obligations under the UNCRC, specifically Article 19 regarding protection from harm and Article 3 concerning the best interests of the child. In contrast, age estimation provides a more appropriate solution for medium-risk environments, where full age verification might constitute an overly intrusive response that could unnecessarily restrict legitimate access.

The European Union's approach to age verification, which focuses on 18+ proof verification, establishes a strong benchmark by effectively balancing privacy protection, user autonomy, and technical robustness. This hybrid model ensures that children are not systematically over-excluded from age-appropriate digital services while simultaneously raising protection standards in contexts where such measures are most critically needed.

Any implementation of age assurance systems must adhere to several fundamental principles that respect both child protection imperatives and broader human rights considerations. Proportionality and necessity must guide all deployment decisions, ensuring that AV and estimation technologies are employed only when genuinely required, with less intrusive alternatives consistently prioritised in the decision-making process. Privacy and data minimisation, as enshrined in GDPR Article 5(1)(c) and Article 25, demand that age assurance mechanisms must never become gateways for identity tracking or comprehensive profiling of users. These systems must operate through anonymous, local, and untraceable architectures that preserve user privacy while achieving their protective objectives.

The effectiveness of any age assurance framework depends critically on accuracy, robustness, and reliability in implementation. Weak or easily circumvented age verification methods, such as simple self-declaration systems, fundamentally undermine the entire protective framework and fail to deliver meaningful safeguards for children. Additionally, principles of non-discrimination and inclusion require that solutions function equally effectively for all children, regardless of disability status, language barriers, nationality, or variations in device access and technological literacy. Transparency and understanding represent equally crucial elements, demanding that children receive clear, age-appropriate information about when age assurance systems are being employed and how their data is, or importantly is not, processed during these interactions.

However, Homo Digitalis wishes to highlight some critical concerns. The risk of age verification systems becoming surveillance tools represents a paramount concern that demands vigilant oversight. While the EU's current approach maintains privacy-preserving characteristics, many third-party AV solutions operating in the market do not adhere to these same standards. This creates substantial risk that age verification could function as a Trojan horse for mass identification and comprehensive user profiling, directly violating UNCRC Article 16, which establishes the fundamental right to privacy for children.

Implementation consistency across platforms presents another significant challenge that current guidelines fail to adequately address. While the guidelines mandate that AV and AE systems must be deployed for high and medium-risk platforms respectively, they provide no clear criteria or systematic scoring mechanisms to define these risk categories. This absence of specific guidance could result in inconsistent and potentially weak enforcement across different digital environments. Therefore, these definitional criteria must be carefully aligned with the standards and methodologies established in comprehensive Risk Review sections of regulatory frameworks.

The absence of meaningful mechanisms for child participation in age verification decisions represents a fundamental oversight that contradicts core principles established in UNCRC Article 12. Despite the central importance of incorporating children's perspectives into decisions that directly affect them, current approaches fail to integrate young people's voices into either the selection or design phases of age assurance systems. This exclusion undermines both the effectiveness and legitimacy of protective measures intended to benefit children themselves.

Finally, even privacy-preserving age verification methods may create significant barriers for under served minors who lack access to smartphones, stable internet connectivity, or government-issued identification documents. Current guidance offers no substantive direction for developing inclusive alternatives that ensure equitable access to age-appropriate digital services regardless of socio-economic circumstances or technological access limitations. Addressing these accessibility concerns is essential for ensuring that protective measures do not inadvertently create new forms of digital exclusion for vulnerable young people who may already face significant barriers to digital participation.

Recommendations

1. **Develop a common risk matrix or scoring framework:** The Commission should provide a standardized risk scoring framework to help platforms classify content, features, and services as low, medium, or high risk — and apply the appropriate age assurance method. Include this risk scoring matrix as an Annex to the guidelines.

2. **Require inclusive fallback mechanisms:** Providers should implement fallback options (e.g. local device checks, parental verification, or educational exemptions) for minors unable to use the primary AV/AE method.

3. **Involve children in the development and evaluation of age assurance:** Platforms and regulators should consult with children of various age groups in the design, selection, and testing of age assurance systems to ensure relevance and usability.

4. **Explicit life cycle management and deactivation of age data**: Platforms must ensure that any age-related credentials, tokens, or metadata are not stored longer than strictly necessary, are subject to automated expiry, and cannot be re-purposed for profiling, cross-site tracking, or behavioural targeting.
5. **Ban on function creep and secondary uses**: Age assurance tools must not be used for identity verification, advertising, or behavioural analysis, and must not be linked to account-based data flows beyond the purpose of verifying age.
6. **Foresee third-party auditing and certification**: Age assurance systems should be subject to mandatory third-party audits and independent certification to ensure compliance with data protection standards, non-discrimination criteria, and technical robustness.This would align with the DSA's broader push for auditable transparency and systemic risk mitigation (Articles 34–35).
7. **Require clear labelling for age-restricted areas**: Platforms should clearly label age-restricted features or content using consistent visual cues and warnings, so children can understand when age assurance applies and why.
8. **Evaluate AV/AE technologies**: Providers should apply ethical and human rights due diligence when evaluating new age assurance technologies, including biometric methods, and prioritise those aligned with the principles of necessity, proportionality, and privacy by design.
9. **Clarify minimum technical standards**: Define minimum accuracy thresholds and performance metrics to ensure platforms are not using ineffective or biased tools.
10. **Prevent re-authentication through correlation attacks**: Require technical safeguards against token reuse, mandate one-time-use credentials or session-based proofs, and enforce no cross-platform correlation policies.
11. **Account for edge cases and adversarial circumvention**: Platforms must implement anti-circumvention checks, such as liveness detection (if facial analysis is used), device.
12. **Define technical redress flows for false positives/negatives**: Provide technical guidance for re-verification mechanisms (e.g. escalation to alternate AV method), and ensure minimal friction for legitimate users to correct errors — without requiring them to disclose more personal data than necessary.
13. **Emphasise low-latency and UX optimisation**: Technical specks should include performance thresholds (e.g. verification in <2 seconds), clear UI standards, and fallback mechanisms for devices with limited performance or connectivity.
14. **Define "state-of-the-art" for AI based age estimation: R**equire tools testing across demographic groups, document model training data sources, and publish bias mitigation strategies. Technical audits should assess not just accuracy but equity of performance.

Registration

General comments

Homo Digitalis considers that the measures outlined in the guidelines demonstrate clear alignment with established children's rights principles and existing legal obligations across multiple dimensions of digital service design and implementation. When service providers explain the necessity of registration processes to minors using clear, accessible, and age-appropriate language, they enhance the fundamental trust relationship between children and digital platforms while simultaneously building confidence among parents and guardians regarding third-party services. This transparency creates a foundation of mutual understanding that facilitates safer access when minors are developmentally ready to engage with particular digital environments.

The commitment to designing registration processes that is easy for users with disabilities to navigate represents a critical component of non-discriminatory practice that minimises risks of unintentional exclusion while empowering all young people to use digital services safely and independently. This inclusive approach ensures that protective measures do not inadvertently create additional barriers for children who may already face challenges in accessing digital technologies and services.

Communication strategies that clearly explain whether children are permitted to use specific services foster a sense of inclusion and agency in decision-making processes, allowing young people to understand and participate meaningfully in determinations that directly affect their digital experiences. This transparency supports the development of digital literacy and responsible online behaviour while respecting children's evolving capacities for autonomous decision-making.

Developing thoughtful measures to address situations where minors attempt to register for age-inappropriate services requires balancing respect for children's natural determination and curiosity with consistent application of protective rules and standards. Effective approaches acknowledge young people's agency while maintaining appropriate boundaries, avoiding punitive responses that might damage trust or encourage deceptive behaviour. Similarly, implementing comprehensive measures to discourage under age users from creating accounts through false information prevents the development of problematic patterns where children learn to routinely misrepresent their age to achieve desired goals, thereby fostering a culture of honesty and appropriate boundary recognition in digital environments.

Ensuring that minors can easily log out of services and delete their profiles and associated data provides young people with meaningful control over their digital presence and personal information. This empowerment supports the development of healthy data management practices while respecting children's evolving understanding of privacy and digital footprints. Such control mechanisms also align with broader data protection principles while acknowledging that children's preferences and comfort levels regarding digital participation may change over time.

The integration of prominent safety features, comprehensive risk identification resources, and accessible support materials provides essential protection and guidance for minors who may not yet possess the experience or developmental capacity to independently recognise potentially harmful online behaviours or situations. These educational and protective elements create layered safeguards that support children's safe exploration of digital environments while building their capacity for independent risk assessment and self-protection as they mature.

Analysis

Minors have limited attention spans, therefore cognitive load on the registration page matters. Most registration processes are trying to be as quick as possible in order for the user to start using the service. While trying to provide as much safety as possible considering we are dealing with minors, there is a chance that the information overload will frustrate both minors and parents and would discourage future use of the service. However, this can also discourage under-age minors from trying to circumvent the process by trying again.

Once minors are approved for the service, introducing safety features during on boarding is more user-friendly. Users are more receptive to safety information when they know they'll be using the platform, rather than while still uncertain about their access

Platforms can use interactive methods instead of long text blocks to share information more engagingly. However, incorporating interactive elements during registration may confuse minors about whether they have actually gained access to the service. Interactive safety content might signal to users that they're already using the platform, when in reality they're still undergoing a registration process that could deny them access.

Recommendations

1. **Develop a targeted on-boarding interactive experience**: once registration is complete with methods such as, e.g., tutorials, scenario-based questions ('What would you do if you land into inappropriate content?').

2. **Incorporate a 'help button':** that would always be visible, in a colour that makes it easily spotted, which can (a) lead to safety resources but also (b) run the tutorial or scenario-based questions again.

3. **Avoid long texts due to limited attention span**

4. **Deploy effective deterrents**: When designing measures to deter minors from lying about their age, ensure that such deterrents are effective, e.g. requiring for a parent's e-mail address or phone number to confirm age of minor

Account settings,online interface design and other tools

General comments

Homo Digitalis adheres to the Commission's preferred approach that sets minors' accounts to the highest level of privacy, safety, and security by default. The guidelines rightly note that minors tend not to change default settings, making defaults critical to their protection (behavioural inertia principle). Such an approach aligns with UNCRC Article 16 (right to privacy), Article 19 (protection from harm), and General Comment No. 25, which stresses that children must be provided by default with the highest safeguards online.

Homo Digitalis believes that privacy-by-default minimises exposure to exploitation, harmful content, or excessive data collection and supports preventive rather than reactive — blocking known risk vectors (like contact from strangers, auto play, and excessive engagement loops) before harm occurs.

Analysis

While the section is largely positive, several concerns emerge that deserve further consideration. The proposed guidelines do not provide clarity on age-differentiated settings, noting only incremental control "according to their age and needs" without providing specifics on how platforms should tier protections across distinct age bands such as under 13, 13–15, and 16–17.

Additionally, there appears to be an over-reliance on individual choice in opt-out scenarios. Although asking minors to confirm or modify choices represents a positive step, some platforms may exploit this as a loophole to nudge children toward privacy-eroding behaviours. Stronger design safeguards, including friction mechanisms, clear warnings, and automatic reversion to default settings, must accompany any choice architecture.

The proposed guidelines also lack requirements for parental involvement in certain high-risk changes. In cases involving enabling live streaming, sharing geolocation, or facilitating interactions with strangers, some form of verified adult consent could be warranted, particularly for children under 13. Furthermore, there is no mention of third-party data flows, representing a significant privacy gap. Default settings should govern not only platform features but also control data exposure to third parties, including advertisers and analytic tools, to ensure comprehensive protection of minors' privacy rights.

Finally, in its tours to school all over Greece, the Homo Digitalis team heard far too many times from the children themselves how they lost track of time due to never ending scrolling and pop ups, notifications and news that made everything look interesting and urgent. Despite the fact that the majority of the children recognised that this ‘addiction’ was partly caused by the *‘design features that are aimed predominantly at engagement that may lead to extensive use or overuse of the platform or the forming of problematic or compulsive behavioural habits*’ they could not always stop. This is why measures that not only enhance children’s awareness but also enable them,nudge them and protect them from spending too much time on the platforms are considered positive.

Recommendations

1. **Tier settings based on age groups**: Providers should adopt a graduated approach to default settings, calibrated by age groups (e.g., under 13, 13–15, 16–17), ensuring that younger children benefit from stricter safeguards.

2. **Prohibit manipulative design that nudges toward lower settings:** Design practices (e.g. dark patterns) that encourage minors to reduce their privacy or increase visibility should be expressly prohibited, in line with the principle of child-appropriate design.

3. **Make parental/adult involvement an option for high-risk actions:** In cases of high-risk setting changes (e.g. live streaming, enabling DMs from unknown users), platforms may offer options for verified adult support or approval, especially for younger users.

4. **Extend default protections to third-party integrations:** By default, minors’ data should not be shared with or accessed by third-party services unless those services are essential and compliant with children’s data protection standards.

5. **Require regular usability testing with children:** Platforms should regularly test whether privacy settings are understandable and accessible to minors of different ages, ideally via co-design or consultation with children

6. **Explicit limitation on data profiling and behavioural advertising:** Default settings should not only restrict interactions but also prevent data profiling and behavioural targeting—especially by third-party ad tech or algorithmic systems. Under UNCRC Articles 16 & 36, children must be protected from exploitation, including for commercial gain.

7. **Cross-device and cross-platform consistency:** Children often access platforms across devices. If settings vary between web and app versions (or reset after log outs), protections are undermined.

8. **Auto-reset to default after a defined period of inactivity or update:** If children adjust settings in a single session, those changes may persist longer than intended or be forgotten. An auto-reset mechanism ensures temporary adjustments don't expose minors to ongoing risks.

9. **Default minimisation of engagement-driven features:** Default protections should include design-level defences against features that exploit children's psychological vulnerabilities — such as infinite scroll, algorithmic loops, or real-time engagement nudges.

10. **Platform-level enforcement accountability:** The DSA's success relies on enforcement. There should be auditable accountability on how platforms apply and enforce default settings for minors.

11. **Clear opt-in requirement for any deviation from defaults:** Changes should be opt-in, not opt-out. This must include clear, accessible, age-appropriate explanation and must avoid nudging or manipulation.

12. **Inclusion of indicators for "privacy grade':**Visual cues (e.g. a traffic light system) can help minors understand how their privacy level changes when adjusting settings.

Recommender systems

General comments

Homo Digitalis believes that the draft guidelines for recommender systems correctly focus on the on mitigating the documented harms of engagement-driven algorithmic ranking systems for minors, by prioritising user agency in digital environments. Notably, the guidelines require the prioritisation of 'explicit user-provided signals' over 'implicit engagement-based signals', and clearly defines what is to be understood in these guidelines by these two terms.

However, Homo Digitalis notes that many other proposed measures remain insufficiently detailed for effective implementation. The high-level nature of these provisions may limit their practical utility for both regulators and industry stakeholders.

Analysis

In particular, Homo Digitalis considers that critical terms such as "safety," "security," "fairness," and "age-appropriate content" must be further defined. This fundamental ambiguity creates significant implementation challenges for development teams who lack concrete benchmarks against which to measure their efforts. The absence of specific metrics or user experience standards risks fostering superficial compliance rather than meaningful implementation of protective measures.

While this may represent a deliberate choice to preserve adaptability and avoid overly prescriptive regulations, it nevertheless creates a disproportionate burden on technical teams who must interpret vague expectations without adequate guidance. The guidelines would benefit from either establishing clear baseline definitions or requiring platforms to publicly articulate and demonstrate their own interpretation of these standards.

Homo Digitalis expresses significant reservations regarding the proposed blanket prohibition outlined in measure 535-8, which restricts the use of "ongoing behavioural data" in recommender systems serving minors. This provision, as currently drafted, may fundamentally undermine the basic operational capacity of recommender systems.

Essential interaction metrics, including watch time and click-through rates, constitute foundational elements for content curation and system optimisation, independent of user profiling activities. These minimal behavioural signals enable platforms to deliver relevant content and maintain system performance standards. The categorical prohibition of such data collection mechanisms risks rendering recommender systems technically inoperable or significantly degraded in their capacity to serve user needs effectively.

If the regulatory purpose is to prevent invasive user profiling practices, this goal should be explicitly stated and precisely delineated rather than implemented through broad prohibitions that may generate unintended consequences. Clear regulatory language distinguishing between essential operational data and problematic profiling activities would better serve both policy objectives and industry compliance efforts.

Additionally, the guidelines fail to provide essential implementation resources, including a taxonomy of technical measures and practical examples of inappropriate or harmful content beyond the isolated example provided in point 44. This complete absence of a living repository of examples, acceptable techniques, or benchmarking protocols may frustrate developers, complicate roll out processes, and cause confusion amongst technical personnel. While avoiding over-restriction of technical choices may be desirable, the lack of any practical guidance creates unnecessary implementation challenges.

As stated earlier,despite placing strong emphasis on testing, adaptation, and ethical standards, the guidelines provide no clarity regarding how strategic and technical measures under section 6.5.1 should be communicated to the public. This omission contrasts sharply with transparency provisions in section 8.4, which explicitly address section 6.5.2 requirements. Additionally, there is no guidance on third-party audits, external reviews, or feedback mechanisms. This inward-facing structure—designed by and for VLOPs—risks undermining accountability and could weaken the overall effectiveness of the guidelines by excluding independent oversight.

Furthermore, Homo Digitalis identifies potential conflicts between the proposed measures (576-590) and established regulatory mechanisms, particularly concerning their integration with existing parental control systems and investigatory safeguards mandated under the DSA.

The provision enabling users to reset recommender systems and the option for minors to select non profiling based recommendations present operational challenges for compliance with audit trail requirements and parental oversight mechanisms. Specifically, the autonomy granted to minors under point 3 to independently choose recommender options that eschew profiling may create tensions with parental supervision frameworks and complicate data collection protocols required for regulatory compliance.

These measures require careful calibration and clear guidance on the hierarchical relationship between minor autonomy provisions and existing parental control obligations would strengthen implementation coherence.

Finally, Homo Digitalis wishes to highlight that the cognitive burden imposed by comprehensive recommendation explanations risks overwhelming younger users, potentially leading to disengagement from safety mechanisms or diminished comprehension of relevant information. This approach may prove particularly problematic for younger minors, whose developmental capacity for processing complex algorithmic explanations remains limited.

Recommendations

1. **Issue a living Technical Annex**: Develop a dynamic technical annex that defines minimum safety baselines, standardised metrics (e.g. exposure-diversity index, true/false positive rates by protected attribute), and exemplar testing protocols. Require providers to publicly declare and justify the specific metric set they adopt, thereby promoting comparability and accountability.

2. **Introduce proportional external oversight**: Mandate independent audits for VLOPs and VLOSEs, ensuring that transparency and information-sharing obligations are sufficient to enable meaningful third-party scrutiny—particularly by civil society and academic institutions.

3. **Clarify the data collection provisions**: Refine the ongoing data-use restrictions to clearly prohibit full-funnel attention logging (e.g. dwell time, infinite scroll traces), while explicitly allowing (i) ephemeral, on-device processing or (ii) aggregated logs subject to differential privacy protections. This would safeguard minors' privacy without undermining essential system functions.

4. **Adopt a shared Risk Taxonomy**: Leverage an existing risk framework—such as the EU Kids Online model or the OECD four-pillar taxonomy—and require its consistent application across risk assessments, system metrics, and compliance reporting. This would foster semantic and procedural interoperability across providers.

5. **Establish a living multilingual registry of risky terms**: Create and maintain a continuously updated, multilingual registry of high-risk terms (e.g. harmful slang, emojis, hashtags), curated by the EU Centre for Algorithmic Transparency. Synchronised updates at regular intervals would help platforms keep pace with evolving online language and content risks.

6. **Ensure alignment with the AI Act**: Include guidance on the intersection of DSA obligations with AI Act requirements for High-Risk AI systems. This would enhance regulatory coherence and provide shared reference points for technical and procedural safeguards.
7. **Design age-appropriate explanation interfaces**: Introduce alternative methods for informing minors why particular content is recommended. These might include visual cues, gamified interfaces, or contextual prompts, drawing from design research in the UK Children's Code and cognitive UX models from JRC and UNICEF. Such mechanisms would better align with children's cognitive capacities and engagement patterns.

Moderation

General comments

Homo Digitalis believes that the proposed measures are efficient and necessary to moderate inappropriate, illicit and harmful content. More specifically (numbers according to measures): a clear definition of harmful content and behaviour for minors makes large platforms and service providers accountable to their own rules by having a strong basis as to why they make a certain decision to bar specific content from the minors' feeds.

This also helps build trust between service providers and parents. Minors will also not be exposed to inappropriate content which also protects them from unwanted exposure and therefore also cultivates trust to the platform. Collaboration with independent experts and civil society is an ideal step to mix law with children's rights advocacy groups and societies, psychologists which gives greater insight into the design of the platform.

Establishing moderation policies and procedures to ensure that harmful content is detected and moderated proves efficient internal governance structures which shows that the service providers commits in practice to moderate content. It also improves transparency, especially to parents of minors who may actually need to access these policies and read what content is permitted for their child to see and what not. Providing criteria which show why the content was moderated also ensures that the platform's algorithms have been built with the best interests of minors in mind and avoid unaligned decisions regarding the permission to display harmful content.

A risk-based approach is always vital when exploring different pathways on a specific activity which may produce different outcomes. Therefore, assessing a specific harmful content based on its impact and likelihood effectively manages higher and more imminent risks. Also, it can create a pattern or method of risk occurrence and aid the platform to easier identify and rate certain risks.

Human review is an effective means for regulation and avoidance of mistakes made by any automated decision making tools. Especially regarding content which exceeds average views, and therefore reaches a big audience, human review is essential to understand any potential threat, common risks, identify viral trends and decide based on the best interests of children.

Implementing effective tech to detect harmful or illicit content before this is made public is an effective measure focusing on prevention rather than implementation of ad hoc measures.

Analysis

Defining harmful content can sometimes become tricky, especially when the service provider and large platform does not operate in Europe which shares the same liberal values. A platform designer outside the EU may consider content which may seem inappropriate to European users as appropriate in the country of origin. Also, marginalised societal groups (LGBTQ+ community) may have different perceptions of what constitutes inappropriate content, while religious groups and societies may have the opposite view. Often, definition of such terms can be subjective.

Sometimes algorithms or other automated detection tools can make mistakes which may lead to under or over blocking. Such mistakes can be made due to several non-detected faults, such as: supposedly friendly content based on keywords or phrases, the covert meaning of which can only be detected by a human brain, disguised hate or sexist speech. Also, there are certain minority languages which the platform may want to serve for purposes of demographics or inclusivity, however there may not be enough moderators to cover the volume of incoming information.

As with all automations and decisions, they may – at times – be subjective. An adult, or even an algorithm, may not have the clearest perception of a minor’s life or threats which may lead to harm to the minor. Also, some correctly identified low risks (e.g. body shaming posts) which at once, do not seem harmful to one individual of a certain age category, may, if show repeatedly and without moderation, lead to greater future harm and have psychological effects.

Similarly with automated decision making tools, humans also have biases which can prove to be discriminatory and arbitrary. This may have as a consequence that some reported accounts may be deactivated with the premise of posing a risk of harm to minors’ privacy, safety and security, whereas at the same time other accounts may remain active based on the determination of another person conducting human review, coming from a different background with different biases. Where algorithms base their decisions on a specific model, humans take decisions based on different factors, and thus, human decisions vary. Also, given that human review in social media platforms is scarce and requires resources, there is a risk of over-reliance on reported content from potentially harmful accounts rather than focusing on moderated content that has not been yet reported.

As with all technologies that ultimately take and adopt decisions which may have effects on natural persons, AI tools and systems may present false positives and false negatives. Considering that the purpose of engaging AI tools and systems to take such decisions is a measure designed to minimise the need for human review, there is a risk that a decision taken by such a system or tool may be definitive and a potential correction of this action (based on user’s contested decision) may require even more resources. Also, platforms using such systems and tools may operate differently than others, not using these tools or using different ones and cause a lack of standardised implementation across similar child-reaching platforms.

Recommendations

1. **Clarify on harmful behaviour**: Considering a clear and finite definition of harmful content may not always be achieved due to conflict of interests or the bridging of cultural gaps, platform providers need to be absolutely clear on what constitutes harmful behaviour as they may, otherwise, run the risk of an algorithm making an ultimate decision as to whether to display specific content, which may lead to further issues.

2. **Coordinate with civil society organisations**: Share the policies and procedures with civil society groups in order to show how a platform moderates its content in order to gather feedback from other specialists or independent experts. Also, testing a moderation mechanism or the person performing the moderation about potential biases.
3. **Provide adult-appropriate training:** developers need to familiarise themselves with modern causes of harm and trauma to minors (by sourcing data from civil society or therapists) in order to be made aware and classify certain behaviour and content accordingly. Also, during the operation of the service, the platform provider may want to gather user and parent feedback based on their own experiences, even outside the platform, in order to better classify harmful content.
4. **Clarify content that 'substantially extends' average number of views**: It could be useful to set a number of criteria that will indicate when content is considered viral reaching to a wider audience in order to ensure that specific resources are efficiently allocated to human review. In cases where human review may take time or the reviewer is unsure of the determination, it may be useful to (a) establish a second review stage where content could be escalated for scrutiny by experts, moderators with child rights advocates experience and (b) decide to blur any content which has not yet been classified as harmful.
5. **Conduct regular audits**: may be needed to test such AI systems and tools in order to ensure their effectiveness in moderation, reviewed based on novel technological criteria for AI moderation tools. Human in the loop mechanisms shall also be explored in order to ensure that effective escalation to human review is achieved.

User reporting, feedback and complaints

General comments

Providing child-friendly and accessible feedback mechanisms supports children's sense of agency and autonomy in reporting as well as enhances control over their privacy and safety. Most importantly, it is children that know better than any other expert, civil society group or children rights' advocate how they feel about the content they are exposed to online. Therefore, they are the most suitable to provide effective and constructive feedback.

Analysis

Given that reporting, feedback, and complaint mechanisms are specifically designed to serve the interests and protection of minor users, Homo Digitalis emphasises that these tools must prioritise child-friendly design principles and accessibility standards. The effectiveness of these protective measures fundamentally depends on their usability and comprehensibility for users across different developmental stages.

Allowing minors to report content (e.g. accounts, groups, pages) as inappropriate is a powerful tool which can, however, be used both in a productive but also in an arbitrary way. In a similar way, allowing reporting go suspected under age accounts may have the intention to protect minors, however it may also lead to targeted actions of certain minors against others.

Recommendations

1. **Develop child-friendly design features and standards**: too complex and detailed report features will only deter children from providing their feedback. Measures to be taken to ensure that such tools are accessible to children, is providing them with visuals throughout the reporting and understandable language. Accessible feedback, reporting and complaint tools should also take into account minors with disabilities.

2. **Prevent arbitrary or revenge reporting**:Minors should be provided with definitions or examples of harmful content or behaviours that are as clear as possible for a minor to understand. Reporting of suspected under age accounts can be restricted to a certain number of reports per day/week/month, in order to deter any false harassment.

3. **Enhance clarity**: Any distinct reporting and feedback tools should be very clear to understand, should not cause ambiguities to minors or confusing them into what and how to report. Adding visual images or emojis on any reporting mechanism may engage minors to report more accurately and provide me with a safe space to communicate their feelings.

4. **Provide tracking tools in the complaints' process**: as well as being provided with a follow up for taking a certain decision (blocking content or user) as well as rationale for not doing so. Minors need to know that their voices are heard and that complaints are not just a bureaucratic practice.

User support measures

General comments

Homo Digitalis strongly supports the proposed measures as they show a commendable understanding of children's needs in digital environments. These measures address critical vulnerabilities children face online, including exposure to harmful content and their inability to find appropriate help when needed. The proposals promote child-friendly support mechanisms and effectively integrate national resources such as INHOPE and Safer Internet Centres into comprehensive protection frameworks. Furthermore, they demonstrate recognition of the unique risks posed by AI tools in digital spaces and empower children with meaningful controls over their online interactions, including blocking, muting, commenting, and group participation features.

This section aligns closely with fundamental principles established in international children's rights frameworks. The measures support UNCRC Article 3 regarding the best interests of the child, Article 12 concerning children's right to be heard and participate in decisions affecting them, and Article 19 addressing protection from harm. Additionally, these proposals are consistent with General Comment No. 25, which specifically addresses children's rights in the digital environment and provides guidance for protecting and promoting these rights in online spaces.

Analysis

Homo Digitalis believes that while the section is strong, a few gaps and implementation risks could be addressed. For instance, the use of non-binding language, particularly the term "should," significantly weakens enforcement mechanisms. For systemic protection of children in digital environments, key safeguards such as access to support tools and the ability to block users must be mandatory rather than discretionary.

Additionally, the guidelines lack essential specifications regarding response time frames and escalation procedures. Platforms should be required to respond to children's reports within clearly defined time limits and must have protocols for rapidly escalating urgent issues such as grooming attempts or self-harm content. Without these temporal requirements, children remain vulnerable during critical periods when immediate intervention could prevent serious harm.

Furthermore, the framework fails to incorporate children's participation in designing support features. Young people's voices should directly inform how help systems are structured and presented, particularly regarding usability and trust factors that determine whether children will actually utilize these protective mechanisms when needed.

While AI warnings represent a positive step, they may prove insufficient without additional safeguards. Simply notifying minors that they are interacting with AI systems does not adequately address the risks of AI-generated misinformation or emotional manipulation, both of which can have serious impacts on children's mental well being and decision-making processes. More comprehensive protections are needed to address these emerging technological risks.

Homo Digitalis acknowledges that guardian tools serve a complementary function in digital child protection, while maintaining that primary responsibility for safeguarding minors rests with online platform providers. We support measures designed to protect children from potential misuse of monitoring capabilities by guardians, particularly the implementation of real-time notifications that clearly inform minors when any surveillance functionality is active.

The effectiveness of parental controls and similar guardian tools varies significantly across age groups and remains particularly limited in contexts where ongoing negotiations occur between parents and children regarding screen time, application permissions, and digital boundaries. These practical limitations reinforce our position—and align with the European Commission's stance—that such tools fulfil an auxiliary rather than primary protective role. While guardian tools may provide some oversight capabilities, they cannot substitute for the fundamental obligation of platform providers to design inherently safer digital environments for young users.

Finally, it should be highlighted that many minors may have limited literacy, be neurodivergent, or speak a minority language. The guidelines need to clarify and interpret the notion of “child-friendly” in inclusive terms.

Recommendations

1. **Make critical support functions mandatory:** Platforms must provide visible, accessible support tools for minors to report harmful content, and enable anonymous blocking/muting of other users.”

2. **Introduce response timelines and escalation protocols**: Reports made by minors should trigger timely responses. Platforms must define clear timelines for acknowledgement and action, with an escalation path for high-risk issues.

3. **Require co-design or usability testing with children:** Support tools should be developed with input from children of different ages and backgrounds, including those with disabilities, to ensure they are intuitive, non-intimidating, and effective.

4. **Enhance AI interaction safeguards:** In addition to labelling AI interactions, platforms should limit the topics AI tools can discuss with minors (e.g., avoid medical or mental health advice) and apply monitoring to detect manipulation or harm.

5. **Ensure accessibility and inclusion:** Support tools must be available in multiple languages and be accessible to children with disabilities. Visual cues, icons, and voice prompts can help ensure understanding and use.

6. **Offer feedback or resolution mechanisms:** Minors should be informed about the outcome of their report (e.g., content removed, user warned), and given the option to provide feedback or request further action.

7. **Include proactive support and digital resilience education:** Waiting until a child seeks help is reactive. Platforms can proactively support children through in-app guidance, digital literacy prompts, and scenario-based education on what to do when facing harm.

8. **Introduce a "trusted adult" or "designated guardian" feature:** Some children are more likely to seek help from a known adult rather than a faceless report mechanism. A designated adult contact option (e.g., teacher, parent, school counsellor) within the platform could provide an added support layer.

9. **Require trauma-informed design and moderation:** Children reporting abuse, exploitation, or distressing content may relive trauma when describing their experience. Support tools should follow trauma-informed principles (e.g., not requiring detailed descriptions, avoiding judgemental language).

10. **Protect privacy when seeking help:** Children may fear being monitored or punished for reporting. Their ability to seek help must be protected by confidentiality assurances, including not alerting abusers (in domestic violence or coercion situations).

11. **Facilitate offline follow-up if needed:** Some cases (e.g. grooming, self-harm threats) require offline intervention by child protection services or police. Platforms must have a clear protocol to refer serious risks to competent national authorities.

12. **Implement user feedback and continual improvement:** Children's needs and risks evolve. Platforms should regularly solicit feedback from child users on how well the support tools meet their needs, and publicly report improvements.

Governance

General comments

Homo Digitalis expresses its support for the proposed governance framework, which demonstrates a mature and holistic understanding of the institutional requirements necessary to embed effective child protection within digital platform operations.

The Commission's approach towards governance extends beyond mere regulatory compliance to encompass the cultivation of organisational cultures grounded in responsibility and accountability toward child welfare. The framework appropriately calls for the establishment of dedicated roles with substantive authority and direct access to senior leadership, ensuring that child protection considerations receive appropriate institutional priority and resources.

Furthermore, the emphasis on child participation represents a fundamental acknowledgement of minors as stakeholders with legitimate voices in shaping the digital environments they inhabit.

This governance framework aligns with the fundamental principles established under the UNCRC, supporting the mandate of Article 3 by prioritising the best interests of the child in all decisions affecting minors, and Article 12 by committing to including minor voices in governance processes. The comprehensive protection focus aligns with Article 19's requirements for safeguarding children from harm, and the framework's digital-specific considerations reflect the contemporary guidance provided in General Comment No. 25 on children's rights in the digital environment.

Analysis

Having said that, several critical areas require clarification, enhanced accountability measures, and more robust commitments to ensure effective implementation.

As stated earlier in this document, the use of "should" rather than "must," significantly undermines the enforceability of essential governance provisions. The assignment of dedicated child safety officers and establishment of direct reporting lines to senior management represent fundamental institutional requirements that cannot be treated as optional recommendations. These core governance structures require mandatory implementation through binding regulatory language to ensure consistent application across platforms and prevent selective compliance.

Similarly, the effectiveness of child safety officers and dedicated teams fundamentally depends on their professional expertise, institutional independence, and substantive decision-making authority. The current guidance fails to establish minimum qualification standards, independence requirements, or decision-making parameters for these critical roles. Without clear specifications regarding professional competencies, reporting structures that preserve independence from commercial pressures, and explicit authority to implement protective measures, these positions risk becoming ineffective or merely symbolic appointments.

Although the framework appropriately encourages child participation in governance processes, it lacks essential guidance for ensuring ethical, inclusive, and meaningful engagement. The current provisions do not address critical implementation considerations, including safeguards against tokenism, methodologies for ensuring diverse and representative child voices, and systematic feedback mechanisms that demonstrate how child input influences platform policies and practices. Effective child participation requires structured protocols that not only solicit input but transparently communicate how that input shapes decision-making processes.

Finally, the proposed governance framework lacks clear requirements for public disclosure of outcomes, including systematic reporting on identified harms, remedial actions taken, and institutional gaps discovered through governance processes. Without mandatory transparency measures and external accountability mechanisms, governance structures risk operating without meaningful public oversight, limiting their effectiveness and undermining public confidence in platform child protection efforts.

Recommendations

Homo Digitalis believes that the proposed governance structure can be strengthen in six main ways:

1. **Make core governance measures mandatory**

2. **Define minimum standards for child safety officers**
3. **Operationalise meaningful child participation**
4. **Require public transparency reports**
5. **Incentivise cross-platform accountability**
6. **Include whistleblowing and internal escalation mechanisms**

Monitoring and evaluation

General comments

Homo Digitalis believes that the proposed measures take the right direction by promoting continuous safety testing, stakeholder consultation and child-centred transparency. Yet they remain programmatic. Without clearer benchmarks, sampling methods and disclosure formats, platforms will implement these duties unevenly and regulators will struggle to verify compliance.

Analysis

The operational efficiency and adequacy of this regulatory framework face significant definitional and implementation challenges. The requirement for "effective monitoring" remains fundamentally undefined, lacking essential reference metrics, standardized incident-logging schemas, or specified review frequencies that would ensure consistent application across platforms. This ambiguity creates a regulatory environment where some platforms could potentially satisfy compliance obligations through minimal quarterly surveys, while others might implement comprehensive real-time monitoring dashboards, with both approaches theoretically meeting the same regulatory standard.

Similarly, the consultation duty imposed on providers suffers from excessive open-endedness, requiring platforms to "regularly" engage with diverse minor populations without providing crucial operational guidance. The absence of clear sampling methodologies, parental consent frameworks, or privacy protection protocols leaves providers without adequate direction for implementing meaningful consultation processes while safeguarding vulnerable participants. This regulatory gap creates both compliance uncertainty and potential risks to the very populations the policy aims to protect.

Furthermore, the adjustment mechanism embedded within the framework delegates critical implementation decisions entirely to provider discretion. While platforms must "adjust" their designs following consultations with minors, the regulation imposes no requirements for documenting intervention outcomes, measuring risk-reduction effectiveness over time, or demonstrating continuous improvement in safety measures. This absence of accountability mechanisms and performance tracking requirements undermines the policy's potential effectiveness and creates substantial challenges for regulatory oversight and enforcement.

Recommendations

1. **Baseline monitoring framework**: mandate core indicators (for example, monthly privacy-incident rate per 10 000 minors, average time-to-resolve reports) and a minimum review cadence (for example, at least quarterly).

2. **Structured consultation protocol**: supply templates covering consent, age-stratified sampling, accessibility checks and anonymised feedback storage, aligned with GDPR and children's-rights guidance.

3. **Documented "adjust-and-prove" loop**: require a public change-log showing each design adjustment, the trigger (consultation, incident data, research) and the measurable effect after N days.

4. **Layered transparency standard**: recommend icon sets or traffic-light badges plus expandable detail; reading-level targets should track EN 301 549 or WCAG literacy criteria.

Conclusion

As these protective measures advance toward implementation, critical questions remain regarding impact assessment and effectiveness evaluation. How will the European Commission measure the tangible difference these interventions make in children's digital experiences? What indicators will demonstrate progress compared to pre-regulation baselines? Establishing robust evaluation frameworks with clear metrics will be essential to assess whether regulatory objectives translate into meaningful protection improvements for young users.

Central to long-term success is comprehensive digital literacy education. Homo Digitalis firmly believes that minors must receive structured education on digital rights, consumer protection, and children's rights throughout primary and secondary schooling. Only through understanding why privacy safeguards matter and how to navigate digital interactions safely can young people become empowered participants in their online lives rather than passive subjects of protection.

Finally, Homo Digitalis, believes that despite well-intentioned policy goals, fundamental incompatibilities with current technical capabilities may be presented, and recommend that future regulatory initiatives incorporate greater consideration of technical feasibility to ensure that policy objectives can be meaningfully achieved within existing technological paradigms, rather than mandating solutions that exceed current technical possibilities.