Yoti Response to eSafety consultation

19 December 2023

Contact details:

Respondent full name:
Julie Dawson, Chief Policy & Regulatory Officer
Florian Chevoppe-Verdier, Public Policy Associate

Email address:
[redacted]

About Yoti

Yoti is a digital identity company that makes it safer for people to prove who they are. Founded in April 2014, we started by creating a secure Digital ID app which gives people a safer and instant way to prove their identity, with no need to show identity documents or share an excessive amount of personal data. Yoti now provides verification solutions across the globe, spanning identity verification, age verification, age estimation, eSigning and authentication. We're a team of over 400 people, working together to shape the future of digital identity.

We're committed to making the digital world safer for everyone. Our seven ethical principles guide us in everything we do and we're held accountable by our independent Guardian Council, whose minutes we publish. With an award-winning social purpose strategy, we're always looking for new ways to explore what (digital) identity means globally. The journey isn't one we're making alone, but with the help of policy advisers, think tanks, researchers, humanitarian bodies and everyday people.

What we are doing and why:

- Transforming the way individuals can prove their age and identity
- Increasing security and privacy of personal data
- Helping to create age-appropriate experiences and safer communities online
- Creating the most reliable and comprehensive identity verification solutions
- Shaking up the way we sign documents

Technology as a force for good - Yoti was founded on seven business principles which guide our actions. Yoti is also a founding UK B Corp meaning we aim to balance profit with purpose.

Security credentials - We commission regular external audits of our business and have been certified to meet some of the world's most stringent security standards, such as ISO 27001 and SOC2 Type II. We are also certified by the UK Government under the UKDIATF.

A transparent, open and honest approach - Yoti publishes regular white papers to build trust and understanding of our technology.

Responses to Consultation Questions

Question 1 Are the requirements for risk assessment in the draft Standards targeted at the right services and at the right points in a service's development journey?
Are the risk factors appropriate?

We would suggest that age appropriate design and review of age assurance should be part of the risk assessment at the outset of service development.

We would welcome the addition of a duty for adult content providers to make the result of their risk assessments publicly available, in order for the regulator to be able to access them, as well as members of the public and non-profit organisations to access them, and bring them to the attention of all responsible regulatory authorities for review.

This is important as it is unclear what degree of transparency will exist during the pre-assessment stage, on which the type of risk assessment that will be conducted depends.

We would also highlight that the proposed categorisation as currently set out on table 1 of this consultation document could risk creating confusion when assessing the nature of a service provider. Indeed, in many instances, a socal, dating or gaming platform can evolve over time to host in part or as its main type of content offering, age restricted and pornographic content.

[In the categorisation social, dating, gaming platforms are in this consultation, adult in the next, however there is known to be large volumes of adult content within social and there are hybrid 'fan sites' and gaming sites which include adult content.]

Most relevant electronic services will fall within a category as set out in Table 1.
They are deemed to have a specified risk profile, and are therefore not required to conduct a risk assessment unless they make a material change to their service.

Question 2 **Do the obligations** on each relevant electronic service and designated internet service category appropriately reflect the above considerations? **Are other considerations relevant?**

Given the widespread evidence emerging in US of social platforms knowingly having under 18s and 13s on their platform, should the service design for social media platforms that are known to attract and keen to target under 13s be required to look for and evidence 'actual knowledge' of their users and to then be required to implement age appropriate design and age assurance as a default requirement.

It is worth considering, should transparency requirements stipulate the gathering and sharing explicitly of the demographic of users and hence considering the relevant risk assessment and what age appropriate design practices need to be put in place in terms of minors.

We welcome the fact that '*eSafety will be able to receive complaints and investigate potential breaches of the Relevant Electronic Services Standard. When assessing whether adopted compliance measures are* ***reasonable****, eSafety will consider a range of factors including the capability and* ***size of a service provider*****.**'

This begs the question, would lower health and safety standards be expected of a small food outlet; would lower child safety standards be expected of a smaller child care facility, would lower technical standards be expected of a small garage providing a car service? It is worth considering, on what basis, is it appropriate that a smaller online organisation should not be expected to provide age appropriate design and safety measures to protect and enable young users to thrive…

In addition, we would also encourage that the complaints and investigation be accessible to civil society organisations and to citizens.

Question 3 Is the test in section 5 workable?* Is further guidance required to assist providers to determine whether this standard, or another code or standard, applies to a particular online service?

*Section 5 of the draft Relevant Electronic Services Standard provides that, **where a single electronic service could potentially fall within the scope of the Relevant Electronic Services Standard and also an industry code or another standard under the Act, the service provider will only be required to comply with one code or one standard in relation to that specific service** (until the relevant industry code or standard dealing with class 2 content comes into effect).

There is an obvious problem of legibility both from an industry & citizen perspective, it is hard to navigate the multiple layers of law, standards, and codes that will apply*. We would encourage all government & regulatory actors involved in the definition of Australia's online

safety regime to collaborate to point to a document or website that could set out more clearly what duties will co-exist.

The codes currently suggest that 'Where a single online service could fall within the scope of more than one industry code or industry standard, the code or standard that will apply is the code or standard that the service's predominant functionality **is most closely aligned with**.'

Enabling providers to determine which code is most closely aligned with their predominant functionality risks that providers could self-assess to place their service under the code with the least requirements on their services, or which positions them best for advertising or investors, at the expense of the safety of their users.

As we have said previously, we believe there would be value in publishing those self-assessments too, in order to enable the regulator and citizens to read, understand and possibly challenge those decisions..

An example of the risk that allowing providers to essentially choose a single code are blended platforms e.g. social media platforms with large volumes of adult and harmful content material, or gaming & fan platforms operating also in the adult content area. For the largest of platforms, they could very feasibly be both a global social media organisation and rank in the top sources of adult content.

In those instances, we would suggest that it would be impractical and potentially unsafe to only leave firms to follow one code. Further, as codes and standards evolve, firms could decide to place new services or move existing services under whichever has the fewest perceived restrictions.

(*e.g. to signpost how to access and understand the test in section 5 as a case in point)

Question 4 Is the technical feasibility exception in the obligation to detect and remove known child sexual abuse material and pro-terror material appropriate? How effective will this obligation be with this exception?

As currently worded, the requirements and exceptions under section 7 ('technical feasibility') risk leading to a race to the bottom, and could incur significant legal uncertainty. We do not believe that the obligation will be effective in the way this exception is currently worded.

Is there not the need simply to understand the age of users and design services to be 'in the best interest of the child' and age appropriate.

E.g. Consumer device manufacturers could consider offering options for devices for young or vulnerable users with a pre-set functionality that does not allow the outbound sharing of nude images.

The key problem associated with this duty is the need for the regulator to define the terms *'technically feasible'* as well as the adjective *'reasonable'* which are employed in section 7 ('Technical feasibility'). In absence of a definition by the regulator or lawmakers, it will likely be left to a court of law to decide, which will incur significant cost and delay for Australian residents/taxpayers. It could lead to a race to the bottom.

In a previous response to the eSafety Commissioner's call for evidence on age verification, we debunked the all too often used myth that online safety technologies, such as age assurance technologies, are too costly for firms to implement. In a response to a public consultation on the Privacy Legislation Amendment Bill 2021 exposure draft, we had also emphasised that when it comes to non account based checks (such as for gaming terminals and adult content sites where users do not have an account), rates can be as low as cents or even fractions of cents per check and that smaller platforms and sites will not be at a competitive or financial disadvantage either.

The regulator should be required to provide clarity as to industry pricing through the use of its information powers. There have been clear instances where platforms and the media have overexaggerated the costs of compliance, as a rationale to not invest in safety tech approaches and deem it economically unviable.

There is a lack of clarity as to whether firms would be required to publish motivated decisions on technical feasibility as well as estimated financial cost. Similarly, it is unclear whether they would be required to share them with the eSafety Commissioner, and if so whether citizens and civil society organisations could challenge those decisions and ask for reviews. There is a risk that decisions could be based on exaggerated information.

In terms of applying protective measures - eg age assurance, is there a requirement for platforms to justify not investing. Is there not a minimum threshold of investment eg a percentage of turnover into safety measures following the, 'polluter pays principle'.

It is not cost prohibitive to small platforms, nor too intricate; age assurance can be implemented within a few hours and at low cost for low volumes. If you took the parallel of small playground equipment manufacturers - they would not be able to bring materials to market that do not meet health and safety standards. Similarly a garage offering car servicing or a nursery providing childcare - even if they are small scale organisations, they are required to meet specific standards. Hence given the risks clearly detailed on online

platforms, surely there should be a parallel level of investment to meet minimum standards, irrespective of the scale of the organisation.

It is also worth the regulator reflecting that investors in the safety tech sector have highlighted that without regulatory certainty, there is a disinclination to investment.

Question 5 Are there other examples of systems, processes and technologies that can detect, flag and/or remove known child sexual abuse material and known pro-terror material at scale, which should be highlighted in the Standards or accompanying guidance?

There is clearly growing evidence of user generated Child sexual abuse material (CSAM) by minors. Despite the existence of nudity detection algorithms and age assurance algorithms, which could prevent the outbound sharing of CSAM at scale by young users, there is no requirement currently for consumer equipment - handset or headset manufacturers to either offer default settings to assess the age of users ahead of the outbound sharing of nude images..

Question 6 Are there any limitations which would prevent certain service providers from deploying systems, processes and technologies to disrupt and deter child sexual abuse material and pro-terror material on relevant electronic services? If there are limitations, how might these be overcome? Is it appropriate for this requirement to apply to gaming services with communication functionality?

We would re-emphasise that online safety technologies, such as age assurance technologies, are a cost of doing business and not too costly for firms to implement nor would they put smaller firms at a competitive disadvantage against larger ones.

As we have emphasised in several of our previous consultation responses to the eSafety Commissioner, rates can be as low as cents or even fractions of cents per check and that smaller platforms and sites will not be at a competitive or financial disadvantage either. Therefore the argument of cost should not be permissible unless in very specific circumstances which should be assessed by the regulator, and consideration made either for public support or support based on an industry wide tax.

It is useful to consider the parallel within the environmental industry where for over thirty years now, the 'polluter pays' principle' (PPP)[1] has become widespread. The polluter pays principle means that, where possible, the costs of pollution should be borne by those causing it, rather than the person who suffers the effects of the resulting environmental or

[1] Sustainable Development Law in the Courts: the Polluter Pays Principle, Australian Parliament House website, Sustainable Development Law in the Courts: The Polluter Pays Principle

societal damage, or the wider community. In this instance the service provider which offers individuals a service, or builds a platform on which it invites persons to consume content would be the responsible party.

The polluter pays principle serves several functions and may be used through different phases of policy making. It can be used upstream in the design of a policy to prevent or deter environmental or societal damage. In cases where pollution cannot be avoided or is caused by accident, the polluter pays principle can be used to restore or redistribute the costs of damage. This helps to incentivise individuals or groups to avoid causing damage and encourage sustainable practices.

The polluter pays principle is applicable where there is evidence of, or potential for harm or a negative environmental effect; and prevention of that harm is not possible or proportionate.

Policymakers should consider:

1. Who the polluter is: the polluter could be an individual, group or sector. Considerations for making this assessment include:
    - What is the driver for the pollution being caused and who is responsible for this? It may be difficult to identify or define the polluter, the source of pollution and the associated cost over time. Policymakers must use judgement to identify who the polluter is and the extent to which the polluter ought to and is able to pay.
    - How does the allocation of responsibility for the pollution cause the most environmental benefit?
    - Who is it fair to expect to pay for the pollution? It may be more effective to distribute the cost across a particular sector responsible for the pollution, rather than place it on an individual or group.
2. How much the polluter should pay: the polluter pays principle should be applied proportionately. This means that the amount the polluter pays should be proportionate to the environmental damage and the wider costs and benefits to society of the activity in question. When deciding how much polluters should pay, policymakers must consider the value of the damage caused by the polluter or the potential polluter, along with the costs and benefits associated with the polluter paying (fully or partially) for this damage. In some cases, full cost recovery may

not be possible or proportionate and in these cases it may be reasonable that the cost is covered through other means.

3. How the polluter should pay: the polluter can pay in a variety of different ways depending on what is appropriate, and how this can act as an incentive or disincentive for action. It may be most appropriate for the polluter to pay directly through fees or charges, or indirectly through regulatory or contractual requirements (which in turn require additional investment to fulfil) to ensure the outcome minimises the environmental damage. In the latter instance, fines or penalties for breaching these obligations may also be appropriate.

Is it appropriate for this requirement to apply to gaming services with communication functionality?

We think that it should and we would advise that age appropriate design considerations are needed across all services, including gaming services with communication functionality.

Risk assessments should use the '4Cs of risks' for minors
Apply age assurance should be required at key inflection points eg upon sign up at and at agreed periods for re authentication that the person has not handed an account to a minor if higher risk activity is being entered into, e.g.

- Contract, purchase or other legal or financial mechanism engaged
- Contact with others - eg when accessing live chat functionality which could lead to requests/incitation to create user-generated child sexual abuse material.
- Content risks of nudity, obscenity e.g. when accessing 18+ rated games
- Conduct risks eg when accessing mixed audience or 18+ rated games, different standards of content moderation should apply

Question 7 Are there other examples of systems, processes and technologies that can disrupt and deter the use of a relevant electronic service to solicit, generate, distribute or access child sexual abuse material and pro-terror material, which should be highlighted in the guidance?

Technology and methodologies exist today, that would help break the cycle by taking the key stage of when child sexual abuse material content is uploaded onto an electronic service, prior to distribution. During the crucial stage of content upload, sites should be required to verify that the person uploaded content is over 18, and that due consent has been collected.
.

During the content moderation stage and after content has been uploaded, age assessment tools should be deployed to ensure that all individuals, co performers in content are over 18 and have given their consent to be in the content.

For content it would be possible to require traceability to a known 18+ individual; to make it clear to individuals that they're accountable for their content

For individuals who have infringed platform guidelines, nudges and education should be provided. It should be clear to users that individuals infringing platform standards can risk proportionate sanctions, including barring. Platforms should be allowed to maintain their platform guidelines and appropriate governance. For instance an individual company may set up their own or take part in a shared register [2]of self excluded or barred individuals according to their platform terms of service, governance and oversight policies. A platform may decide to outline and put in place proportionate sanctions, for instance to not automatically allow subsequent anonymous re registration from certain individuals where community guidelines have been breached.

Question 8 Do you agree with the monthly active user threshold for the investment obligation? Are there other appropriate thresholds that should be considered to ensure the obligation is proportionate to the size and reach of the relevant electronic service?

We would disagree with the proposal of 1 million monthly active users as a minimum threshold to be required to comply with obligations. Small, niche and hidden sites can host very harmful content and provide even greater risks, by operating slightly under the radar, being listed as shell companies and frequently changing their urls. High risk activities at any volume should require compliance.

The justification for differential obligation between small and large companies seems to be based on what they do already rather than the impact of the harmful consequences. This is a quantitative assessment of harm - how many people, rather than how badly they can be impacted.

Unless there is a clear definition of what is a reasonable programme of risk assessment, investment and development, from our experience again there is a risk of a 'race to the bottom' or total disregard for the implementation of basic safety tools.

It would be unwise to say that a smaller platform does not need to undertake a robust risk assessment, and if they don't assess their own risk adequately, then they also don't have to

[2] https://www.technologycoalition.org/newsroom/announcing-lantern

comply with all the measures in the codes. This could lead to a small but high risk service not being required to address the illegal content duties in the same way. This would render the Online Safety Act less effective.

Taking again the parallel with the retail of food - all organisations are required to respect the same minimum food hygiene standards[3] if they are serving food to the public, from the smallest cafeteria to a mass catering establishment and to be open for inspection. Strict standards would also apply to settings offering in person childcare or play facilities.

As per the previous answer, to question 6, with regards the 'polluter pays'; should the regulator deem that additional support is needed for smaller sites of under 1 m users, then that should be considered by the regulator and solutions proposed e.g. a github technical repository with case studies of effective compliance, examples of appropriate safety measures, free templates, free pre recorded webinars to supporter smaller scale operators to achieve cost effective adherence to agreed minimum standards.

To take an example, a small adult content site with under 1 million users, per year, could access age assurance options for under [£2k sterling] AUD3.7k per annum which can be implemented in 2-4 hours. This is hardly a disproportionate burden. Businesses do not pay for Over 18 shares from individuals using the Yoti reusable digital identity app

In principle, all organisations offering a service which poses risks in terms of CSAM, whatever their size should be in scope…

Question 9 Are the end-user reporting requirements workable for the relevant service providers? Are there practical barriers to implementation?

We would agree that end user reporting should apply to all service providers.
It should be specified that end user reporting should be in plain language, to meet age appropriate comprehension level for demographics using the service, including minors.

The Australian Government Style Manual[4] suggests '*that about 44% of adults read at literacy level 1 to 2 (a low level), People at a reading level 1 read at a primary school equivalent level. They can understand short sentences'.* Hence setting a reading age of 9 years of age, could

---

3 https://www.foodstandards.gov.au/publications/safefoodaustralia#:~:text=2A%20is%20a%20new%20food,activities%20are%20being%20adequately%20managed.

4 https://www.stylemanual.gov.au/accessible-and-inclusive-content/literacy-and-access#:~:text=Reading%20levels%20in%20Australia&text=In%20Australia%3A,5%20(the%20highest%20level). #

be a suggestion. This can be obtained by reviewing text with readability tools such as Flesch Kincaid[5].

Question 10 Should the requirement on certain relevant electronic services to respond to reports of class 1A and class 1B material on their service be limited to a requirement to take 'appropriate action'?

We would welcome more clarity on what the regulator would deem to be 'appropriate action' in order to answer the question. Again this should not be left open to interpretation, or that would again risk the 'race to the bottom'.

Question 11 What are your views on the likely compliance costs and, in particular, the impact of compliance costs on potential new entrants?

In our view all new entrants of all sizes should be required to comply with obligations, as per the answers above. Small niche sites can host very harmful content and provide even greater risks, by operating under the radar.

Where age appropriate design and age assurance is concerned, as the industry trade body and providers have confirmed, the costs are not prohibitive.

It is worth adding that costs will remain competitive where there is a healthy ecosystem of providers, driving innovation and competition. However, the marketplace of independent third party age providers will be disrupted if age checks are mandated to be undertaken at app store level. This is perhaps a topic worthy of discussion also by the relevant Competition & Consumer Commission and the Global Online Safety Regulators Network, hosted by the eSafety Commissioner.

Question 12 Is there any additional information eSafety should consider in determining the Relevant Electronic Services Standard?

It would be helpful for both public and industry transparency for the regulator to provide clarity as to the flexibility in terms of enforcement and blocking powers of the esafety Commissioner and to understand the engagement with the widest range of ancillary service providers - e.g. payment processing providers, ISPs, advertising networks. It may also be useful to consider at the Global Online Safety Regulators Network the work of various regulators in terms of engaging with ancillary service providers, for instance sharing the

[5] https://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid_readability_tests

experience of using dynamic blocking orders in terms of thwarting url hopping and illegal sports betting[6].

Question 13 Are the categories in Table 2 sufficiently clear for designated internet service providers to identify which category they fall within and therefore what obligations apply? What are the benefits and/or challenges of the categories as they are currently proposed?

The list of tiers as currently drafted do not seem to reflect the severity of the impact that content can have on individuals, rather it is confusing to understand the difference between *'pornography sites'* content and *'entertainment that would be classified at least R18+'*. It is also difficult to understand why those two are not grouped together.

This could lead to confusion on what tier various types of adult content would fall under, as pornographic content could be simultaneously deemed to be high impact and classified. We would welcome more clarity and explanation of table 2 - are these examples or an exhaustive list?

To give an example, a 'fan site' may badge itself as a social site for 18+ offering general entertainment. However a given performer or channel may offer explicit pornographic content. A global social media site may offer large volumes of adult content materials. At what stage of percentage of channels on a 'fan site' does that require re classification as a pornography site..

We would suggest the creation of a help tool, or dynamic decision tree to assist organisations to understand in which category they would fall. It would be helpful to road test this tool with industry bodies and civil society organisations.

Question 14 Are the section 9(5) matters in the draft Standard appropriate and sufficiently clear to help designated internet service providers accurately self-assess which tier their service falls within?

We would highlight that in order for firms to comply with the requirement to effectively and accurately assess the ages of end-users, they need to implement age assurance technologies. In our view self declaration or self attestation would not be sufficient for this purpose. There is overwhelming evidence that this cannot be relied on. This is mirrored in the incoming international standards, where any use case which would provide risks in terms of content, contact, conduct, contract should not rely upon self assertion.

---

6 chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://rm.coe.int/mapping-report-on-national-remedies-against-online-piracy-of-sports-co/1680a4e54c

Self declaration would only be appropriate for the lowest risk use case, such as a child selecting the level of maths problems on an edtech site, that they wish to engage in.

(g) safety by design guidance and tools published or made available by a government agency or a foreign or international body;

The regulator should set out or signpost companies to a list of tools and guidance documents which it deems acceptable and equivalent, in order to ensure that assessments are made consistently by all industry actors.

Question 15 eSafety is seeking to place requirements on service providers that are best-placed to prevent the use of generative AI features to create and disseminate class 1A and class 1B material. Does the proposal achieve this?

In terms of dissemination it would be worth engaging with a wide range of ancillary service providers and following the classic adage of 'follow the money' and including the requirements to extend to the full range of ancillary service providers to also include payment processors, ad networks, search engines to not support illegal activity on their platforms.

Payment providers such as MasterCard, VISA, PayPal have a reputation to uphold and have at their disposal the ability to require clarification of who is uploading content, obtaining consent and their age. See MasterCard AN5196[7] Revised Standards for New Specialty Merchant Registration Requirements for Adult Content Merchants.

According to TRM Lab's report titled 'Illicit Crypto Ecosystem Report'[8] and contrary to the thinking that CSAM commerce takes place using cryptocurrency, '*...true CSAM vendors seldom publicly promote their activity and continue to favor traditional finance channels*' (TRMlabs report)

To cite the recent high profile case in Spain where teenagers used the Nudify app to create deep fake nude images of teenage girls using generative AI[9], *'When ABC News reviewed the app, it offered a premium paid service that listed payment methods such as Visa, Mastercard, and Paypal. These payment methods, along with several others, were removed after ABC News reached out.'*

---

[7] https://segpay.com/MC_RevisedStandardsForNewSpecialtyMerchantRegistrationRequirementsForAdultContentMerchants.pdf

[8] https://www.trmlabs.com/report

[9] https://abcnews.go.com/US/mobile-apps-fueling-ai-generated-nudes-young-girls/story?id=103563734

In the instance of the Nudify app, it was openly listed on the Google Play marketplace. The regulator should engage with app marketplaces and search engines to require them to delist apps that allow the nudification of minors or the the distribution of non-consensual intimate imagery (NCII).

Payment providers and regulators could both require that Content Moderation companies switch on ongoing monitoring of all content to screen for consented content and review for any underage images, using AI age assessment tools to require instant notification when underage images are detected to review for CSAM.

The regulator should require ancillary service providers to undertake real time scanning for NCII, prohibit the upload of NCII on platforms and services, including the soliciting of NCII or advocating for the production or redistribution of intimate imagery without a victim's consent.

The international regulator community could consider a joint awareness campaign and tools to educate the public (both adults and minors) that creating AI generated CSAM or NCII is a crime and can lead to criminal prosecution.

Question 18 In relation to high impact generative AI designated internet services, do the proposed obligations (in particular, the section 21 obligation to 'detect and remove' and the section 23 obligation to 'disrupt and deter' child sexual abuse material and pro-terror material) provide appropriate safeguards? Are there specific challenges to deploying these measures in a generative AI context?

Again international co-operation is required to join forces. For instance, it would be useful if the Global Online Safety Regulators network could liaise with the proposed expert commission recommended by US Attorney Generals.

54 US Attorney General offices[10] have recommended that the US Congress should establish an expert commission and ongoing working group to study the means and methods of AI that can be used to exploit children specifically and to propose solutions to deter and address such exploitation.

The regulator should consider expanding existing restrictions on CSAM to explicitly cover AI-generated CSAM.

---

10 https://www.naag.org/wp-content/uploads/2023/09/54-State-AGs-Urge-Study-of-AI-and-Harmful-Impacts-on-Children.pdf

The regulator should consider requiring built-in filters that bar the creation of CSAM. A number of tools do have built-in features to bar the creation of pornographic content e.g. Midjourney, DALL-E, and Adobe Firefly.

In terms of applying appropriate safeguards with regards the use of AI tools, it is worth the regulator considering the actions of the Italian government which is requiring age assurance for access to the use of AI tools,

The Italian Data Protection Authority (GPDP) , required OpenAI to set up a mechanism in place to stop underage users accessing the service, which in its view "exposed minors to absolutely unsuitable answers with respect to to their degree of development and self-awareness" .

To quote the INHOPE Foundation, '*Children must be made aware that creating self-generated CSAM is not only illegal but can have devastating consequences for their safety and mental health*.'

It is worth the regulator having clear information gathering tools to require transparency as to where training data comes from and how an AI tool is delivering information to its users. The Global Online Safety Regulators Network and data protection regulators network may wish to share information on these topics.

Question 19 In relation to machine learning model platform services, do the proposed obligations (in particular, the section 23 obligation to 'disrupt and deter') provide appropriate safeguards? Are there specific challenges to deploying these measures?

There is a longstanding debate globally as to the relative rights of citizens and the relative impact of harms. No doubt, in the same vein as with E2E encryption and age assurance, there will be those that take the view that adults' rights to privacy outweigh childrens' rights to protection.

Question 21 Do sections 16 to 19 effectively reflect the considerations on minimum compliance measures outlined on pages 10-11 of this discussion paper?

This begs the question, what is the process for new approaches to be reviewed and added to the 'minimum requirements list'?

For instance, other example of 'safety technologies' for prevention of revenge pon / non consensual image abuse (NCII) of adults to ensure that all adults in the images or videos

are the correct people, are aged over 18+ and that each one has consented to be in the content on user to user platforms:

- E signature to gain consent plus age assurance from each person in the content

An example of technology for detection by content moderation companies

- The use of AI facial age estimation to assess age of each user in the content
- The option for face matching to recognise consented images, and maintain self exclusion /watch lists or bar lists

Where there are no faces detected, then it may not be possible to link consent to an individual, unless there are other unique identifying features.

Question 23 Is the technical feasibility exception in the obligation to detect and remove known child sexual abuse material and pro-terror material appropriate? How effective will this obligation be with this exception?

Innovation and investment in implementing approaches will only occur where there is a minimum floor and clear regime of monitoring and enforcement from the regulator.

www.yoti.com