5RIGHTS FOUNDATION

Ofcom: Protecting people from illegal harms online

5Rights Foundation, consultation response

February 2023

Overview

5Rights acknowledges the scale of the task in preparing for the enforcement of the UK's new online safety regime and welcomes Ofcom's efforts in delivering the draft illegal harms code of practice so soon after Royal Assent.

Ofcom has clearly sought to understand many of the risks associated with illegal harms and mitigation methods currently in use to take down illegal content. Some thought has also been given to how default settings can prevent children from becoming victimised online.

The proposals however fall far short of what is needed to deliver on the promises of the Online Safety Act, and demonstrate a highly concerning lack of consideration and alignment with existing UK regulation and best practice as concerns children, as well as with legislative intent. In particular:

1. The proposals give far greater consideration to the interests and costs to business than the costs to the many victims who have come to harm because of the commercial imperatives of tech companies. While the Act requires regulated services take a "proportionate" approach to fulfilling their duties, Ofcom is also required to look at the severity of harm. This one-sided accounting is completely out of spirit with legislative intent, which as stated in Section 1 of the Act is to, "make the use of internet services regulated… safer for individuals in the United Kingdom".[1] With regard to children, is also runs contrary to international law requirements to ensure the "best interests of the child" are a primary consideration in all actions concerning them.[2]
2. Ofcom has also placed undue focus on the size of services rather than their risk, creating a regime that will exempt many services from comprehensive duties. Small is not safe, and companies with 7 million users are not large –they are behemoths. We would argue that any company with more than 2 million UK users is large. We suggest that Ofcom carry out some polling of the general public, since we believe that most of the public did not imagine that companies such as Fortnite and Roblox could potentially be out of scope of the regulation. Ministers also gave assurances that the size of a service should not be used a reason for them to have fewer duties. As Lord Minister Parkinson of Whitley Bay said: "a small platform that is a font of illegal content cannot use the excuse of its size as an excuse for not dealing with it."

---

[1] Online Safety Act 2023, https://www.legislation.gov.uk/ukpga/2023/50/enacted

[2] UN Convention on the Rights of the Child (1989), https://www.unicef.org.uk/what-we-do/un-convention-child-rights

3. Ofcom's risk register posits that for the majority of illegal offences in scope of the legislation – including grooming, encouraging suicide and harassment, stalking, threats and abuse offences – the business model is not a risk factor, but various functionalities including recommender systems are. There is a wealth of evidence that functionality designed to keep attention is intrinsically connected to the business model. Exempting business models from scrutiny effectively legitimises commercial practices that are known to create risk and harm, once again completely out of spirit with legislative intent. As set out by Lord Minister Parkinson of Whitley Bay: "Obligations on services extend to the design and operation of the service. These obligations ensure that the consideration of risks associated with the business model of a service is a fundamental aspect of the Bill."[3]
4. Considering the risk associated with a product, feature or functionality before it has been introduced and mitigating harm ahead of its introduction is the norm in most other sectors, and a fundamental principle of safety by design, as required for services already under the UK's Age Appropriate Design Code.[4] Yet Ofcom has chosen to only require this kind of ex-ante assessment for the largest services or as a secondary measure in the risk assessment proposals. Risk assessments must assess risk. And new Codes that fall below the bar of existing Codes not only fail to add value, but risk creating confusion and the watering down of established best practices.
5. We are very concerned that the draft code of practice overly focuses on ex post facto measures rather than outcomes-based standards which would promote safety by design and encourage innovation in safety. 5Rights considers that measures should be as far as possible expressed in processes that iterate until the goal has been reached, thereby driving creative solutions and innovations, while furthering investments in online safety. Again, the intent of the legislation is to promote safety by design, and again, this up-stream approach is central to existing law and best practice for children, from the UN Convention on the Rights of the Child General comment No. 25[5] to the Age Appropriate Design Code.[6]
6. Ofcom's proposals interpret online safety 'measures' as tools rather than systems and processes, which could mean that companies are judged compliant with the regulation even when the desired outcome has not been

---

[3] Lord Parkinson of Whitley Bay, 19 July 2023, https://hansard.parliament.uk/Lords/2023-07-19/debates/63B4EB59-CF63-4E1D-8C6E-6D1901175AE1/OnlineSafetyBillhighlight=%22business+model%22#contribution-BA7FE81A-F30A-4D93-BEC7-CE7B8FFD2525

[4] ICO (ND) Introduction to the Children's Code, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code

[5] UN Convention on the Rights of the Child (1989), General Comment No.25 on children's rights in relation to the digital environment (2021), https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

[6] ICO (ND) Introduction to the Children's Code, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code

achieved. In the context of established international legal frameworks[7][8], existing regulation on children's privacy[9], best practice[10] as well as the substantial body of independent expert opinion, there is no justification for this interpretation. No environment is entirely risk free, but a safety by design approach will ensure platforms and services are equipped to identify and mitigate risk in their systems and processes.

7. On child user measures, we are confused as to why these would only be required of services with a high risk of grooming. Ofcom has noted in its draft risk register that children and young people are uniquely at risk of many of the illegal offences in scope of the regulation. Default settings should be applicable for all offences, and these should be alignment with hard-won existing regulation in the Age Appropriate Design Code.
8. We are deeply concerned that Ofcom's fixation on technical evidence – which tech companies' control – rather than the evidence of outcomes, could lead to online safety regressing in the UK. There is substantial and valid evidence which sets out the full scale of harm and risk online, especially to children, including 5Rights research Pathways[11], Disrupted Childhood[12], Risky by Design[13], in addition to the testimony of children themselves.

Overall, we are very concerned that the intersection of how Ofcom has interpreted 'measures', 'evidence', 'proportionality' and 'risk' in these proposals will contribute to a regime that will slow down safety measures rather than drive innovation in safety standards. This is counter to the stated aims of this legislation. It is critical that this approach be substantially reviewed and refocused on corporate responsibilities for up-stream due diligence and outcomes that do not cause harm, especially to children, in line with established law and best practice.

Consultation response

Volume 2: The causes and impacts of online harm

---

[7] General comment no. 25 on children's rights in relation to the digital environment (2021) https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

[8] OECD (2022) *Companion Document to the OECD Recommendation on Children in the Digital Environment,* OECD Publishing, Paris, https://doi.org/10.1787/a2ebec7c-en

[9] ICO (ND) Introduction to the Children's Code, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code

[10] IEEE Std. 2089-2021 (2021) IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, https://5rightsfoundation.com/static/ieee-2089-2021.pdf

[11] 5Rights Foundation (2021) *Pathways: How digital design puts children at risk,* https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf

[12] 5Rights Foundation (2023) *Disrupted Childhood: The cost of persuasive design,* https://5rightsfoundation.com/in-action/disrupted-childhood-the-cost-of-persuasive-design-2023.html

[13] 5Rights Foundation (2020) *Risky-by-Design,* https://www.riskyby.design/introduction

Question 1: Do you have any comments on Ofcom's assessment of the causes and impacts of online harms? Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.

Ofcom has undertaken considerable investigation into the causes and impact of illegal harms online, which 5Rights welcomes. However, we note there are gaps in the research which should be reflected in the first iteration of this work to ensure services can undertake thorough risk assessments and comply with the code of practice.

Safety by design: functionality and the business model

The draft risk register largely correctly identifies functionality that can impact the level of risk associated with becoming victim to illegal harms, but is inconsistent with regard to how business or revenue model poses a risk of the same harm. This is despite the fact that there is very substantial evidence that the two are intrinsically linked.

In the case of Grooming (6C), for example, while the analysis correctly identifies that network recommender systems can help facilitate the risk of grooming it does not identify that the revenue model can pose a risk. In fact, the analysis suggests "no evidence was found suggesting that revenue models are a risk factor in the facilitation and commission of these offences". This claim is also posited for:

- 6B. Terrorism offences ("There is limited evidence on how the different revenue models may affect the risks of harm, related to terrorism")
- 6D. Encouraging or assisting suicide (or attempted suicide) or serious self-harm offences
- 6E. Harassment, stalking, threats and abuse offences
- 6G. Controlling or Coercive Behaviour (CCB)
- 6H. Drugs and psychoactive substances offences
- 6I. Firearms and other weapons offences
- 6J. Unlawful immigration and human trafficking offences "very little evidence"
- 6M. Intimate Image Abuse
- 6N. Proceeds of Crime offences
- 6Q. False Communications Offence
- 6R. Epilepsy trolling offence
- 6S. Cyberflashing offence

Much of the functionality listed in the risk register is the product of persuasive design strategies of tech companies whose primary aim is to keep users on the service, which has been documented in detail by a wide range of research, including for children in our report Disrupted Childhood: The cost of persuasive design (2023).[14] Most user-to-user services deal in the currency of personal data which is sold to advertisers for profit, and need users to spend more time on to extract or infer more information valuable to advertisers who then use that data to target groups and individuals with products.

[14] 5Rights Foundation (2023) *Disrupted Childhood: The cost of persuasive design*, https://5rightsfoundation.com/uploads/Disrupted-Childhood-2023-v2.pdf

5Rights Pathways research[15] illustrated how business objectives drive the design features and functionality which can lead children to harm. For example, friend/follower suggestion functionality intended to reduce friction actually facilitates connection between children and adults, which presents a risk of grooming. Researchers found that many services did not have 'default privacy settings' which meant that child avatars were contacted by unknown adults within minutes of arriving on a service, and sent direct messages from strangers containing pornographic content. This functionality has been created to encourage connection and keep users using the service, and does not take into account the risk to children (or adults).

Research by Amnesty International[16] reinforced that data-driven, surveillance-based "reckless" business models have been exploited and used for profit by tech companies. The collection, storage and analysis of data means that advertisers are able to target users – including children – and guide them to more extremist content. The most profound example of this has been in Myanmar where Facebook's paid advertising features have been used to amplify mass-violence by spreading "dehumanising and inciting posts targeting the Rohingya."[17] Research by Tech Transparency Project[18] suggested that YouTube was profiting from white supremacist bands in its advertising, as well as creation of auto-generated "topic" channels – typically only created for artists with a "significant presence" – which in some cases could promote action such as incitement to violence.

Evidence from qualitative research[19], and testimony from former tech developers[20] and whistleblowers[21] tells us that most of the most common functionalities present on user-to-user services are designed to meet the commercial objective of those services. As designers themselves acknowledge, "reducing attention will reduce revenue."[22]

The investigation attached to the New Mexico Attorney-General case brought against Meta[23] in January 2024 demonstrated the true extent of how business models can facilitate harm. During the investigation, the profile of a fictional 13-year-old girl that had falsely represented her age to open her account was recommended, messaged by, and directed to perpetrators of CSAM/grooming despite the fact Meta's algorithm had

---

[15] 5Rights Foundation (2021) *Pathways: How digital design puts children at risk,* https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf
[16] Amnesty International (2023) *'I Feel Exposed': Caught in TikTok's Surveillance Web, https://www.amnesty.org/en/documents/POL40/7349/2023/en*
[17] Amnesty International (2022) *Myanmar: The social atrocity: Meta and the right to remedy for the Rohingya,* https://www.amnesty.org/en/documents/asa16/5933/2022/en
[18] Tech Transparency Project (2023) *Profiting from hate: Platforms' ad placement problem* https://www.techtransparencyproject.org/articles/profiting-from-hate-platforms-ad-placement-problem
[19] 5Rights Foundation (2021) *Pathways: How digital design puts children at risk,* https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf
[20] Embury-Dennis, T. (2017) *Man who invented 'Like' button deletes Facebook app over addiction fears.* The Independent, https://www.independent.co.uk/tech/facebook-like-inventor-deletes-app-iphone-justin-rosenstein-addiction-fears-a7986566.html
[21] Draft Online Safety Bill, oral evidence session, 25th October 2021, https://committees.parliament.uk/oralevidence/2884/html
[22] 5Rights Foundation (2021) *Pathways: How digital design puts children at risk,* https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf
[23] NM case 1:23-cv-01115-MIS-KK, *Attorney General v Meta Platforms,* Document 36-1 (2024) Case Studies 1 & 2: https://storage.courtlistener.com/recap/gov.uscourts.nmd.496039/gov.uscourts.nmd.496039.36.1.pdf

seemingly recognised her age and was targeting her with ads aimed at 13–16-year-olds. This runs contrary to Meta's own commitment to prevent CSAM on Facebook.[24,25] The account quickly amassed the maximum number of friends (5,000) and over 6,700 followers, of which the majority were adult males. Instead of protecting her safety, Meta suggested setting up a business account which would allow her to monetise her content, thus enhancing Meta's revenue – failing to interpret her audience and its network recommendations as a red flag.

Services amassing data for content recommender systems can lead users – particularly children – down a loop of illegal content. An investigation by Amnesty International[26] showed that, in one particular instance, it took only 67 seconds for TikTok to suggest content recommending posts related to anxiety, depression, self-harm and/or suicide.

In addition to the specific cases that have been noted in public, we point to the vast number of behavioral psychologists employed by the sector.[27] It is widely known that the three pillars of digital design are engagement, reach and time spent, and that technologists and psychologists are focused on those outcomes – for purely commercial reasons.[28] Most functionality is designed to these three pillars. Persuasive design strategies are pervasive across all user-to-user services and many search.

How the business model of services influences risk was also discussed at length during to passage of the Act. Lord Minister Parkinson of Whitley Bay set out that the consideration of risk associated with the business model was a fundamental aspect of the regulation.

> "My Lords, this is not just a content Bill. The Government have always been clear that **the way in which a service is designed and operated, including its features and functionalities, can have a significant impact on the risk of harm to a user.** That is why the Bill already explicitly requires providers to ensure their services are safe by design and to address the risks that arise from features and functionalities.
>
> "The Government have recognised the concerns which noble Lords have voiced throughout our scrutiny of the Bill, and those which predated the scrutiny of it. We have tabled a number of amendments to make it even more explicit that these elements are covered by the Bill. We have tabled the new introductory Clause 1, which makes it clear that duties on providers are aimed at ensuring that services are safe by design. **It also highlights that obligations on services extend to the design and operation of the service.** These obligations ensure that

---

[24] NM case 1:23-cv-01115-MIS-KK, *Attorney General v Meta Platforms,* Document 36-1 (2024) 115, Figures 23-27, https://storage.courtlistener.com/recap/gov.uscourts.nmd.496039/gov.uscourts.nmd.496039.36.1.pdf
[25] Facebook (2021) *Preventing child exploitation on our apps,* https://about.fb.com/news/2021/02/preventing-child-exploitation-on-our-apps
[26] Amnesty International (2023) *Driven into Darkness: How TikTok's 'For You' feed encourages self-harm and suicidal ideation,* https://www.amnesty.org/en/documents/POL40/7350/2023/en/
[27] Dunckley, V. L. (2018) *How the tech industry uses psychology hook children.* Psychology Today, *https://www.psychologytoday.com/us/blog/mental-wealth/201810/how-the-tech-industry-uses-psychology-hook-children*
[28] 5Rights Foundation (2021) *Pathways: How digital design puts children at risk,* https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf

> the consideration of risks associated with the business model of a service is a fundamental aspect of the Bill."[29] (Lord Parkinson of Whitley Bay, 19 July 2023)

Given the intrinsic connection between these functionalities and the business model, it is inaccurate and highly problematic to suggest there is no connection. Doing so will impact the basic usefulness of risk assessments undertaken using this risk register, as services will only be asked to consider the potential risk of the actions of bad actors and not how services can create risk. Indeed, Ofcom's analysis is behind the curve to such a degree that it will likely hamper existing safeguards that include engagement, reach and time spent metrics as a matter of course.

Question 2: Do you have any views about our interpretation of the links between risk factors and different kinds of illegal harm? Please provide evidence to support your answer.

5Rights welcomes the work Ofcom has undertaken seeking to establish the links between risk and illegal offences online. While we believe this is thorough in some places, we wish to raise the following gaps in Ofcom's thinking for which there is evidence.

6B. Terrorism offences

Ofcom has failed to recognise that age is a risk factor with regard terrorism offences. This is despite evidence of cases which demonstrate how young people are at particular risk of radicalisation online.

Due to their still developing cognitive abilities, including their emotional regulation and moral development, children are uniquely at risk of radicalisation online. According to official data, 14% of those arrested for terrorism offences in year ending 31st March 2023 were 17 and under.[30] Indeed, the UK Government's guidance - *Understanding and identifying radicalisation risk in your education setting* - identifies online radicalisation as a growing risk to children.[31]

We share the concerns of Jonathan Hall KC, the government's independent reviewer of terrorism legislation, who has raised concerns of this gap in the risk register.[32]

6C. Grooming, 6.F. Hate offences, 6.E harassment

Despite much discussion on this technology during the passage of the Act, there is scant mention of virtual reality spaces in this consultation. We note that it is only

---

[29] Lord Parkinson of Whitley Bay, 19 July 2023, https://hansard.parliament.uk/lords/2023-07-19/debates/63B4EB59-CF63-4E1D-8C6E-6D1901175AE1/OnlineSafetyBill#contribution-BA7FE81A-F30A-4D93-BEC7-CE7B8FFD2525
[30] Home Office (2023) *National statistics: Operation of police powers under the Terrorism Act 2000 and subsequent legislation,* https://www.gov.uk/government/statistics/operation-of-police-powers-under-the-terrorism-act-2000-quarterly-update-to-march-2023/operation-of-police-powers-under-the-terrorism-act-2000-and-subsequent-legislation-arrests-outcomes-and-stop-and-search-great-britain-quarterly-u#:~:text=Among%20those%20who%20were%20arrested,those%20in%20older%20age%2Dgroups
[31] Department for Education (2023) *Understanding and identifying radicalisation risk in your education setting,* https://www.gov.uk/government/publications/the-prevent-duty-safeguarding-learners-vulnerable-to-radicalisation/understanding-and-identifying-radicalisation-risk-in-your-education-setting
[32] Bentham, M. (2024) *Online safety rules have 'blind spot' to radicalisation of children, terror watchdog warns,* The Standard, https://www.standard.co.uk/news/uk/ofcom-child-safety-rules-online-bill-radicalisation-terror-b1134556.html

included in the risk assessment if the service is a gaming service, which is not the only type of service where virtual reality spaces can be found. The metaverse is not a game, but a digital world that encompasses many different features and services.[33]

While many of the risks to children in particular in the metaverse are the same as those found in other digital spaces, such as social media or games, there are certain risks that are unique to or exacerbated in these virtual worlds. Heightened sensory experience through the use of haptic technologies (beyond text and image interaction) creates a different sensorial experience with the potential to intensify feelings of emotional and physical distress.[34] If a child's avatar is physically assaulted or if a stranger whispers into their avatar's ear this will have a more direct impact on the child as the interaction is immersive and personal. Psychotherapist Nina Jane Patel logged into Horizon Venues metaverse which was launched by Facebook in 2020 and created her avatar. "Within 60 seconds of joining — I was verbally and sexually harassed — 3–4 male avatars, with male voices, essentially, but virtually gang-raped my avatar and took photos — as I tried to get away they yelled — 'don't pretend you didn't love it'." She has since documented having panic attacks and mental health consequences as a result of this assault.[35]

During the passage of the Online Safety Act, a number of parliamentarians raised the question of whether the metaverse would be within scope of the Act, and were reassured by Lord Minister Parkinson of Whitley Bay that it would be:

> "The metaverse is in scope of the Bill, which, as noble Lords know, has been designed to be technology neutral and future-proofed to ensure that it keeps pace with emerging technologies—we have indeed come a long way since the noble Lord, Lord Clement-Jones, the noble Lords opposite and many others sat on the pre-legislative scrutiny committee for the Bill. Even as we debate, we envisage future technologies that may come. But the metaverse is in scope."[36] (Lord Parkinson of Whitley Bay, 12 July 2023).
>
> "The noble Baroness, Lady Finlay, raised important questions about avatars and virtual characters. The Bill broadly defines "content" as "anything communicated by means of an internet service", meaning that it already captures the various ways through which users may encounter content. In the metaverse, this could therefore include things such as avatars or characters created by users. As part of the user-to-user services' risk assessments, providers will be required to consider more than the risk in relation to user-generated content, including aspects such as how the design and operation of their services, including functionality and how the service is used, might increase the risk of harm to children and the presence of illegal content. A user-

---

[33] Ball, M. (2020) *The Metaverse: What it is, where to find it, and who will build it,* MatthewBall.co, https://www.matthewball.vc/all/themetaverse
[34] Lavoie, R. et al. (2021) *Virtual experience, real consequences: the potential negative emotional consequences of virtual reality gameplay*. Virtual Reality. Vol. 25. https://link.springer.com/article/10.1007/s10055-020-00440-y
[35] Patel, N. J. (2021) *Reality or Fiction?* Medium, https://medium.com/kabuni/fiction-vs-non-fiction-98aa0098f3b0
[36] Lord Parkinson of Whitley Bay, 12 July 2023, https://hansard.parliament.uk/Lords/2023-07-12/debates/166C8F0B-D314-4AF6-A7FB-341766931E1F/OnlineSafetyBill?highlight=metaverse#contribution-5F4BD7E7-3D26-4B6A-B81D-AA868B234646

> to-user service will need to consider any feature which enables interaction of any description between users of the service when carrying out its risk assessments."[37] (Lord Parkinson of Whitley Bay, 17 July 2023).

Moreover, the overarching requirement in section 1 of the Act that services should be safe by design is not reflected by leaving out virtual worlds. We are concerned by the omission, and that Ofcom is focusing on the model of services currently operating rather being tech-neutral and driving design change to benefit users, particularly children.

Volume 3: How should services assess the risk of online harms? Governance and accountability

Question 3: Do you agree with our proposals in relation to governance and accountability measures in the illegal content Codes of Practice? Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.

Greater internal accountability and stronger governance will be central to the success of the online safety regime in the UK and 5Rights supports Ofcom's commitment to focusing on this measure in the first months of the illegal content Codes of Practice coming into force.

Accountability throughout the workforce

We support proposals which recommend having staff in place who are responsible for reporting on compliance to senior boards, written statements of responsibilities for senior members of staff, training for staff involved in design and operations on compliance, and codes of conduct for staff members.

However, while it is true that accountability placed on senior staff is central in ensuring compliance, design decisions that impact the safety of children should not be the sole responsibility of top management. It is important that standards for safety are implemented at all levels of the organisation, to ensure it is thoroughly understood and adopted throughout the organisation.

In the Institute of Electrical and Electronics Engineers Standards Association (IEEE SA) Standard P2089 on the Age Appropriate Digital Services Framework[38] a number of roles are identified which are essential to delivering a safe by design service. While this relates to age appropriate design, the research and expertise which underpins it speaks to the objectives of the illegal harms duties in the Online Safety Act and we would encourage Ofcom to consider recommending these roles be required as best practice for compliance.

---

[37] Lord Parkinson of Whitley Bay, 17 July 2023, https://hansard.parliament.uk/Lords/2023-07-17/debates/1F1A09C2-293E-42A1-8DD5-D4C114971FEF/OnlineSafetyBill?highlight=metaverse#contribution-B62F2561-59A1-42CC-93D7-0852436DE35B

[38] IEEE Std. 2089-2021 (2021) IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, https://5rightsfoundation.com/static/ieee-2089-2021.pdf

At paragraph 8.185, Ofcom recommends: “services which identify considerable risk to users will require more checks and balances in place to ensure that they are effectively managing and mitigating identified risks. This includes aligning their staff policies with their approach to risk management, to ensure that employees across an organisation are aware of a service’s duty to manage illegal content effectively on an on-going basis. These services will also likely benefit from communicating expectations around the importance of managing these risks to all staff.” It would be appropriate and proportionate to recommend aligning staff policies with the approach to risk management for all services, not just large or multi-risk services.

Tracking illegal content and reporting to senior staff

We support proposals for the tracking of evidence of new kinds of illegal content on services, and unusual increases in particular kinds of illegal content and for this to be reported to be senior management. However, this measure does not include an obligation to take it to the board, or for the board to take action upon receiving such a report – this should be the case. To date tech companies have been known to have harbored internal research and information regarding the harms on their services which they have not acted upon.[39] This measure could be strengthened by making clear that where risks are identified a new risk assessment must be undertaken and mitigation put in place.

We also have an overarching and critical concern that Ofcom has interpreted 'measures' as tools rather than systems and processes – this may mean that companies are considered compliant even when the service is risky by design and the desired outcome has not been achieved. In this instance, and more generally, Ofcom should put processes that are iterative and have outcomes so that the measure or mitigation is not deemed adequate until the outcome has been achieved. This would be in line with government assurances from the dispatch box over a five-year period which promised a systems and processes regime.

Question 4: Do you agree with the types of services that we propose the governance and accountability measures should apply to? Please explain your answer.

Online safety measures should be proportionate to the risks present on the service, including subject and functionality, irrespective of size, intent or variables such as potential cost. This consultation places excessive focus on the costs to services, and not the severity of the harm. This will exacerbate and encourage companies to continue to privatise the benefits of their service and outsource the costs to the public purse, for example, health, justice and education services. It will also give further competitive advantage to the large service who are much more able to absorb the costs. This is particularly concerning in the context of content and activities that have been deemed illegal.

Costs

[39] Gayle, D. (2021) *Facebook aware of Instagram’s harmful effect on teenage girls, leak reveals,* The Guardian, https://www.theguardian.com/technology/2021/sep/14/facebook-aware-instagram-harmful-effect-teenage-girls-leak-reveals

We are concerned that Ofcom places an undue focus on the cost to the services complying with the measures, and not the cost to the potential victims of offences and illegal content the service causes. This focus runs counter to the stated objectives of the Act – to see services made safer by design – (Section 1)[40] and ignores the fact that regulation can in fact support fledgling companies as they look to scale-up and become commercially viable.

For example, the submission from the coalition of civil society organisations on Violence Against Women and Girls (VAWG) sets out the societal and economic impact that services which facilitate access to harmful content and offences against women and girls. While any regulator is duty bound to be 'proportionate' it seems that the current draft does not take a victim centric approach and may in certain ways provide a regressive regime in which companies that already do more pull back their investments and those that do nothing make the minimum investment. As stated elsewhere in this response, 5Rights considers that measures should be as far as possible expressed in processes that iterate until the goal has been reached, thereby driving creative solutions and innovations, while furthering investments in online safety.

Size and functionality

We are concerned that the consultation on the whole neglects to address how the regulation will deal with small high-risk services. While user-to-user services with large user bases and high functionality can lead to fast dissemination of harmful content, smaller platforms with fewer users and less functionality are not by definition less risky. Indeed, many of them may focus on harmful content.

There are a number of forums dedicated to single issue themes included in the illegal harms and offences, which we do not believe this code of practice will adequately cover. While one forum relating to suicide has been blocked by Ofcom this service now provides advice for UK users on how to circumvent the ban.[41] The current draft proposals suggest that Ofcom would not be able to act as firmly on others.

During the passage of the Online Safety Act, parliamentarians spoke on the floor of the House of the Lords on this issue, and were reassured by the Lord Minister that they would be included:

> "I want to be clear that a small platform that is a font of illegal content cannot use the excuse of its size as an excuse for not dealing with it."[42] (Lord Parkinson of Whitley Bay, 19 July 2023)

Given the offences in scope of this part of the regulation Ofcom has failed to reflect the Minister's promise nor is it in line with other sectors. For example, Food safety regulations apply to any business which "deals in food", which could include businesses from hot dog vans to a five-star restaurant.[43] Similarly, all employers, regardless of the

---

[40] Online Safety Act 2023, https://www.legislation.gov.uk/ukpga/2023/50/enacted
[41] Smith, T & Crawford, A. (2023) *Suicide forum blocked to most UK users after Ofcom pressure,* BBC News, https://www.bbc.co.uk/news/uk-67374129
[42] Lord Parkinson of Whitley Bay, 19 July 2023, https://hansard.parliament.uk/lords/2023-07-19/debates/63B4EB59-CF63-4E1D-8C6E-6D1901175AE1/OnlineSafetyBill#contribution-DF692200-402C-4DB9-B156-17495CD0E59D
[43] Gov.UK (ND) *Food safety - your responsibilities, https://www.gov.uk/food-safety-your-responsibilities*

size or nature of the workplace, have a duty to protect the health, safety and welfare of employees under the Health and Safety at Work Act.[44] It should follow that where is risk of illegal harm or harm to children, online safety regulation applies to any and all online services, irrespective of nature or size.

The final drafting of the illegal offences code must deal with smaller high-risk services which host illegal content and facilitate illegal activity.

Question 5: Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to requiring services to have measures to mitigate and manage illegal content risks audited by an independent third-party?

Third-party auditing of measures to address illegal content should be a part of the first iteration of this regulation.

The tech industry has a poor record of transparency with regard to known risks on their services. In 2021, Facebook whistleblower Frances Haugen gave testimony to the US Congress on how the company sat on internal research that their service was actively harming teenage girls. Her written testimony included: "I saw that Facebook repeatedly encountered conflicts between its own profits and our safety. Facebook consistently resolved those conflicts in favor of its own profits. The result has been a system that amplifies division, extremism, and polarization — and undermining societies around the world. In some cases, this dangerous online talk has led to actual violence that harms and even kills people. In other cases, their profit optimizing machine is generating self-harm and self-hate — especially for vulnerable groups, like teenage girls. These problems have been confirmed repeatedly by Facebook's own internal research."[45] This followed reports from the Wall Street Journal on leaked Facebook internal research which found that Instagram, "make[s] body image issues worse for one in three teen girls" and "thirty-two per cent of teen girls said that when they felt bad about their bodies, Instagram made them feel worse".[46]

The lack of third-party auditing in other regulated industries also points to the need for this measure. Recent reporting into the water industry has found that the self-monitoring system in place – operator self-monitoring – may have led to illegal levels of toxic pollutants being released into rivers.[47]

As part of this measure, it will be important for the regulator to set out what is the bar of illegality, so that expectations and common standards are understood. Another whistleblower Arturo Béjar revealed that Meta was, "creating its own homework" and down-playing the prevalence of harmful content - because they only reported on their

---

[44] Health and Safety at Work etc. Act 1974, https://www.legislation.gov.uk/ukpga/1974/37/contents
[45] Haugen, F. (2021) Statement to United States Sub-Committee on Consumer Protection, Product Safety, and Data Security, 4th October 2021, https://www.commerce.senate.gov/services/files/FC8A558E-824E-4914-BEDB-3A7B1190BD49
[46] Wells, G. et al. (2021) *The Facebook files: Facebook knows Instagram is toxic for teen girls, company documents show,* Wall Street Journal, https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739
[47] Blakely, R. (2023) *Illegal river pollution goes unspotted under flawed testing,* The Times, https://www.thetimes.co.uk/article/pollution-breaches-missed-under-flawed-monitoring-by-water-companies-clean-it-up-zq66cd7g0

own definitions.[48] The purpose of having a regulatory regime is to bring to an end self-regulation that has damaged society and the health and wellbeing. In the proposal as it currently stands, Ofcom has not fully met the severity of the harms.

Given the severity of the harms this consultation discusses, independently assessed auditing of illegal content risks would be appropriate.

Question 6: Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to tie remuneration for senior managers to positive online safety outcomes?

This is a corporate responsibility not an individual one. The way to drive a safety by design model is to set out process measures that have clear outcomes and allow companies to iterate to those outcomes.

Service's risk assessment

Question 7: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.

The risk assessment proposals place too much onus on the proportionality and cost to services rather than focusing on supporting services to carry out holistic and robust assessments. This runs counter to the legislative aims of the Act – that "services regulated by the Act are safe by design" (Section 1(3))[49] - and risks failing to support regulated services to adequately understand how to prevent illegal harm and content being created, in addition to removing it.

Overall, in identifying risk, Ofcom focusses excessively on ex post facto assessment (user reports), whereas ex ante assessment (product testing, consultation with experts – the basics of safety by design) are primarily reserved for larger services or as a secondary / additional assessment.

The risk assessment also does not ask services to look at the context of the system or service and its purpose and instead is asking services to look at risks in isolation. This will not lead to holistic assessments nor does it fulfil the cumulative harm that has been widely evidenced and often promised. During the passage of the Bill, the government specifically rejected the accepted model used by EU and OECD of the four C's framework of risk.[50] In order to fulfil the promise of a systems and process regime, it is incumbent on Ofcom to ensure that they take full account of the full range of risks to which children are exposed. The illegal harms risk assessment would benefit from these risks being explicitly referenced as children can experience harm differently to adults.

Proportionality

---

[48] Hendrix, J. (2023) *Transcript: Senate hearing on social media and teen mental health with former Facebook engineer Arturo Bejar,* Tech Policy Press, https://www.techpolicy.press/transcript-senate-hearing-on-social-media-and-teen-mental-health-with-former-facebook-engineer-arturo-bejar

[49] Online Safety Act 2023, https://www.legislation.gov.uk/ukpga/2023/50/enacted

[50] Lord Parkinson of Whitley Bay, 27 April 2023, https://hansard.parliament.uk/Lords/2023-04-27/debates/958CAC63-A345-45E8-9DE3-7CBA46611DCA/OnlineSafetyBill#highlight=%22contract%22#contribution-F60F2353-5744-4DCA-91FF-4AB21F061606

As previously raised, Ofcom has placed too much onus on the cost to services of complying with the regulation rather than the cost of the outcome of harm to victims and victims families, or third parties such as schools, NHS or Police that then pick up the pieces. While the Act requires regulated services take a "proportionate" approach to fulfilling their duties, and indeed requires Ofcom to look at resources, Ofcom is also required – among other issues – to look at the severity of harm. The one-sided accounting is out of spirit with the legislative intent.

Draft Risk Profiles

Ofcom should ensure that services be prepared to demonstrate that they considered their own internal data and knowledge with regards to risks on their services, in addition to the draft risk profiles. The risk profiles will struggle to keep up with emerging risks if not updated regularly (for example, generative AI and risk associated with its use to produce child sexual abuse material (CSAM) which Ofcom has chosen not to include, is a growing and serious risk)[51].

Annex 5 notes that Ofcom has not included generative AI in the register of risks or risk profiles as the evidence is still developing. Given the severity of the risk and harm which is already known[52,53,54], particularly with regard to its use in creating CSAM, it is imperative that Ofcom commits to collecting the requisite data to add this to the register as a matter of urgency.

Draft risk assessment guidance

We are concerned that the approach to risk assessment is too focused on ex post facto measures and not ex ante risk assessment, which would help services to be safer by design, as per the aims of the Act.

Evidence inputs

To robustly consider the risk of illegal harm and content, services must be guided to holistically assess their risk – both current and potential. We are concerned that the listed *core evidence inputs*, which can serve as the only evidence considered, do not adopt a thorough or ex ante approach. This approach will not incentivise or encourage services to carry out thoughtful and holistic assessment of their risks, but rather internalise business practice as it now is and then exempt many services from complying.

Given the history of whistleblowing, NGO exposés, court cases that reveal willfully careless practice – it does seem that Ofcom has been blind to the level to which companies will go to prevent change. This, coupled with Ofcom suggesting that there is

---

[51] Thiel, D. (2023) *Investigation finds AI image generation models trained on child abuse,* Stanford University Cyber Policy Center, https://cyber.fsi.stanford.edu/news/investigation-finds-ai-image-generation-models-trained-child-abuse#:~:text=An%20investigation%20found%20hundreds%20of,generated%20nude%20images%2C%20including%20CSAM

[52] Internet Watch Foundation (2023) *How AI is being abused to create child sexual abuse imagery, https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf*

[53] Badcock, J. (2023) *AI-generated naked images of dozens of Spanish girls shared around schools,* The Telegraph, https://www.telegraph.co.uk/world-news/2023/09/20/ai-generated-images-schoolgirls-deepfakes-spain

[54] de Guzman, C. & Henshall, W. (2024) *As tech CEOs are grilled over child safety online, AI is complicating the issue,* Time, https://time.com/6590470/csam-ai-tech-ceos

no evidence that the business model impacts on harm and the emphasis on preventing cost to the company, appears to suggest that Ofcom is seeing its role as managerial rather than seeking culture change.

Risk Profiles: While the risk register and risk profiles – *core evidence inputs* - are an important resource for services to understand the risk, and a legislative requirement, it should not serve as the only evidence required in light of the speed at which services, features and functionality can develop – and sometimes cause harm. The introduction of a new Snapchat feature, 'Speed Filter' for example was quickly connected to a number of car accidents.[55] Although it was subsequently removed by Snap, it serves as an example of how the industry does not systematically consider features and products ahead of their introduction and the harm this can potentially cause.

User complaints, including user reports: User complaints and reports may hold some relevant evidence but it is important to highlight that not all victims of offences online report its occurrence. The draft document would benefit from highlighting that many children and women, for example, who are particularly vulnerable to the offences and illegal content do not report to services. A UCL study on sexual violence online found that 51% of young people who had experienced unwanted sexual attention or had an image shared without their consent did not report this to the service. When asked why, they responded that they do not think reporting works.[56,57] Similarly, it is common for victims of child sexual abuse to not report abuse or grooming (online or offline) as they often feel ashamed, a fact that is highlighted by Ofcom in this consultation.[58]

It will also be necessary for Ofcom to create an 'inbox'. The Act would have benefited from an independent complaints mechanism, but in the absence of that, we have already found that parents cannot report on behalf of their child, or that companies receiving complaints are rejecting them even though they involve life threatening challenges. It will be necessary for Ofcom be aware of these cases since it otherwise risk becoming regulation by tragedy and headline.

Product testing

Considering the risk associated with a product, feature or functionality before it has been introduced, and mitigating harm ahead of its introduction, is the norm in most other sectors, and a fundamental principle of safety by design. No environment is entirely risk free, but a safety by design approach will ensure platforms and services are

---

[55] Godwin, C. (2021) *Snapchat removes controversial speed filter,* BBC News, https://www.bbc.co.uk/news/technology-57522146

[56] Ringrose, J. et al. (2021) *Understanding and combatting youth experiences of image-based sexual harassment and abuse*, UCL/School of Sexuality Education/University of Kent/Association of School and College Leaders, https://www.ascl.org.uk/ASCL/media/ASCL/Our%20view/Campaigns/Understanding-and-combatting-youth-experiences-of-image-based-sexual-harassment-and-abuse-full-report.pdf

[57] Internet Matters (2023) *"It's really easy to go down that path": Young people's experiences of online misogyny and image-based abuse*, https://www.internetmatters.org/wp-content/uploads/2023/09/Internet-Matters-Online-misogyny-and-image-based-abuse-report-Sep-2023-2.pdf

[58] Hamilton-Giachritsis, C. et al. (2017) *"Everybody deserves to be happy and safe": A mixed methods study exploring how online and offline child sexual abuse impact young people and how professionals respond to it*, NSPCC/University of Bath/University of Brimingham/CEOP, https://learning.nspcc.org.uk/media/1123/impact-online-offline-child-sexual-abuse.pdf

equipped to identify and mitigate risk. We are concerned that this does not appear in the *core evidence inputs.*

"Product testing" for instance is suggested as an *enhanced evidence input* for only the largest and riskiest services, and only suggested if the *core evidence inputs* have not provided clarity on the risks. Best practice from technical standards, including the IEEE Standard for an Age Appropriate Digital Services Framework[59], demonstrates that recommending product testing before a feature, functionality or service is launched would be a more effective and holistic approach to risk assessment. Similarly, "consultation with internal experts on risks and technical mitigations" is also suggested as an *enhanced evidence input* but would be appropriate for a *core evidence input* for the largest services. Ofcom's call for evidence received responses from Google, Yoti and Trustpilot who "highlighted the importance of product testing and signaled that this is common practice across the industry in services of a certain size." It would therefore not be disproportionate to include this as a required *core evidence input*.

Ofcom should think again and make certain that ex ante evidence forms a key part of the risk assessment. In particular, Ofcom should include 'product testing' or risk assessment before features are introduced, a *core evidence input* for risk assessments.

Risk assessing for child users

Ofcom states that "the purpose of the risk assessment is to improve your understanding of how harm could take place on your service and what safety measures you need to put in place to protect users, especially children" (Annex 5, p35). However, this risk assessment does not take into account the specific vulnerability of child users to illegal harms.

Services must assess the full range of harms children experience online many of which are created by the design features and commercial decisions of services, which are often the true drivers of harm. For example, it is often the pursuit of maximising user-engagement that fuels the amplification and targeting of harmful content to children.[60]

During the passage of the Act, the government specifically rejected the accepted model used by EU and OECD of the four C's framework of risk. In order to fulfil the promise of a systems and process regime, it is incumbent on Ofcom to ensure that they have covered all the risks that are covered by this normative framework. The illegal harms risk assessment would benefit from this being addressed as children can experience harm differently to adults.

Assessing risk requires consideration of several interlocking and interdependent factors. Services need to consider the complexity of risk when assessing their service and features. Approaches to risk assessment and mitigation should be based on "the best available information and scientific insights."[61]

---

[59] IEEE Std. 2089-2021 (2021) IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, https://5rightsfoundation.com/static/ieee-2089-2021.pdf
[60] 5Rights Foundation (2021) *Pathways: How digital design puts children at risk,* https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf
[61] Digital Services Act 2022/2065, Preamble, https://www.eu-digital-services-act.com/Digital_Services_Act_Preamble_81_to_90.html

In particular, the risk assessment should include consideration of:

- How certain features in combination may exacerbate risk
- How some risks are not immediately obvious but may create significant harm over time
- How children with different levels of vulnerability or resilience may respond
- How both individuals and groups experience risk

The IEEE framework previously mentioned, sets out in detail a good process for risk assessment. Ofcom should not allow anything lesser to become a norm.

Record keeping and review guidance

Question 10: Do you have any comments on our draft record keeping and review guidance? Please provide the underlying arguments and evidence that support your views.

Services should be expected to keep records of their compliance and risk assessment as part of the regime. It is good in principle that this should be clear and kept in a durable manner.

In their yearly reports, services should be required to specifically report against the harms to children and their prevalence set by association with regulatory standards not company standards

Question 11: Do you agree with our proposal not to exercise our power to exempt specified descriptions of services from the record keeping and review duty for the moment?

Yes.

It would be prudent for Ofcom to require this of all services in scope as the regulation comes into force.

Volume 4: What should services do to mitigate the risk of online harms

Question 12: Do you have any comments on our overarching approach to developing our illegal content Codes of Practice?

We are very concerned that the intersection of how Ofcom has interpreted 'measures' 'evidence' ‘proportionality’ and ‘risk’ - contribute to a regime that will slow down safety measures rather than drive innovation in safety standards.

See our full response at question 16.

Question 13: Do you agree that in general we should apply the most onerous measures in our Codes only to services which are large and/or medium or high risk? Please provide the underlying arguments and evidence that support your views

No.

The severity of risk should be the first consideration, followed by the scale of risk, and then how the risks intersect to create new risks. As previously raised, we are concerned that Ofcom has placed too much onus on the cost to services of complying with the

regulation rather than the cost of the outcome of harm to victims. While the Act requires regulated services take a "proportionate" approach to fulfilling their duties, and indeed requires Ofcom to look at resources, Ofcom is also required – among other issues – to look at the severity of harm.

We would urge Ofcom to focus on proportionality with regard to the severity and risk of harm to users, particularly children when considering what is appropriate for compliance. Unnecessary regulatory burden can be avoided by additional support, proportionate mitigation strategies and enforcement activity from the regulator, as set out in the Regulators' Code.

Question 14: Do you agree with our definition of large services? Please provide the underlying arguments and evidence that support your views.

No.

While we understand that the child safety duties will have to be undertaken by all companies, we were, as was our broader network, shocked by the definition that Ofcom has provided, particularly as it copies the EU DSA in this regard but not in ways in which it offers a stronger regime.

We suggest that Ofcom carry out some polling of the general public, since we believe, that most of the public did not imagine that companies such as Fortnite and Roblox could potentially be out of scope of the duties. In creating a regime that exempts so many services from comprehensive duties, Ofcom is in danger of creating a backward step in online safety. Small is not safe, and companies with 7 million users are not large, they are behemoths. We would argue that any company with more than 2 million UK users is large.

We would also like to see further detail given as to how regularly services need to review their user-base to determine when they have become a large service, and ensure they begin complying with extra measures as soon as possible.

Question 15: Do you agree with our definition of multi-risk services? Please provide the underlying arguments and evidence that support your views.

Attaching risk solely to particular offence disincentivises services from thinking holistically about how systems, features and functionality can lead users to harm online and create risk.

We would also raise the proliferation of small services which may be dedicated to one subject considered within the scope of the offences set out in the Act. For example, the IWF has seen growth in a number of websites commercialising the production of AI generated CSAM[62], and there are also examples of forums dedicated to suicide and self-harm ideation which have been linked to deaths.[63]

Our answer to Question 14 indicates that many services that should be in scope are not, by associating risk with particular offences rather than the outcome of a risk assessment further undermines the possibility of understanding how a company could effectively – often very simply – mitigate that risk.

---

[62] Internet Watch Foundation (2023) *How AI is being abused to create child sexual abuse imagery, https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf*

[63] Crawford, A. & Smith, T. (2023) *'Failure to act' on suicide website linked to 50 UK deaths*, BBC News, https:/www.bbc.co.uk/news/uk-67082224

Question 16: Do you have any comments on the draft Codes of Practice themselves?

Overall, we are very concerned that the code of practice overly focuses on ex post facto measures rather than outcome-based standards which would promote safety by design and encourage innovation in safety. 5Rights considers that measures should be as far as possible expressed in processes that iterate until the goal has been reached, thereby driving creative solutions and innovations, while furthering investments in online safety.

Measures which promote safety by design would be more effective at supporting services to comply with the illegal content duties, and delivering on the legislative aims of the Act.

Ex post facto approach

We are concerned that Ofcom has interpreted 'measures' as tools rather than systems and processes. This may mean that companies are compliant even when the desired outcome has not been achieved. In this instance, and more generally, Ofcom should set out processes that are iterative and have outcomes so that the measure or mitigation is not deemed adequate until the outcome has been achieved. This would be in line with government assurances from the dispatch box over a five-year period which promised a systems and processes regime.

The code recommends a number of measures which are predominantly ex post facto, for example content moderation, user reporting, and CSAM hash matching. By being too prescriptive, and not making the standards of the Code outcomes based, Ofcom risks disincentivising services to innovate in the name of safety to produce measures which are more technically effective and cost effective. This also risks the code becoming out of date quite quickly.

The Age Appropriate Design Code (AADC), which sets out how in scope services should treat children's data, is a mixture of outcomes-based and prescriptive standards. Since its inception, major technology companies, from Google to TikTok to Meta, have brought in hundreds of design changes – such as defaulting children's accounts private, turning off notifications and direct messaging and ensuring they have transparent - in order to meet the requirements. The AADC has encouraged tech companies to innovate their services in the name of safety, enhancing children's experiences online.

We are concerned that the lack of outcomes-based standards in the code is counter to the stated aims of the codes of practice. During the passage of the Act, Lord Minister Parkinson of Whitley Bay made clear that the Codes should be outcomes based, and should not be too prescriptive as this would impact smaller services ability to comply: "We must also recognise the diversity and innovative nature of this sector. Requiring compliance with prescriptive steps rather than outcomes may mean that companies do not use the most effective or efficient methods to protect children."[64]

---

[64] Lord Parkinson of Whitley Bay, 2 May 2023, https://hansard.parliament.uk/Lords/2023-05-02/debates/C4ADB2FF-C4AE-4BEA-8E30-A341ECF32822/OnlineSafetyBill?highlight=%22innovative%20nature%20of%20this%20sector%22#contribution-035C9CD3-85B6-468A-AC05-CA34E24320D5

Small services

We are concerned that many of the measures in the Code will only be required of the largest services with high or medium risk. As we have set out elsewhere, proportionality must be based on the severity of risk and not size or other variables such as the costs to companies of complying.

Age assurance

Regarding the child user measures, we are very concerned that Ofcom will not require these of services who do not currently undertake any age assurance, leaving many high-risk services out of scope. Ofcom must ensure there is coherence with the AADC which sets out a 'likely to be accessed' by children threshold test services are already required to meet if they let children onto their service.

Child user measures

We are confused as to why these measures will only be required of services with a risk of grooming, as volume 2 sets out how age is a factor in many of the offences in this part of the regulation. Ofcom should require default privacy settings for all services, as is required under the AADC, and look to expand this list as per the additional functionality set out in our response.

Question 17: Do you have any comments on the costs assumptions set out in Annex 14, which we used for calculating the costs of various measures?

The cost assumptions for human moderators do not appear to take into account that many services outsource these roles to other countries where salaries can be considerably lower.[65] Ofcom should consider this in its cost assumptions for those roles.

Content moderation (User to User)

Question 18: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.

Effective content moderation is an important component of tackling harm online and is already used widely in the sector. However, content moderation and reporting mechanisms are not always effective and there are a number of issues associated with these tools.

In principle, as an ex post fact measure, content moderation should not be prioritised over upstream measures to prevent illegal content and activity from be allowed to take place. Content moderation should also be underpinned by transparency and support for those needing to report content.

We would like to raise the following concerns:

Measure 1: At paragraph 12.80, Ofcom recommends: "Where the provider is satisfied that its terms and conditions for the service prohibit the types of illegal content defined in the Act which it has reason to suspect exist, consider whether the content is in

---

[65] Arsht, A. & Etcovitch, D. (2018) *The human cost of online content moderation,* Jolt Law (Harvard), https://jolt.law.harvard.edu/digest/the-human-cost-of-online-content-moderation

breach of those terms of service and, if it is, take the content down swiftly." Given the legislative requirement to take down illegal content, we are unsure why the service would only need to remove the content if its terms and conditions expressly detailed that it is not allowed. While services must uphold their terms and conditions, the service is still subject to the law which would override this.

Measure 5: Performance targets are widely used and can be useful but we would raise that internal performance targets can be skewed where services set their own standards. As set out in an earlier response, through whistleblowers we know that Meta would set its own standards for harmful content to play down its the prevalence.[66] Ofcom should set outcomes for what these should achieve.

Measure 6: We welcome the measure that moderators should have training but, as set out in our call for evidence response, we would urge Ofcom to include that this training includes how to identify risks to child safety, including knowledge of risks to different groups of children and the full range of content and activity that is illegal or might be harmful to a child. Human moderators are core to understanding risk to children.

Content moderation (Search)

Question 19: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views

Effective content moderation is an important ex post facto component of tackling harm online and is already used widely in the sector. However it should not be prioritised over up stream measures to prevent illegal content and activity from be allowed to take place. Effectively applied content moderation should serve as one tool amongst many for services to tackle a key measure for compliance but it is not always consistently applied across services and we are concerned that the draft code does not make reference to minimum standards of use.

Automated content moderation (User to User)

Question 20: Do you agree with our proposals? Do you have any views on our three proposals, i.e. CSAM hash matching, CSAM URL detection and fraud key word detection? Please provide the underlying arguments and evidence that support your view.

As members of the WeProtect Global Alliance, we work with governments, civil society, and industry to develop policies and solutions to protect children from sexual exploitation and abuse online. We would encourage Ofcom to consider the emerging research from the work of the coalition to inform these essential measures to tackle CSEA and CSAM online.

---

[66] Sellman, M. (2024) *Meta tries to hide suicide posts from under-18s,* The Times, https://www.thetimes.co.uk/article/meta-changes-will-stop-children-seeing-harmful-posts-q00vbb7cp

Question 21: Do you have any comments on the draft guidance set out in Annex 9 regarding whether content is communicated 'publicly' or 'privately'?

The guidance is unclear on what constitutes 'publicly' and 'privately' communicated content. It is important that Ofcom provides clarity on this to ensure that the codes of practice, which rely heavily on discerning public and private content, can be effectively complied with. For example, the guidance does not provide a threshold for how many users to should be able to access content for it to be considered 'public', leaving this up to the judgement of services. This will lead to inconsistent application.

This guidance would benefit from test case examples for services and a reminder of their duties under GDPR[67] and the Age Appropriate Design Code.[68]

Do you have any relevant evidence on:

Question 22: The accuracy of perceptual hash matching and the costs of applying CSAM hash matching to smaller services;

Question 23: The ability of services in scope of the CSAM hash matching measure to access hash databases/services, with respect to access criteria or requirements set by database and/or hash matching service providers;

Question 24: The costs of applying our CSAM URL detection measure to smaller services, and the effectiveness of fuzzy matching for CSAM URL detection;

Question 25: The costs of applying our articles for use in frauds (standard keyword detection) measure, including for smaller services; and

Question 26: An effective application of hash matching and/or URL detection for terrorism content, including how such measures could address concerns around 'context' and freedom of expression, and any information you have on the costs and efficacy of applying hash matching and URL detection for terrorism content to a range of services

As members of the WeProtect Global Alliance, we work with governments, civil society, and industry to develop policies and solutions to protect children from sexual exploitation and abuse online. We would encourage Ofcom to consider the emerging research from the work of the coalition to inform these essential measures to tackle CSEA and CSAM online.

User reporting and complaints (U2U and search)

Question 27: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.

Effective content moderation is an important ex post facto component of tackling harm online and is already used widely in the sector. However, it should not be prioritised over up-stream measures to prevent illegal content and activity from being allowed to

[67] ICO (ND) *UK GDPR guidance and resources,* https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources

[68] ICO (ND) Introduction to the Children's Code, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code

take place to begin with. Effectively applied content moderation should serve as one tool amongst many for services to tackle a key measure for compliance but it is not always consistently applied across services and we are concerned that the draft code does not make reference to minimum standards of use.

We support the response of the NSPCC with regard to these recommended measures.

User reporting and complaints (U2U and search)

Question 28: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.

We broadly agree with the proposals Ofcom has made with regard to reporting and complaints. In particular we welcome that Ofcom has taken onboard our concerns around children's use of reporting and complaints mechanisms and the challenges that they can face, and welcome the additional work in chapter 18 on how to support child users.

In particular we welcome that:

- Reporting should allow for contextual information to be supplied in reporting and complaints. This will be especially welcome for complaints regarding harmful content to children, where in isolation it may not meet a threshold but in broader context is harmful.
- Reporting should be available to non-registered users. Through our consultation with the Bereaved Families for Online Safety, the inability to only raise a complaint as a registered user or when logged in was raised a key issue with currently adopted reporting mechanisms.

Our consultation with the Bereaved Families for Online Safety, who have all contended with opaque and challenging reporting systems, raised the following concerns:

- Transparency on judgements: As Ofcom alludes to, there is a lack of transparency in many services as to how to make a complaint. Further to this however was an issue raised by the group that it is not always transparent how decisions are made regarding content, and even where it is perceived to be in contravention of the services terms of service, the service disagrees that is. Ofcom should require that services must provide information as to why a certain decision has been made.

- Updates on reporting: An acknowledgement of the report and a timescale for resolution is a good practice and we welcome this measure. However, the group expressed their concern that where reporting or complaints is related to harm or risk to children, the service's lack of update or further detail can cause additional anxiety. We note in paragraph 16.108, Ofcom has decided to not recommend that users should be provided updates or be given information as to where they are in a database of complaints. Ofcom should look to recommend this for reports from children or regarding child users, where subsequent contextual information has been provided. Emphasis should be

made on establishing what 'good' looks like, offering easily accessible ways for both children and adults to make complaints.

- Reporting child-safety issues: Much of the reporting and complaints process is now automated without adequate access to a human in the system, making it difficult for those concerned about the impact of content on vulnerable individuals – such as parents of children – to raise such concerns urgently. A reporting system should be able to find a person quickly when a child is involved, and thereafter take necessary steps to ensure their safety. Automated systems also often fail to consider the nature by which content has been displayed and to whom, so contextual judgements cannot be made. In addition, for non-registered users, services must be required to clearly provide guidance on how to report without the need to set up an account.

- Right of appeal: While guidance details how services must offer appeals to users or interested persons who may have had content removed unfairly, it is does not include recommendations for users to appeal decisions not to remove content. Ofcom should recommend services that offer a route to appeal these decisions where it involves harm or risk to a child. One parent told us that "We could have the best reporting structure in the world… but they say it meets their community standards and then there is no right for recourse like we would get in a healthcare system… there should be parity with the real world."

Ofcom can encourage greater transparency of reporting mechanisms and complaints by requiring companies to publicly disclose the number and nature of reports made each year on their service. Having these answers in the public sphere will be a pivotal instrument in driving change and creating industry best practice. Additionally, we believe this information should also be presented at board level.

Given much reporting is monitored using AI, it should be standard to provide a drop down set of questions about the nature of the complaints in language that a child might understand, and have an option with a box with an open answer. Only by making a comprehensive reporting structure will Ofcom understand the know the level of harm.

Overall, it should be noted that content reporting by users is not by itself an effective mean to ensure the safety of users, in particular children, and should only be considered as a complementary measure to other safety by design measures. We would also flag that Ofcom should be mindful of not shifting the responsibility of content moderation onto users whilst this responsibility should lie with the services. In relation to flagging and reporting harmful content, children have already stated that they felt there was no point in doing so.[69] Although it is good to provide means for users to identify harmful content or content that should be regulated for children, this should

[69] Office of the Children's Commissioner (2022) *Digital childhoods: a survey of children and parents, September* 2022, https://assets.childrenscommissioner.gov.uk/wpuploads/2022/09/cc-digital-childhoods-a-survey-of-children-and-parents.pdf

only be complementary to providers taking appropriate measures to ensure safety on their service.

Terms of service and publicly available statements

Question 29: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.

Accessible and easy to understand published terms, and privacy policies in particular (which set out how personal data will be used), are central to giving users agency and knowledge regarding the agreement they are entering into when they use a service.

They provide important information such as how personal data is used, reporting mechanisms, and what is and isn't allowed on the service. They are central to helping children in particular understand their rights.

We support Ofcom's recommend measures and standards with regard to terms of service and statements, particularly that it is outcomes based.

While we agree that the reading age of the terms should be understandable to the youngest person able to agree to them, we would note that this would not necessarily mean that a 13-year-old, for example, would understand the contract they were entering in to. We recommend that Ofcom considers the IEEE Standard for an Age Appropriate Digital Services Framework which includes how services can effectively produce age-appropriate published terms.[70] In general, information provided to children on a service must be age-appropriate, designed so that the information they contain comprehensible, an appropriate length, clearly presented, easy to find, introduced at the right moments, and understandable to all young people, no matter who they are, how old they are, or where they come from. Services should obtain meaningful consent of their terms.

A central requirement of the online safety regime that services must uphold their published terms. We would recommend that Ofcom should map how services have changed their public terms since 2022 over time to ensure that companies are not downgrading their terms to comply rather than upgrading their systems and processes to meet the spirit of the Act.

Question 30: Do you have any evidence, in particular on the use of prompts, to guide further work in this area? Please provide the underlying arguments and evidence that support your views.

Please see our response to question 29.

[70] IEEE Std. 2089-2021 (2021) IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, https://5rightsfoundation.com/static/ieee-2089-2021.pdf

Default settings and user support for child users (U2U)

Question 31: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.

U2U default settings and support for child users measures are welcome but we note these are only required for services with a high risk of grooming and large services with a medium risk of grooming.

Default settings for children which promote their safety and privacy are a regulatory requirement under GDPR and the Age Appropriate Design Code (AADC)[71], recommended by the OECD[72] and set out in General Comment 25 on children's rights in relation to the digital environment.[73] As the online world is not built with children in mind, default settings help ensure they have experiences relevant to their evolving capacities and needs. To this end we would strongly argue that these default measures should be in place for all services – not only those with a high risk of grooming.

Ofcom's requirements relating to default settings must not undermine the default settings many services must already apply to child users in the AADC. Default settings must take account of each of the offences in scope of the Act.

Interim child user assessment

While an interim measure ahead of age assurance guidance being published, we strongly disagree that services that do not currently ask for or record the age of a user should can exempt from these measures. We look forward to further guidance on age assurance but would stress that this Code should not undermine the protections child users already receive under the AADC. Alignment with the AADC's 'likely to be accessed' threshold is critical. This is the assessment in use by services currently, many of which will be in scope of the Online Safety Act.

Default settings

"Children using a service are not presented with prompts to expand their network of friends, or included in network expansion prompts presented to other users":

Network recommender systems are a risk particularly in relation to grooming and harassment. Putting in place default settings which protect child user accounts from being promoted to adult user accounts is a welcome measure to tackle this. Ofcom should require these settings being turned off by default, with no option to turn them on for younger groups of children.

"Children using a service are not included in publicly visible lists of who users are connected to, and lists setting out who child users are connected to are not displayed to other users.":

---

[71] ICO (ND) Introduction to the Children's Code, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code
[72] OECD (2022) *Companion Document to the OECD Recommendation on Children in the Digital Environment,* OECD Publishing, Paris, https://doi.org/10.1787/a2ebec7c-en
[73] UN Convention on the Rights of the Child (1989), General Comment No.25 on children's rights in relation to the digital environment (2021), https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

This is already a regulatory requirement for services under the AADC.

"Where services have functionality which allows users to formally connect with one another (e.g. become 'friends') they should ensure that people cannot send direct messages to children using the service without first establishing such a connection.:"

We strongly support this measure. Direct messaging by adults to children is a known risk and there is evidence of how this is used to facilitate offences including grooming and harassment. 5Rights *Pathway*s research found on some user to user services children's accounts were being added to 'Group Chats' user accounts they did not follow on which paid-for pornography was shared.[74] Ofcom should require that these settings are turned off entirely for younger groups of children, with no option to turn back on.

"For services with no user connection functionality, child users are provided with a means of actively confirming whether to receive a direct message from a user before it is visible to them, unless direct messaging is a necessary and time critical element of another functionality, in which case child users should be presented with a means of actively confirming before any interaction associated with that functionality begins.":

We strongly support this measure.

"'Automated location information displays', which automatically create and display the location information for child users, are switched off.":

This is already a regulatory requirement for services under the AADC.

Support for children using a service

Information provided to children on a service must be age-appropriate, designed so that the information they contain is comprehensible, an appropriate length, clearly presented, easy to find, introduced at the right moments, and understandable to all young people, no matter who they are, how old they are, or where they come from. We welcome that Ofcom has included the importance of the timings of prompts, and that they must be presented clearly and a format children can understand.

With regard to prompts for children, we would raise a general concern of their use in persuasive design strategies. They can also serve as a dark pattern which push child users into taking a particular course of action they might not otherwise have taken.[75] Research by the Norwegian Consumer Association found that services attempt to persuade users into accepting certain settings through specific wording and presentation of prompts.[76] Ofcom should make clear that information provided to child users should be empowering, relaying the facts and risks, and nothing should persuade children to lower the protections offered to them by default settings.

---

[74] 5Rights Foundation (2021) *Pathways: How digital design puts children at risk,* https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf

[75] 5Rights Foundation (2023) *Disrupted Childhood: The cost of persuasive design*, https://5rightsfoundation.com/uploads/Disrupted-Childhood-2023-v2.pdf

[76] Forbrukerrådet (2018) *Deceived by design: How tech companies use dark patterns to discourage us from exercising our rights to privacy,* https://storage02.forbrukerradet.no/media/2018/06/2018-06-27-deceived-by-design-final.pdf

We welcome that Ofcom has taken onboard 5Rights evidence as to the importance of timely prompts and warnings for child users, particularly before an action is about to be taken which could lead them to harm.

Question 32: Are there functionalities outside of the ones listed in our proposals, that should explicitly inform users around changing default settings?

We agree with the functionalities listed in the proposals but believe that Ofcom should expand the list to cover a number of additional functionalities that should be turned "off" by default for children in order to keep them safe from illegal harm.

Public profiles

Many services are often have default settings that mean that children are searchable and discoverable by all users, including perpetrators of grooming and other illegal activity. Although some services have implemented privacy by default for profiles - as required by the AADC, it is not always applied consistently and, in instances like Instagram, it is applied to arbitrary ages such as "under 16."[77] Children's profiles, anyone under 18, should be private by default, and never be visible or searchable to *all* other users of the service.

In-game gifts/gifting in-app

In-game/in-app gifting should not allow adults to access children and create circumstances by which they can be exploited. As highlighted in 5Rights *Risky by Design* research, groomers have the ability to leverage children by offering in-game/in-app gifts or currency that can then be used to coerce them into participating in a criminal activity.[78] Children are highly susceptible to commercial pressures, in particular relating to online games, and can easily be taken advantage of by perpetrators.[79] The Sunday Times reported that criminals are using Fortnite in order to groom children as young as 12 to become 'drug mules' by sending children V-Bucks, the virtual currency of Fortnite, before asking them to store drugs or deliver them to customers.[80] Ofcom must ensure that proposals limit the ability of groomers to reach children through in-game/in-app gifting by setting in-game gifts/in-app gifting to 'off' by default.

Livestreaming and video-sharing

5Rights *Risky by Design* research also highlights the greater risk children are at through livestreaming and video-sharing features.[81] Specifically, livestreaming in personal spaces like bedrooms creates an ease of access for groomers to enter a young person's life and build up trust. Livestreams and video-sharing also typically allow for any user to comment on children's posts. Ofcom should require services have livestreaming off by default for child users.

---

[77] UK Safer Internet Centre (2021) *Instagram makes new under 16 profiles private by* default, https://saferinternet.org.uk/blog/instagram-makes-new-under-16-profiles-private-by-default
[78] 5Rights Foundation (2020) *Risky-by-Design,* https://www.riskyby.design/introduction
[79] 5Rights Foundation (2020) *Risky-by-Design,* https://www.riskyby.design/introduction
[80] Mararike, S. (2021) *Dealers are using Fortnite treats to groom children as drug mules,* The Times, https://www.thetimes.co.uk/article/dealers-are-using-fortnite-treats-to-groom-children-as-drug-mules-pq7rkftv2
[81] 5Rights Foundation (2020) *Risky-by-Design,* https://www.riskyby.design/introduction

‘Quick add’

Ofcom should have greater consideration for specific ‘quick add’ systems as they allow children to expand their network at a rapid rate that leaves them at significant risk to being contacted by malign actors. 5Rights *Pathways* research illustrates how ‘quick add’ functions – such as this feature on Snapchat – are being used by children, with one child using it to add people in his area or mutual friends, regardless of whether they were known to him personally.[82] This, built in alongside the pressures many children face to appear popular in the online world, means that bad actors have a greater chance of reaching children quickly. We believe that Ofcom should recommend that ‘quick add’ features are ‘off’ by default for children.

Targeted advertising

Ofcom’s draft register notes that advertising can help facilitate many of the illegal harms in scope of the legislation. Targeted advertising is facilitated by services which collect user data to build ‘profiles’ that can be used to target highly individualised adverts. Some may promote products that are detrimental to a child's health and wellbeing, which breaches the AADC, including products and services that are age-restricted. This kind of advertising should be turned off for children’s accounts and it should be clear when content is sponsored or paid-for.[83]

Autoplay

Many services including YouTube have autoplay features where video or audio content plays without initiation from the user when autoplay is enabled.[84] This content is informed by recommendation algorithms and can become increasingly narrow in focus or amplify extremist content.[85] Similarly to how YouTube turns this off for users under 18, this should off by default across all services for child users.

Ephemeral content

Some services have features which allow for content that expires after a certain amount of time.[86,87] These posts ‘disappear’ before they can be fact-checked and 5Rights research has found young people find them difficult to report.[88] Features which allow for ephemeral content should be turned off by default for child users.

Question 33: Are there other points within the user journey where under 18s should be informed of the risk of illegal content? U2U default settings and support for child users

---

[82] 5Rights Foundation (2021) *Pathways: How digital design puts children at risk,* https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf
[83] 5Rights Foundation (2020) *Risky-by-Design,* https://www.riskyby.design/introduction
[84] YouTube Help (ND) *Autoplay videos,* https://support.google.com/youtube/answer/6327615?hl=en#:~:text=The%20Autoplay%20feature%20on%20YouTube,play%20after%20a%20video%20ends
[85] Whittaker, J. et al. (2021) *Recommender systems and the amplification of extremist* content, Internet policy Review, Vol. 10(2), https://doi.org/10.14763/2021.2.1565
[86] Instagram (ND) *Instagram Stories,* https://about.instagram.com/features/stories
[87] BeReal (ND) https://bereal.com/en
[88] 5Rights Foundation (2020) *Risky-by-Design,* https://www.riskyby.design/introduction

We welcome that Ofcom has taken onboard 5Rights evidence as to the importance of 'just in time' prompts and warnings for child users, particularly before an action is about to be taken which could lead them to harm.

As set out above support for child users and default settings should be required for the full range of harms and for all services in scope of the regulation.

In addition to the requirements Ofcom has made for prompts on the user journey, we would also suggest periodic reminders for child users in cases where default safety settings have been turned off at the request of the child. For example, if a child has turned part of their profile to public, services should seek to remind them it is still public and other users can still see them. One way this could be done is by requesting confirmation of their default settings turned off at the end of every session.

Recommender system testing (U2U)

Question 34: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views

We welcome measures which would support an ex-ante approach to risk assessment of products, as a key feature of safety by design.

Services should undertake a risk assessment of their recommender systems before they are applied to the service. This would help the service to understand potential harm and put in appropriate changes and mitigations as required.

5Rights has developed a four-step algorithmic oversight model which is tech neutral and can be applied to a number of different measures, including redress and content moderation.[89] We will be supplying more evidence in relation to the children's safety code of practice on this issue.

Question 35: What evaluation methods might be suitable for smaller services that do not have the capacity to perform on-platform testing?

Services should undertake a risk assessment of their recommender systems before they are applied to the service. This would help the service to understand potential harm and put in appropriate changes and mitigations as required.

5Rights had developed a four-step algorithmic oversight model which is tech neutral and can be applied to a number of different measures, including redress and content moderation.[90] We will be supplying more evidence in relation to the children's safety code of practice on this issue.

We are aware of design features and parameters that can be used in recommender system to minimise the distribution of illegal content, e.g. ensuring content/network balance and low/neutral weightings on content labelled as sensitive

---

[89] 5Rights Foundation (2021) *Children's Rights and AI Oversight - 5Rights position paper on the EU's Artificial Intelligence Act,* https://5rightsfoundation.com/uploads/Childrens-Rights-and-AI-Oversight---5Rights-position-on-AI-Act-Oct-2021.pdf
[90] Ibid.

Question 36: Are you aware of any other design parameters and choices that are proven to improve user safety?

See responses to questions 34 and 35.

Enhanced user control (U2U)

Question 37: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views

We support measures that would give users including children more control over their experiences online, including the ability to block and mute content and other users.

Online tools to help children exercise their data protection rights are recommended as a key measure of the Age Appropriate Design Code. Standard 15 of that Code states services must "Provide prominent and accessible tools to help children exercise their data protection rights and report concerns." Tools provided by services should be age appropriate and easy for children to use and commensurate to their evolving capacities.[91]

However, it should be noted that promotion of user controls and online tools by users is not by itself an effective mean to ensure the safety of users, in particular children, and should only be considered as a complementary measure to other safety by design measures. We would also flag that Ofcom should be mindful of not shifting the responsibility of safety to children- this responsibility should lie with the services. Although it is good to provide means for users to control what content or users they can see, this should only be complementary to providers taking appropriate measures to ensure safety on their service.

Question 38: Do you think the first two proposed measures should include requirements for how these controls are made known to users?

With regard to child users, existing regulation dictates that online tools and controls should be displayed prominently. Standard 15 of the ICO's Age Appropriate Design Code sets out how this can be done children of different age groups. ICO guidance recommends online tools "should highlight the reporting tool in your set up process and provide a clear and easily identifiable icon or other access mechanism in a prominent place on the screen display."[92]

The IEEE Standard for an Age Appropriate Digital Services Framework also sets out how services can best and most comprehensively apply this design strategy for children.[93]

Question 39: Do you think there are situations where the labelling of accounts through voluntary verification schemes has particular value or risks?

---

[91] ICO (ND) Introduction to the Children's Code, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code
[92] ICO (ND) Introduction to the Children's Code, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code
[93] IEEE Std. 2089-2021 (2021) IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, https://5rightsfoundation.com/static/ieee-2089-2021.pdf

While we agree with Ofcom's assessment that services should provide more transparency regarding how verification schemes are operated, the guidance on this measure does not raise the unique vulnerability of children to bad actors online. The recommendations would benefit from making this clearer.

Children's developing cognitive abilities means that they cannot always distinguish between reliable and unreliable information online.[94] As Ofcom has found, verification schemes can be used by bad actors to impersonate official sources and mislead users. Reporting on X Verification in particular found these schemes are particularly vulnerable to scams.[95] According to Ofcom's own research, nearly a quarter (23%) of children claimed to be confident in their ability to identify what is real or fake online but could not correctly identify a fake social media profile when presented with one.[96] With this in mind, children are more susceptible to the risks of fraud and bad actors and Ofcom should ensure services have regard to this in their operation.

Any information, context or prompts services put in place to create more transparency around how users can gain verified status must also must be age-appropriate, designed so that the information they contain is comprehensible, an appropriate length, clearly presented, easy to find, introduced at the right moments, and understandable to all young people, no matter who they are, how old they are, or where they come from. The IEEE Standard for an Age Appropriate Digital Services Framework also sets out how services can best and most comprehensively apply this design strategy for children.[97]

User access to services (U2U)

Question 40: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.

Given the lack of consistency and evidence as to how this measure is currently being used by the services, we would recommend that Ofcom uses its information-gathering powers to understand how regulated services are approaching blocking or using strikes with regard to users who have acted in contravention with the law or it terms of service to help inform future iterations of the Code.

This should consider challenges such as the use of VPNs and how to prevent these users from creating new profiles to tackle perpetrators from using burner accounts. This measure would benefit from specific guidance which provides criteria to support services to implement this proportionately and consistently.

This should also inform Ofcom's proposals regarding users who have shared CSAM.

---

[94] Vosloo, S. (2021) *Digital misinformation/disinformation and children: 10 things you need to* know, UNICEF, https://www.unicef.org/globalinsight/stories/digital-misinformation-disinformation-and-children
[95] Burgess, M. (2022) *Elon Musk's Twitter is a scammer's paradise,* WIRED, https://www.wired.co.uk/article/twitter-blue-check-verification-buy-scams
[96] Ofcom (2023) *Children and parents: Media use and attitudes,* https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2023
[97] IEEE Std. 2089-2021 (2021) IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, https://5rightsfoundation.com/static/ieee-2089-2021.pdf

Services should work with the relevant enforcement authorities regarding individuals they suspect of having committed offences on their platforms.

Do you have any supporting information and evidence to inform any recommendations we may make on blocking sharers of CSAM content? Specifically:

Question 41: What are the options available to block and prevent a user from returning to a service (e.g. blocking by username, email or IP address, or a combination of factors)?

Please refer to our response to question 40.

What are the advantages and disadvantages of the different options, including any potential impact on other users?

Please refer to our response to question 40.

Question 42: How long should a user be blocked for sharing known CSAM, and should the period vary depending on the nature of the offence committed?

Please refer to our response to question 40.

Question 43: What steps can services take to manage this risk? For example, are there alternative options to immediate blocking (such as a strikes system) that might help mitigate some of the risks and impacts on user rights?

Services should work with the relevant enforcement authorities regarding individuals they suspect of having committed offences on their platforms.

Service design and user support (Search)

Question 44: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.

Reporting for predictive search

Predictive search functionality is a known risk to harmful and illegal content and we agree that users including children should be able to access prominently displayed reporting mechanisms for illegal and harmful content. However, as set out elsewhere in this response, promotion of user controls and online tools by users is not by itself an effective mean to ensure the safety of users, in particular children, and should only be considered as a complementary measure to other safety by design measures. Ofcom should be mindful of not shifting the responsibility of safety to users including children - this responsibility should lie with the services. Although it is good to provide means for users to control what content or users they can see, this should only be complementary to providers taking appropriate measures to ensure safety on their service.

With regard to warning messages for users searching suicide content or CSAM, we urge Ofcom to consider the response of the NSPCC.

Cumulative Assessment

Question 45: Do you agree that the overall burden of our measures on low risk small and micro businesses is proportionate? Please provide the underlying arguments and evidence that support your views.

Please refer to our response to Question 4 on costs and proportionality.

Question 46: Do you agree that the overall burden is proportionate for those small and micro businesses that find they have significant risks of illegal content and for whom we propose to recommend more measures? Please provide the underlying arguments and evidence that support your views.

Please refer to our response to Question 4 on costs and proportionality.

Question 47: We are applying more measures to large services. Do you agree that the overall burden on large services proportionate? Please provide the underlying arguments and evidence that support your views.

Please refer to our response to Question 4 on costs and proportionality.

Statutory Tests

Question 48: Do you agree that Ofcom's proposed recommendations for the Codes are appropriate in the light of the matters to which Ofcom must have regard?

We urge Ofcom to consider the statement of the Online Safety Act Network in response to this consultation, of which 5Rights is a signatory.[98]

Volume 5: How to judge whether content is illegal or not? The Illegal Content Judgements Guidance (ICJG)

Question 49: Do you agree with our proposals, including the detail of the drafting? What are the underlying arguments and evidence that inform your view?

Ensuring that services are safe by design is a key objective of the Act and this must be part of the implementation process. We are concerned that Ofcom has failed to reflect safety by design principles in the discussion of the meaning of illegal content in Volume 5, with guidance geared to identifying individual items of content (as per Paragraph A1.14 in Annex 10) as opposed to also considering systems and processes that allow illegal content to manifest. This is despite the fact that, at Committee stage in the House of Lords, the minister said:

> "To be clear, the duty requires platforms to put in place proportionate systems and processes designed to prevent users encountering content. I draw my noble friend's attention to the focus on systems and processes in that. This requires platforms to design their services to achieve the outcome of preventing users encountering such content. That could include upstream design measures, as

[98] Online Safety Act Network (2024) *Ofcom's illegal harms consultation: Emerging concerns,* https://www.onlinesafetyact.net/uploads/osa-network-ofcom-illegal-harms-sign-on-feb-2024.pdf

> well as content identification measures, once content appears on a service."[99] (Lord Parkinson of Whitley Bay, 27 April 2023)

Further, Volume 5 and the draft guidance has an overwhelming focus on ex-post measures (see Paragraph 26.43 and Paragraph A1.16 in Annex 10) and does not consider ex-ante measures or safety by design measures, and thus does not assess how measures are improving safety. We believe Ofcom must review this guidance to ensure that more measures than content takedown – a practice which many companies already employ – are applied, in particular regarding the design and operation of the service, as relates to all duties in the Act.

With regard to the SSH offence detailed in Volume 5, we strongly urge Ofcom to consider the response of the Molly Rose Foundation to this consultation.

Question 50: Do you consider the guidance to be sufficiently accessible, particularly for services with limited access to legal expertise?

Please refer to our response to question 49.

Question 51: What do you think of our assessment of what information is reasonably available and relevant to illegal content judgements?

Please refer to our response in Question 49.

Volume 6: Information gathering and enforcement powers, and approach to supervision.

Information powers

Question 52: Do you have any comments on our proposed approach to information gathering powers under the Online Safety Act? Please provide the underlying arguments and evidence that support your views.

Ofcom's information gathering powers are central to the effective implementation of this legislation and ongoing regulation. While it is important these powers are used proportionally, it is important that Ofcom see its regulatory role as one which can support services of all sizes to respond to these notices. To this end, Ofcom should set out how it intends to support smaller services to respond and comply with these notices, rather than suggest they may not need to comply.

This volume does not reference the powers in Section 101 of the Act regarding *Information in connection with an investigation into the death of a child.* This section would benefit from clear guidelines on how the information gathering powers will apply to support the understanding of relevant persons involved.

Enforcement powers

[99] Lord Parkinson of Whitley Bay, 27 April 2023, https://hansard.parliament.uk/lords/2023-04-27/debates/958CAC63-A345-45E8-9DE3-7CBA46611DCA/OnlineSafetyBill#contribution-96C5009F-1379-4FA0-B5FF-B7FC725CB2DB

Question 53: Do you have any comments on our draft Online Safety Enforcement Guidance? Please provide the underlying arguments and evidence that support your views.

It is welcome that, as per the legislation, the decision to take enforcement action will prioritise where the service has contravened its child safety duties.

With regard to paragraph 29.39(b), in the determination that children can access part of or all of the service, Ofcom must ensure that the age verification the service has in place meets the standard set out the age assurance guidance. Although a service may have age verification in place, depending on its quality, it does not automatically mean children cannot access the service.