July 2024

**TikTok response to Ofcom’s Online Safety Act consultation:
Protecting children from harms online**

Executive summary

TikTok welcomes the opportunity to contribute to Ofcom’s consultation on the protection of children online. We are one of the few platforms to have worked with Ofcom as a regulated video-sharing platform since 2020 and are committed to continued collaboration under the Online Safety Act (OSA) to ensure young people can safely enjoy the full benefits of being online.

Protecting our users, particularly our younger users, is a top priority for TikTok. We have invested heavily to build in safety by design on our platform, and ensure it remains a fun and safe environment that inspires creativity and brings joy (as we set out further below). We are proud of these innovative efforts, and pleased to see that many of Ofcom’s proposed measures in the draft Code of Practice reflect our existing practice. However, we recognise there is no finish line when it comes to safety, and we continually look for ways to enhance our protections. We hope that the views, information and insights we are able to share with Ofcom in this consultation response contribute towards creating and sustaining a safe online environment for children.

We have focused this response on key themes within the consultation and the response is structured by reference to those themes, as summarised below. Where we have identified a need for further information or clarity as to how certain aspects of the draft Code and draft guidance will operate in practice, or suggested an amendment to the documents, we have outlined this under each theme in the body of the response. In summary:

1. **Services should be given flexibility to rely on the most innovative and effective solutions:** The draft Codes contain a number of prescriptive measures which require services to take specific steps rather than setting out the outcome to be achieved and giving them the flexibility to design compliance methods that best achieve it. Where services are experienced in trust and safety innovation, it is particularly important to be able to add new tools and more effective methods as they become available. This will enable services to carefully design controls tailored to the risks and features of their particular service, while still achieving the same aims.

2. **The approach to age assurance must be risk-based and proportionate:** We would welcome further clarity on what methods Ofcom will regard as “highly effective age assurance” and in particular:
    - confirmation that a multi-layered approach to age assurance is encouraged and that sufficiently accurate age estimation models would be regarded as highly effective;
    - a greater differentiation in what would be regarded as highly effective age assurance for services with different levels of risk, such that only those services that pose the highest safety risk to children are required to use the

most intrusive “hard” forms of age assurance, such as photo identification at account sign-up; and

- ○ clarification that services can assess the accuracy of age assurance tools by reference to quantitative performance targets the service has set.

3. **Ofcom’s approach to interpreting the categories of content harmful to children should be evidence based and avoid incentivising over-removal:** Ofcom’s ‘Children’s Register of Risks’ appears to interpret a number of the primary priority (PPC) and priority (PC) categories of content in a manner that is overly broad and could incentivise services to err on the side of removing content even where it may be disproportionate to do so based on the strength of the available evidence. This could have a severe impact on access to information for both child and adult users in the UK.

4. **Recommender systems should be recognised as not only a risk factor but also a means of promoting high-quality content and online safety:** Our recommender systems, which are underpinned by robust trust and safety strategies, ensure that younger users on our service see relevant, diverse and age-appropriate content, such as our STEM feed, which we discuss in more detail below. These potential benefits of recommender systems, and their role in protecting users from harm, should be recognised in the Register of Risks, Risk Profile and draft Code.

5. **The Governance and Accountability measures should not subject individual employees to undue public scrutiny:** We would ask Ofcom to clarify that the person who must be named as accountable to the most senior governance body for compliance with the safety duties should only be named internally or confidentially to Ofcom. Requiring this person to be named publicly could expose them to undue pressure and risk and keeping their name confidential would ensure services designate the most appropriate person based on their role, not their existing public profile.

6. **Ofcom must ensure that measures are designed in consideration of the principles of protecting users’ rights to freedom of expression and privacy:** In particular, under data protection law, services must ensure that the amount of personal information they collect about a person to verify or assure their age is proportionate. In recommending that services with differing levels of risks should apply the same standard of ‘highly effective age assurance’, Ofcom should ensure that it has appropriately considered the principle of user privacy as required under the OSA. We would therefore welcome clarification from Ofcom that the necessary consultations are being undertaken with the Information Commissioner’s Office (ICO) in respect of the proposals and that the final proposals will be confirmed to be compliant with the relevant data privacy laws.

7. **Ofcom’s approach should be aligned with regulators in the UK and, where possible, in other jurisdictions:** We strongly support Ofcom’s recommendation in certain parts of the consultation that services should align with existing international best practice (for example, in reviewing children’s risk assessments annually in alignment with the EU’s Digital Services Act). Similar to our point above regarding the

consideration of privacy rights, however, Ofcom should continue to utilise existing cooperation channels to collaborate with regulators in the UK, as well internationally, ensuring alignment with similar regulations (such as the EU's Digital Services Act) where possible and taking feedback from other regulators into account when developing the final versions of Codes of Practice and guidance.

How TikTok keeps children safe online

At TikTok, we are committed to creating a safe platform for our community, and believe this work is especially important for our younger users. We have invested heavily to put in place measures that are focused on ensuring that those under 13 do not access our platform, and that users aged 13 - 17 are able to enjoy a safe, positive and developmentally appropriate experience.

We adopt a multi-layered approach to age assurance, and implement a variety of measures directed towards preventing children under 13 accessing the service. This starts with a minimum age for downloading our app in the Apple and Google app stores, and once users are using the platform, our measures continue to seek to identify children under 13 and those that are 13 - 17. TikTok trains its safety moderation team to be alert to signs that an account may be used by a child under the age of 13. We also use other information as provided by our users, such as keywords and in-app reports from our community, to help detect potential underage accounts. When our safety team reviews an account and a moderator confirms the account appears to belong to an underage user, the account will be suspended.

If the user believes that we have made a mistake in determining their age, they may submit an appeal to the ban by confirming their age through submission of either an ID, a selfie with a trusted adult if they are 13-17, or a temporary credit card authentication if they are over 18. Users over 18 also have the option of using facial estimation to confirm their age.

TikTok has published this figure of suspected underage removals regularly since June 2021, and remains the only major platform to do so. In Q1 2024 TikTok removed 21,639,414 accounts suspected to be under 13. While no other peer platform publishes this statistic, previous Ofcom reports have highlighted this figure is significantly higher than other video-sharing platforms.

Outside of age assurance, some of TikTok's other industry-leading measures include:

- **Default settings:** as teens start to build a presence online, it's important for them to learn about and review their privacy settings and controls. We implement a number of default privacy settings for users under 18, and limit access to some features depending on the user's age.
- **Screen time limits:** a default 60-minute screen time limit for users under 18, with the ability for parents to set stricter limits using Family Pairing (TikTok's parental control tool), as well as muting notifications for younger users at nighttime.
- **Content levels**: assigning content maturity scores, creating a system which organises content based on thematic maturity, similar to systems used in TV and film

classifications, and automatically preventing content with overtly mature themes from reaching younger users.
- **Content Authenticity:** TikTok is the first video sharing platform to require all realistic AI-generated content (AIGC) to be clearly labelled, and the only major platform to integrate the Coalition for Content Provenance and Authenticity (C2PA) 'Content Credentials' into TikTok. Content Credentials automatically labels AIGC as it is uploaded, and enables content originating from TikTok to be automatically labelled if it is shared on other services.
- **STEM feed:** a dedicated feed focusing only on Science, Technology, Engineering, and Mathematics content, all fact-checked by independent third parties, and enabled by default for all users under 18.
- **TikTok Youth Council:** a global council of young people with experience of using our platform, providing their input and perspectives to enable us to be better positioned to make changes that create the safest possible experience for our community.

We are pleased to see that many of the measures proposed by Ofcom in the draft Code of Practice align with our existing efforts, and hope that some of the best practices we have shared here can be implemented more broadly across the industry to create an online environment where young people can express themselves and access information safely.

Ensuring appropriate flexibility in the Codes of Practice

The OSA sets out the 'comply or explain' principle (ss.23 and 49 of the OSA), which requires services to either implement the measures set out in the Code and be deemed compliant, or implement alternatives and record how this achieves equivalent compliance with the duties imposed on platforms by the OSA. Ofcom currently does not reflect this principle in the draft Code and guidance.

The comply and explain principle is crucial to the online safety regime as no platform is the same: the types of harm, their manifestation, and the best mitigation, may all differ accordingly. However, it is not clear how the ‘comply or explain’ principle would work in practice given the prescriptive nature of the Code, both in terms of the risk profiles outlined and the recommended measures put forward. Services may be incentivised to adopt the prescriptive (but possibly ill-suited and less effective) measures in the Code for the benefit of the assurance they will be deemed compliant with the Act.

For example, Measure CM4 requires providers to take certain requirements into account when prioritising content for review. This includes a requirement to ‘*prioritise content for review in a way which minimises circumstances in which the number of child users encountering a particular item of content that is harmful to children increases exponentially over a period of time*’. Although the virality of a piece of content is an important factor in the prioritisation of content for review, there may be circumstances in which it is important to prioritise reviews based on other factors (for example, content that may be more harmful but viewed by fewer people). This may differ depending on the functionalities of the service, or on the type of content being reviewed.

Given this, there is a risk that the granularity and specificity of the proposed measures is taken to imply that no other measures could be sufficient. Coupled with Ofcom's strong enforcement powers, this means that there are strong incentives to focus compliance efforts on achieving the standards set out in the Code rather than developing solutions that may be best suited for a particular service. In the context of Measure CM4, for example, providers should be able to use their on-platform experience to develop a content review policy which prioritises content according to the most relevant factors for the nature of content available on the service.

Developing new trust and safety systems can be an extremely resource intensive procedure, involving thousands of employees across the globe, which is a disproportionate requirement in a situation where less intensive methods can be implemented to achieve the same outcome. The prescriptiveness of some of the measures increases the risk of creating a homogenous approach amongst industry and stymying trust and safety innovations, as even if services wanted to introduce new measures that go beyond the Code, there would be continued uncertainty as to whether they will be deemed sufficient to comply.

Recommendations

- Amend the Code to be more outcomes-focused and less prescriptive, for example, in relation to how services design content prioritisation frameworks, relative to risks that they have identified.
- Ofcom should publish public statements, including in the accompanying guidance, that set out that Ofcom supports alternative measures that services take to comply with the requirements in the OSA, and that trust and safety innovation is encouraged.
- Amend the Code such that they facilitate the continued testing of new systems, for example through the application of a Good Samaritan principle, or a regulatory sandbox model.

Highly effective age assurance

Annex 10 of the draft Code sets out criteria that service providers should review to ensure that the age assurance process is highly effective, and provides various examples of age assurance measures that could be considered highly effective. The draft Code provides for the circumstances in which, and the services on which, highly effective age assurance should be used.

We consider that further guidance should be provided in relation to what will constitute "highly effective age assurance", and in particular, whether Ofcom plans to assess "effectiveness" quantitatively. We recommend that Ofcom clarify that services should self-assess whether their age verification is "highly effective" on the basis of internal statistics, for example through comparing the performance of the age assurance tools against performance targets that the service has set based on Ofcom's guidance as to what constitutes highly effective age assurance.

In the current guidance, "highly effective age assurance" appears to be interpreted in the same way for all services, regardless of whether the service is medium or high risk for PPC / PC. The current draft Code requires that any service that is at medium **or** high risk for one or more types of PPC / PC must use the same types of highly effective age assurance. There is therefore no differentiation between services that are at "medium" and "high" risk for a particular type of harm. This may lead to the unintended consequence that there is no regulatory incentive for high risk services to reduce risk. Similarly, the current draft Code fails to differentiate between services that are risky across multiple categories, with those that are risky for just one category. This is in tension with the requirement in the OSA that the Code must have regard to "*the principle that more effective kinds of age assurance should be used to deal with higher levels of risk of harm to children*" (Schedule 4, para 12(2)(d)).

We therefore consider that, in alignment with the intention of the OSA, Ofcom should recognise that the type of age assurance necessary for a service may differ depending on its risk-level, and the different types of content that are high risk. For example, the age assurance required for a service that is high risk across a number of PPC / PC categories is likely to be different from what is necessary for a service that is medium risk across one or two of the categories. What constitutes "highly effective" age assurance should therefore be flexible in line with this risk-based approach. In order to ensure Ofcom's approach is proportionate and aligns with data minimisation best practices, the most intrusive forms of age assurance, such as the use of “hard” age assurance like photo identification on account sign-up, should be necessary only for the riskiest services.

The current approach taken in the Code risks unfairly and unnecessarily impacting users' privacy rights by relying on data intensive measures, or measures that may not be accessible to a significant number of users, to determine the age of users. We ask that Ofcom continue to work with the ICO to ensure that the approach taken in the Code aligns with broader data protection and privacy requirements, and we expand on this issue below.

Furthermore, age estimation models should in principle be highly effective, if sufficiently reliable and accurate, for use on the vast majority of services. We therefore recommend that Ofcom confirms that age estimation models can be used as a highly effective age assurance method, where evidence indicates that the model being used is reliable and accurate. These methods can be used in conjunction with other forms of age assurance, such as self-declaration when signing up.

Recommendations

- Ofcom should clarify that services should self-assess whether their age assurance is “highly effective” on the basis of internal statistics.
- Ofcom should ensure the Code reflects the fact that assurance must be risk-based and proportionate, by sufficiently differentiating between the highly effective age assurance required on:
    - Services that are at medium risk of PPC / PC, and those at high risk and;
    - Services that are at risk of one category of PPC / PC, and those at risk of multiple.
- Ofcom should coordinate with the ICO and other regulators to ensure the approach taken in relation to age assurance aligns with broader data protection and privacy requirements.
- Ofcom should confirm that age estimation models should in principle be considered highly effective age assurance for use on the majority of services.

Definitions of harmful content

The draft Code sets out that different measures apply to different services depending on the type of risk that they pose in relation to content that is harmful to children. Ofcom also proposes to include two categories of content - “body image content” and “depressive content” - as non-designated content (NDC), and in some cases to extend the applicability of the measures to this type of content. We recognise that there are circumstances in which this type of content can be harmful to children, and our policies to address this content are drafted by trust and safety experts with deep subject matter expertise in their field, supported by input from global experts.

Every piece of content posted on TikTok is reviewed by our automated moderation technology, and our systems are designed to capture all types of Community Guidelines violations, which include all types of PPC / PC. However, Ofcom’s current proposals do not recognise the limitations in relation to proactive monitoring that occur in relation to context-dependent content types, nor the risks that may arise from inaccurate moderation if these limits are ignored. In particular, Ofcom’s proposal to extend the measures to certain specified categories of NDC risks introducing uncertainty into the approach taken by services.

While we welcome the steps that Ofcom are taking to assist services in preventing children from encountering harmful content, we note that some types of harmful content are more difficult to proactively identify and remove than others, because the assessment of whether that content is violative is context-dependent. For example, pornography is likely to be much easier to identify – both through automated and human means – than eating disorder content, particularly in relation to recovery content (we note that in Ofcom's view content "*that in some way is intended to be, or is presented or described as, recovery content*" will meet the definition of eating disorder content). We therefore consider that the way in which Ofcom has currently sought to interpret different categories of PPC and PC is likely to result in technical difficulties that will adversely affect compliance.

Furthermore, by designating specific categories of content as NDC which is subject to measures in the Code, Ofcom appears to be elevating such content to priority content, which has a statutory basis. We consider that Ofcom should not be designating new categories of content in this way, but, if it assesses that it is necessary, should use the appropriate statutory processes to incorporate these categories into the requirements under the Act. The current approach to include categories as NDC appears at odds with the statute, which sets out that services are to identify NDC themselves through the children's risk assessment.

We recommend that Ofcom re-assess its current proposal that NDC includes body image content and depressive content. Ofcom notes that it does not currently have sufficient evidence to define the categories of content more clearly, nor to determine the relationship between these kinds of content and the material risk of significant harm. Given the need for Ofcom to act proportionately, and the unintended consequences that may result from requiring services to remove content where there is no clear link to harm, it is critical that Ofcom recommends measures where evidence (whether established through engagement with subject matter experts and providers, or through higher quality independent research) suggests that they are effective in mitigating the identified risk.. It is only through an understanding of the harms and their manifestation that appropriate measures to tackle harm can be introduced. The iterative nature of the Code allows the latest evidence base to be reflected, and this approach prevents a situation where risk profiles are outlined without measures to mitigate their potential harm, or the misapplication of measures which could have unintended consequences.

In relation to the specific categories that Ofcom has identified, we consider that these are not appropriate in their current form. These categories are highly subjective: depressive content in particular is incredibly difficult to define, and whether a piece of content constitutes depressive content is frequently context-dependent. For Ofcom's requirements to be effective, they must be technically feasible. Ofcom should therefore more closely consider the technical difficulties associated with the implementation of this broad category of content, that relies on services making contextual judgements, which are difficult to automate, about whether a piece of content falls into the category. Extending the measures to these imprecise categories may inadvertently lead to an unnecessary level of enforcement that restricts the ability of children to access non-harmful and beneficial content.

Over-regulation in this area could lead to a disproportionate level of enforcement and result in the removal of, for example, recovery content, thereby restricting healthy conversation of recovery content about this issue, as well as unduly interfering with freedom of expression.

Recommendations

- Ofcom should recognise the nuances inherent in types of content, and provide services with the latitude to interpret categories of PPC and PC using proportionate systems and processes, to avoid resulting in over-removal and risk adversely affecting freedom of expression.
- Ofcom should not be designating new categories of content as NDC, but, if it assesses that it is necessary, should use the appropriate statutory processes to incorporate these categories into the requirements under the Act.
- In the event that Ofcom does determine that it should be designating categories of content as NDC, it should only recommend measures and set out guidance where it has a sufficiently strong evidence base to conclude that the measure is effective in achieving the desired outcome.

Recommender systems

Ofcom positions recommender systems in its draft Code and guidance as being a risk factor that can increase the risk that users encounter content that is harmful to children. However, it is important to acknowledge there are positives to some of the platform features outlined as risk factors in the Risk Profiles and restricted by the draft Code, as well as potential risks, in order to consider proposed measures proportionally. This is particularly true in relation to recommender systems, which are not necessarily harmful means of disseminating content.

Recommender systems are a key means by which platforms can ensure users are provided with diverse, engaging and high-quality content that is still relevant to them. For example, TikTok’s recommendation system works to intersperse recommendations that might fall outside people's expressed preferences, offering an opportunity to discover new categories of content. Our systems will not generally recommend two videos in a row made by the same creator or with the same sound, which enriches the viewing experience and can help promote exposure to a range of ideas and perspectives on our platform.

Recommender systems can therefore avoid people being kept in “content bubbles” which may result in a greater degree of harm, for example by diversifying their experience. In contrast, platforms that are driven by ‘social systems’ (i.e. where the content presented comes solely from accounts that the user follows) may result in users being exposed to extremely low levels of content diversity (which in turn, can have the effect of confirmation bias).

Trust and safety systems built around recommender systems can reduce potential risks, providing greater avenues for services and users to restrict or control content (for example by making content ineligible for recommendation), and represent new opportunities for media literacy content to be effectively targeted in an manner that is consistent with the user experience, deepening engagement.

Recommendation

- Ofcom should recognise in the Code, and in particular in its Register of Risks, that recommender systems are not intrinsically harmful to children, but can be used to diversify the content children encounter, to their benefit.

Governance and accountability

Measure PCU A2 in the draft Code requires services to name a person accountable to the most senior governance body for compliance with the safety duties protecting children and reporting and complaints duties. However, it is unclear from the draft Code whether this individual needs to be named internally, confidentially to Ofcom, or publicly. While we recognise the need for accountability in carrying out the requirements of the Act, requiring services to name individuals publicly risks placing undue personal pressure on the named individual, without any clear or evidenced benefit for the protection of children.

There may be negative unintended consequences that arise as a consequence of this individual being publicly named, and potential impacts to their privacy and safety. Companies compete globally amongst themselves and other sectors to attract the best talent to lead different functions. We want the best to develop and lead our trust and safety functions, and effective governance arrangements should complement that aim. The requirement for certain executives to be named publicly would run contrary to that aim.

Recommendation

- Ofcom should clarify that the individual must be named by the service internally and, if Ofcom considers it to be necessary, in a confidential communication to Ofcom.

Balancing effective measures with the principles of user privacy and freedom of expression

We welcome Ofcom's statement in the consultation that it has sought to strike a fair balance between securing adequate protections for children from harm, and the rights to privacy and freedom of expression of individuals (both adults and children). Despite this, there continues to be the potential for conflict between the measures that Ofcom are recommending, and the privacy framework established by the ICO. In the ICO's response to Ofcom's Illegal Harms Consultation, we note that the ICO raised a number of important points of alignment with data protection law, and we have identified similar areas of concern in this Consultation.

On age assurance, for example, Ofcom's proposals appear to conflict with the ICO's Children's Code, which warns against using age assurance where it may be disproportionately intrusive, or may result in the exclusion or discrimination of already marginalised groups. The ICO Code also explicitly states that the definition of "highest possible" certainty on age of users for high-risk services does not extend to measures which "*are not currently technically feasible*" or which pose *"a significant and disproportionate economic impact on their business*". By contrast, as outlined at the start of our response, Ofcom's draft Code suggests that all services, once caught by the age assurance recommendations, must implement the same standard of 'highly effective age assurance', regardless of the differing levels of risks that such services may present to users.

Recommendations:

- Ofcom should align the draft Codes with applicable data protection law, and in particular, the ICO's Children's Code.

Regulatory alignment

While Ofcom recognises that global and shared regulatory standards are beneficial, it currently does not set out how the draft Code and guidance align with existing regulatory frameworks both within the UK and in other jurisdictions.

TikTok supports innovations like the Digital Regulation Cooperation Forum and the Global Online Safety Regulators Network, which work to deliver a coherent approach to online regulatory matters. This is important given the wide-ranging impact that measures introduced in one jurisdiction may have on the both other jurisdictions, and on the work of different regulators in the same jurisdiction. Ofcom rightly acknowledges that global and shared standards are constructive, and encourage more innovation. While this is a helpful starting point, we believe more can be done to improve regulatory alignment and the recognition of shared standards and expectations.

Implementing shared standards and expectations across regulatory regimes can give confidence to services that measures they are taking will be deemed compliant by a range of regulators, helping to drive a 'race to the top'. It could also improve the pace at which online safety innovation takes place, by giving services the confidence that some technologies have been agreed between different regulators, which will encourage services to commit to investing in them. Crucially, it would help to avoid a regulatory trade-off: where one regulator recommends or mandates measures which could actively conflict with requirements from other regulators.

Recommendation

- Ofcom should clarify whether and how existing cooperation mechanisms between regulators are engaged in the design of the Code and guidance.
- Ofcom should continue to utilise existing cooperation channels to collaborate with regulators in the UK and internationally, ensuring alignment with similar regulations where possible and taking feedback from other regulators into account when developing the final versions of Codes of Practice and guidance.

Conclusion

As highlighted throughout our response, TikTok remains supportive of the OSA, the draft Code and guidance and Ofcom's broader work to create safer online experiences, particularly for children. We hope the feedback and suggestions we have provided in this response are helpful to Ofcom. TikTok has been, and continues to be, an industry leader in online safety, placing the safety of our users at the heart of our approach and continuously innovating to introduce new safety and user empowerment tools.

We are pleased to have played a part in developing Ofcom's thinking through our years of regulation under the video-sharing platform regime, and hope some of the best practices shared can be adopted across the wider sector to develop an online environment that allows young people to express themselves and access content of interest in a safe manner.

We remain committed to engaging with Ofcom through the finalisation of the Codes and guidance to ensure an effective regulatory regime that achieves our shared objective of improving online safety.