



Pinterest Response to Consultation: Protecting people from illegal harms online
23 February 2024

Introduction

The following is Pinterest's response to the consultation published on 9 November 2023 by the Office of Communications (Ofcom) on how online platforms should approach their new duties under the Online Safety Act (OSA) relating to illegal content, including draft Illegal Content Codes of Practice for User-to-User Services (the "Codes"). Pinterest welcomes the opportunity to respond to the consultation, as we share Ofcom's goal of fostering a safe and positive online environment and support its objective of creating codes of practice that are proportionate and adaptable to the varied platforms that are subject to the OSA.

Pinterest is a visual discovery engine where users (or “Pinner”) around the world go to get the inspiration to create a life they love. Browsing and saving visual ideas on Pinterest’s service helps Pinner imagine what their future could look like, and go from inspiration to reality.

Pinterest shows Pinner visual recommendations called Pins. Pins are created by both individuals and businesses by uploading photos or videos or bookmarking content from the web, and providing a text caption. Pinner can save and organise these recommendations into collections called boards. Some of our most common use cases include food, beauty, home decor, and travel.

Pinterest has a longstanding commitment to making our platform a safe place for everyone, which is the right thing for the people who use Pinterest, and aligns with our goals as a business. We view a positive, inspiring atmosphere as essential to our user experience, and accordingly, we have developed industry-leading content policies that broadly prohibit, among other things, pornography; any content that contributes to sexualization or sexual exploitation of minors; content promoting suicide or self-harm; hateful or violent content; and health, climate, or civic misinformation.¹ We enforce these policies through automated tools, manual review, and hybrid approaches that combine elements of both, and we work with trusted third parties and the wider industry to enhance our work.

These efforts are highly effective in drastically limiting the prevalence and reach of harmful content on Pinterest, but we know that the work needed to achieve our goal of providing a safe and positive platform for our users will never end. We’re committed to keeping Pinterest a positive corner of the internet. That is why we continually review and update our safety policies, tools, and technologies. Keeping our community safe is an industry-wide challenge as online trends and technological solutions continue to evolve, and we remain committed to exploring the best ways to achieve this goal. Indeed, as a founding signatory of The Inspired Internet Pledge, Pinterest has publicly pledged to centre its commitment to making the internet a healthier place through its product, thought leadership, policies, and strategic giving.²

We expect that we will already be compliant with significant segments of the Codes thanks to our stringent content safety policies, robust content moderation systems, and strong measures for the protection of minors. However, as detailed herein, we believe certain sections of the Codes either require clarification or impose requirements that are

¹ Pinterest’s Community Guidelines can be accessed at: <https://policy.pinterest.com/community-guidelines>

² For more information on The Inspired Internet Pledge, please see: <https://inspiredinternet.org/signatory/pinterest/>

not effective, practicable, or proportionate in the context of Pinterest's functionality and common use cases.

This response is divided into two sections, with the first containing general comments that are applicable throughout Ofcom's proposals, and the second consisting of responses to specific questions posed by the consultation.

General Comments

Pinterest's general comments address (1) the importance of proportionality in drafting and applying the Codes; (2) whether and how platforms may implement measures other than those included in the Codes; and (3) a suggestion regarding the materials to be included in future consultations.

The Importance of Proportionality in Codes of Practice

Pinterest is encouraged by Ofcom's stated intention that the Codes be implemented in a manner that is proportional to platforms' differing risk profiles and operational capacities. In order to preserve competition in the digital sector and allow platforms to implement measures that are suitable and effective for their own platforms, we agree that it is crucial for the Codes to be interpreted and applied in a manner that is tailored to the OSA's objectives and ensures compliance is practicable for the various platforms in the law's scope.

The proportionality of the Codes' measures may be assessed by taking account of a range of factors, including the size of the platform, its functionalities and use cases, the risk to users, existing mitigation measures, relevant industry standards, users' rights, and the platform's resources compared to the cost and effectiveness of the proposed measures. In view of these factors, it's clear that platforms are exceedingly diverse, and accordingly the Codes should allow flexibility for platforms to implement only the safety measures that are practical and effective in light of their unique designs and risk profiles, and to do so in a manner that is suited to the features and functionality of their platforms.

In order for small- and medium-sized platforms to continue to innovate and grow, it is important for these platforms to avoid taking steps that are not necessary to promote their users' safety, considering the substantial aggregate costs of compliance with the Codes' measures. Even platforms that may be considered "large" due to their user base may have vastly different resources and capabilities than other "large" platforms due to differences in their business models or stages of development. As such, we recommend

a nuanced approach that accounts for platforms' respective features, risk profiles, and resources, to ensure that Ofcom's regulatory framework is tailored to the relevant circumstances and avoids undue burden on platforms wherever possible.

Implementation of Alternative Measures

We recognize that Ofcom has emphasised that the Codes consist of recommendations, and that platforms may comply by implementing alternative measures, as the ultimate goal of the online safety regime should be for platforms to implement mitigations that are suitable and effective for achieving their intended purpose. However, some measures throughout the Codes appear to be highly prescriptive, imposing particular measures or technologies rather than setting an outcome to be achieved and allowing regulated services the flexibility to implement proportionate solutions that are tailored to their platforms. For example, the Codes of Practice recommend that platforms utilise keyword detection to moderate fraud. [REDACTED - CONFIDENTIAL].³

While this measure, along with others in the Codes, is put forth as a recommendation, it is not clear how platforms can demonstrate that their alternative measures are achieving a similar (or better) outcome, or how Ofcom would assess whether those alternative measures are sufficient to comply with the Codes. Ofcom's recognition of alternative approaches to compliance is welcome, and while we appreciate Ofcom's desire to maintain flexibility in assessing them, guidance on the criteria to be applied to alternative measures would help platforms confidently plan for compliance with the Codes and streamline the implementation process across platforms. This would also prevent the Codes' "recommended" measures from effectively becoming required, as platforms would otherwise feel pressure to default to them to avoid risk of enforcement. Certain of these recommended measures will not be effective or practicable for all entities in scope of the OSA, and in fact may be suitable only for well-resourced incumbent platforms.

Volume of the Consultation Materials

We appreciate Ofcom's extensive research and transparent approach in developing the proposed Codes, and we welcome the opportunities to engage with Ofcom thus far. While we commend the thoroughness of Ofcom's approach, the sheer volume of consultation materials, exceeding one thousand pages, presents challenges.

Despite Pinterest's eagerness to engage and the generous consultation timeline provided by Ofcom, the task of reviewing, discussing, and providing constructive feedback within the given timeframe requires significant resources. It is important to

³ Our views on this recommendation are noted in further detail below, under Question 16.

acknowledge that in light of the extensive materials, the consultation may receive limited participation from medium-sized and smaller platforms due to capacity constraints and time limitations. Looking ahead, we would welcome thoughts on how to reduce the administrative burden on all companies, and any ways in which Ofcom can ensure that the smallest companies, which may not even have any employees in policy roles, can engage with this process to ensure the whole ecosystem is able to have a voice.

Responses to Consultation Questions

Approach to Illegal Conduct Codes of Practice

Question 15: Do you agree with our definition of multi-risk services?

We appreciate Ofcom's decision to account for a platform's risk profile as well as its size in determining whether additional requirements will apply. This approach helps ensure proportionality in the measures applied, and is an improvement over other regulatory frameworks, such as the DSA, where a platform's user base is alone sufficient to subject a platform to heightened compliance obligations. We are concerned, however, that in practice, the risk thresholds are set at a level that will place a disproportionate burden on many platforms, particularly those categorised as "multi-risk."

As currently drafted, the definition of a "multi-risk" platform in the Codes would likely lead to the most onerous measures being applied to platforms whose risk profiles do not merit such treatment. As a general principle, we do not agree that being at medium risk of at least two harms necessarily entails a significant overall risk, particularly when circumstances such as a platform's design, use cases, and existing mitigation measures are taken into consideration.

The draft Codes' approach to risk assessment is largely premised on the idea that certain platform functionalities carry increased risk. The mere presence of any functionality, however, should not necessarily lead to the conclusion that it poses a risk of harm, and indeed such an approach would likely lead to most platforms being categorised as "multi-risk."

While functionalities may have some inherent risk, that should not be the end of the analysis. The level of risk that functionalities actually pose to users is dependent upon a number of platform-specific factors. One such factor is the design of the platform. For example, a platform's recommender system may be tuned to show content that users find interesting or useful based on their activity on the service, as opposed to content that is viral or controversial. The former type of recommender system may pose less

risk than the latter type, and is beneficial to users in that it helps them identify information useful to them amidst a sea of irrelevant user-generated content. In addition, to the extent a functionality may carry some risk if left untreated, it is important to consider what controls and safeguards a platform already has in place to mitigate that risk, such as content safety guidelines and content moderation systems, as these will result in a smaller, residual risk for each functionality.

If platforms were to be deemed higher-risk based merely on their functionalities, we are concerned that such an approach would effectively eliminate risk profile as a factor in determining which obligations apply, as the outcome of OSA's framework would be similar to the DSA's in that risk is assessed purely on platform size, particularly considering that the Codes identify a platform's user base as factor that contributes to a high risk level. This approach would also give the Codes a more prescriptive character, since more platforms will need to meet the highest compliance burdens merely due to their functionalities, instead of adapting and responding appropriately to the actual risks posed by their particular service.

We would welcome a more holistic, nuanced approach that accounts for the role of platforms' design and risk mitigation measures. This would more accurately reflect the true risk profiles of regulated platforms, and help foster a safe and positive online experience while avoiding undue burden on platforms that already provide a low-risk service to their users.

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

Pinterest has the following comments on proposals in the draft Codes of Practice for user-to-user services:

- 4A. Having a content moderation function that allows for the swift take down of illegal content

This measure states that a platform's content moderation systems should be designed to "swiftly take down illegal content of which it is aware." We agree that harmful content should be actioned expeditiously and make best efforts to do so, but while turnaround time is a valuable metric, we do not believe it should be treated as a determinative factor in whether platforms' content moderation systems are effective.

Since the ultimate goal of content moderation is to prevent users' exposure to harmful content, we place more emphasis on how many impressions an item of content received before it was actioned (i.e., how many users viewed the content). This

emphasis on reach is aligned with the functionality of Pinterest’s platform, in which content is displayed to users based on their personal interests, rather than prioritising content that was posted recently or that has “viral” status.

In our global Transparency Report, we report data on the reach of Pins deactivated for violating each of our policy categories, calculated as the number of unique users that saw each deactivated Pin during the reporting period for at least 1 second before it was deactivated.⁴ In most cases, this data indicates that very few Pinner’s view content before it is deactivated.

As speed is not the sole measure of whether a content moderation system is effective, we believe this measure should be applied flexibly, with consideration given to whether platforms are creating a safe environment for users.

- 4C. Performance targets

Here, Ofcom proposes that platforms set performance targets for how quickly illegal content is removed and for the accuracy of content moderation decisions, adding that platforms should balance the need to quickly action illegal content against the need to make accurate moderation decisions. While we do not object overall to the setting of performance targets, we would caution against their use as a determinative measure of effective content moderation.

As Ofcom itself alludes to, performance targets for making decisions quickly as well as accurately can be in tension with each other, particularly where a small number of edge cases can skew average turnaround times because additional analysis is required to make an accurate decision.⁵ Strictly measuring platforms against performance targets can also create perverse incentives for platforms to over-enforce their policies in an effort to act “quickly” or to reject valid appeals in order to show higher “accuracy” rates, instead of engaging in careful review of nuanced content.

In enforcing this measure, we encourage that Ofcom allow companies to set performance targets that make sense in the context of their own platforms and content moderation systems, and that performance against them should be measured with an eye towards the overall success of content moderation efforts, rather than a reductive consideration of whether targets were met.⁶

⁴ <https://policy.pinterest.com/transparency-report>

⁵ Also see our response to Code 4A, on the use of swift content moderation actions as a performance metric generally.

⁶ For example, Pinterest’s content moderation efforts place a strong emphasis on reach, i.e., how many impressions an item of content receives before it is actioned. See response to Code 4A.

- 4H. Detection of CSAM URLs

This measure states that platforms should employ automated tools to detect URLs which have been identified as hosting CSAM. [REDACTED - CONFIDENTIAL].

- 4I. Keyword detection regarding articles for use in frauds

Ofcom proposes that platforms employ keyword detection technology to analyse content on their service to identify content that is “likely to amount to an offence concerning articles for use in frauds.” [REDACTED - CONFIDENTIAL].

Pinterest’s Community Guidelines prohibit harmful or deceptive practices such as exploitative financial practices, counterfeiting documents, and impersonation; as well as any content that reveals personal identification, private contact information, online login information, or financial information. This content can be reported by users [REDACTED - CONFIDENTIAL].

[REDACTED - CONFIDENTIAL]

- 5C. Appropriate action – sending indicative timelines

This measure states that platforms should acknowledge receipt of complaints regarding content and “provide the complainant with an indicative timeframe for deciding the complaint.”

Depending on the complexity of the report at issue, there may be a range of turnaround times that are within industry standards. A given report may be processed later than this average range, or much sooner. We believe that principles of transparency would be better served by allowing users to track the status of their reports and receive notifications as that status is updated. Taking such an approach would also be an opportunity for Ofcom to ease platforms’ compliance burden by aligning with measures required by international regulations, namely the DSA, which does not require complainants to be provided with a timeframe for deciding their complaint, but does contain requirements concerning notifications to users on the status and outcome of complaints.⁷

⁷ See DSA Arts. 16, 17, 20.

- 5I. Dedicated reporting channels

In this measure, Ofcom proposes that platforms maintain a dedicated reporting channel for a group of designated trusted flaggers. Pinterest is amenable to prioritising reports from trusted flaggers, and appreciates Ofcom’s proactive approach in identifying the trusted flaggers at this stage. Rather than requiring a dedicated channel, however, we view this as another area for potential alignment with the DSA, under which trusted flaggers may use the same channel as users for reporting illegal content and the only requirement to platforms is that these reports be given priority.⁸ We believe this approach is sufficient to give due priority to notices from trusted flaggers, without imposing any additional burden on platforms to maintain additional communications channels.

- 9A. User blocking and muting

Ofcom proposes that users should have the option to “block” or “mute” other users. When another user is blocked, the blocked users should not be able to send direct messages to a blocking user and vice versa, the blocking user will not encounter content posted by a blocked user, and if these users were previously connected, they will no longer be connected. Where another user is muted, the muting user will not encounter any content posted by the muted user on the service, unless the muting user visits the profile of the muted user.

While we are aligned with the goals of these features, we believe this specific proposal is overly prescriptive, and its objectives are met by Pinterest’s existing safety measures. Pinterest has a “block” feature which allows a Pinner to prevent any other user from following them, messaging them, or interacting with their Pins. Pins from blocked users may still appear in searches and feeds, though, and if the blocking user had saved any Pins from the blocked user to their profile, they would need to be manually removed.⁹

We believe Pinterest’s blocking function achieves the intended purpose of avoiding unwanted interaction with another user. While it is still technically possible to encounter content saved by the blocked user in searches or in feeds, it would be extremely unlikely. As opposed to platforms that are built around networks of friends, family members, or other real-world communities, Pinterest recommends content to users from a corpus of billions of Pins based on their interests, so the odds of viewing content from any particular Pinner (who is not being followed) is exceedingly low. Accordingly,

⁸ See DSA Art. 22.

⁹ For more information, please see our Help Center at: <https://help.pinterest.com/article/block-or-unblock-someone>

requiring Pinterest to develop a “mute” function would be disproportionate in light of the functionality of Pinterest’s platform.

- 10A. Removing accounts of proscribed organisations

This measure states that platforms should remove a user account from their service when they have reasonable grounds to infer it is operated by or on behalf of a proscribed organisation, based on content reported by a user or otherwise reviewed by the platform. The determination that the account is operated by or on behalf of a proscribed organisation can be based on at least two of the username, profile image, or profile information; or it can be inferred if a “significant proportion of a reasonably sized sample” of the latest content posted by the user consists of proscribed organisation content.

Pinterest’s Community Guidelines prohibit content and accounts that encourage, praise, promote, or provide aid to dangerous actors or groups and their activities, including extremist, terrorist, and criminal organisations. We work with industry, government and security experts to help us identify such groups, and this content will be actioned accordingly if it is reported or otherwise identified by Pinterest.

While we are committed to tackling content that supports proscribed organisations, we are concerned that this measure imposes proactive monitoring obligations on platforms, particularly the proposal regarding sampling of recent content to review for proscribed organisation content. Proactive monitoring of content across an entire platform would be extremely burdensome and disproportionate, particularly for platforms such as Pinterest which are not at high risk for violent or extremist content.

Particularly in light of the government’s commitments to upholding the eCommerce Directive’s liability protections and preserving its “notice and take down” regime,¹⁰ we propose that Ofcom clarify the extent to which this measure imposes proactive monitoring obligations, as opposed to merely recommending criteria on which to evaluate whether content is being operated by or on behalf of a proscribed organisation. If indeed Ofcom is proposing proactive monitoring, we strongly recommend that this measure should be applied proportionately, based on a platform’s risk for this specific harm.

¹⁰ For example, see: <https://www.gov.uk/guidance/the-ecommerce-directive-and-the-uk>

Default Settings and User Support for Child Users

Question 31: Do you agree with our proposals?

While we agree with many of these proposals, we believe some of them may not be additive to existing protections that platforms, including Pinterest, already have in place for teens, and in some cases may actually detract from teen safety. Accordingly, we encourage that such measures only be applied where they would be effective in promoting teen safety overall in the context of a given platform.

Among the recommendations for default settings, Ofcom recommends that platforms do not display lists of users that teens are connected with.¹¹ For Pinterest and other platforms that do not include a user's age in their profile, this feature may actually increase risk to teens, as the absence of Follower and Following lists would identify them to other users as being under 18 years old.

Ofcom also recommends that teens be provided with a notice every time they receive a message from another user for the first time, reminding them that this is the first direct communication with the other user and advising them of their options regarding receiving the message or blocking the user. This measure would not be effective on Pinterest, however, as a mutual following relationship is required for a teen to receive a direct message from another user, making such a notice redundant.

Enforcement Powers

Question 53: Do you have any comments on our draft Online Safety Enforcement Guidance?

Section 29.14 states that “[w]hile we will consider what is reasonable on a case-by-case basis, all services should expect to be held to full compliance within six months of the relevant duty coming into effect.”

For most if not all service providers, compliance with the Codes' requirements will entail substantial changes to their product, policies, and content moderation systems. As such, compliance will require significant planning and resource allocation by numerous internal teams at each company. Depending on the duty at issue, six months may be a very short turnaround to effectively plan, design, develop, test, and implement the required changes. Accordingly, we encourage Ofcom to remain flexible in determining

¹¹ On Pinterest, this would consist of “Following” other users or having “Followers.” These relationships may not be mutual. For example, a Pinner may follow another user without being followed back, but they may also follow each other.

whether individual platforms' compliance timelines are reasonable, particularly considering the complexity of the requirement and the resources of the platform in question.

Conclusion

Pinterest is grateful for the opportunity to contribute to the important work being done by Ofcom in developing the Codes. We welcome further engagement as part of this process, and if it would be of assistance, we would be happy to elaborate on or discuss any of the points raised in this response or any other matters of interest to Ofcom.

* * *