Response to Ofcom Call for Evidence
Second Phase of Online Safety Regulation: Protection of children
21 March 2023

I.   Introduction

On 10 January 2023, the Office of Communications (Ofcom) published its call for evidence soliciting evidence on risks of harms to children online and how they can be mitigated, in order to assist Ofcom as it prepares codes of practice on protection of children and risk assessment guidance under the Online Safety Bill.

Pinterest shares Ofcom's objective to combat harmful content and promote a safe online experience for teens, and we welcome the opportunity to provide information about our content safety policies and content moderation measures, particularly as they relate to younger users.

A.   Introduction to Pinterest

Pinterest is a visual discovery engine where 450 million users (or "Pinners") around the world go to get inspiration for their lives. Pinterest shows users visual recommendations called Pins. Pins are created by both individual and business users by uploading photos or videos or bookmarking content from the web, and providing a text caption.

Pinterest selects Pins to recommend to users based on interests they have communicated to Pinterest, using a variety of sources to determine what to recommend to a user, including the user's past search and browsing behavior on Pinterest and the interests they select when signing up for the service. A Pinner who is interested in vegetarian cooking will see more Pins linking to recipes, while one who shows an interest in travel or modern interior design will see a feed more tailored to those interests. Pinners can then save and organize these recommendations into collections called boards.

While Pinners can interact by sending Pins, creating group boards, and sending messages, communication is not the purpose of the service. Instead, Pinterest's mission is to inspire users to live a better life and provide the means for them to do more than just dream. Browsing and saving visual ideas on Pinterest's service helps Pinners imagine what their future could look like, which helps them go from inspiration to reality--to cook that recipe, throw that party, buy that outfit, and so on.

Pinterest is free to use, and generates revenue through the sale of digital advertising. Pinterest provides a suite of advertising products and businesses of all sizes and across many industries can use Pinterest to achieve a diverse set of goals, from building brand awareness, to increasing online traffic, to driving sales.

B. Pinterest's Commitment to User Safety

Pinterest has a longstanding commitment to creating a positive corner of the internet. We continually invest in our policies, products and partnerships to support the safety and wellbeing of our community because it's the right thing for the people on our platform.

We use various inputs to assess the risk from harmful content, including reports from users, information from outside experts and organizations, and trends observed on other platforms. We then mitigate those risks in a number of ways, including setting policies that clearly prohibit harmful content, deploying robust reporting and content moderation tools, providing wellness resources to users, and partnering with leading outside experts and organizations to inform our policies and practices. All of these practices are constantly evolving to keep up with new behaviors and trends.

We share the UK government's commitment to addressing online safety and view the Online Safety Bill as a continuation of our commitment to making Pinterest an inspiring and welcoming place for everyone. We also agree with the UK government that "online safety is a shared responsibility between companies, the government and users." We believe it is important for platforms to collectively tackle illegal and harmful content and prevent it from simply moving between platforms. We hope that the Online Safety Bill in the UK and Ofcom, as the proposed independent regulator, achieve a system which has user safety and risk management at its heart. Cooperation between platforms in achieving online safety is critical in our view, as a greater degree of inter-platform collaboration will be essential to prevent the spread of illegal content online.

We are happy to continue engaging with the UK government and Ofcom on the Bill, particularly to help ensure that codes of conduct and risk assessment requirements are aligned with requirements in other jurisdictions and proportionate to platforms' size and risk profile, thereby allowing small- and mid-sized platforms to remain competitive with their larger counterparts.

II.   Protection of Teen Users on Pinterest

We seek to build a safe and inspiring platform for all users through a combination of broad-reaching content policies and robust moderation efforts to identify and deactivate harmful content (see Section III., below). Teen users certainly benefit from these efforts to keep all Pinners safe, but we also have measures in place that are particularly focused on protecting them. These include provisions in our Community guidelines that are focused on teens, age gating technology to block underage users, high-privacy default user settings, safe messaging features, and a ban on targeted ads for teens in the UK.

A. Community guidelines Related to Teen Users

Our Community guidelines broadly prohibit various types of content that are harmful to users of all ages, including, among others: content promoting suicide or self-harm; hateful or violent content; content promoting disordered eating; and health, climate, or civic misinformation.

While all content that violates our Community guidelines would be considered harmful to teens, there are specific areas of our policies that focus on harms that uniquely affect younger users. For example, our Community guidelines ban bullying and harassment, as well as harmful pranks or challenges that risk imminent physical harm or extreme emotional distress, especially if showing or encouraging the participation of minors. *[REDACTED - CONFIDENTIAL]*

We also prohibit any content containing sexualization or sexual exploitation of minors, like grooming, sexual remarks or inappropriate imagery. We have a strict zero-tolerance policy for any content or behavior that may exploit or endanger minors and when we find such content, we will not only deactivate it, but report it to the relevant authorities where appropriate. To help improve our ability to combat the sexual exploitation of minors, we recently launched two new features. First, we are enabling users to report boards and accounts in their entirety (in addition to individual Pins), so that we can receive reports of individually innocuous Pins that are collected in a policy-violating manner. Second, we have expanded the list of reasons in our in-product reporting to include intentional misuse of content involving minors, to encourage users to report exploitative content and behavior, even if the content alone does not contain nudity or explicit sexual references.

B. Age Gating Technology

Pinterest requires users in the UK to be 13 or older to create an account. At account registration, we require new users to provide their date of birth, using a "neutral" prompt that does not indicate that the user needs to be a certain age to join Pinterest. When a user inputs an age below 13, Pinterest informs them they are not eligible to join, using a neutral message to discourage any false declarations of age. *[REDACTED - CONFIDENTIAL]*
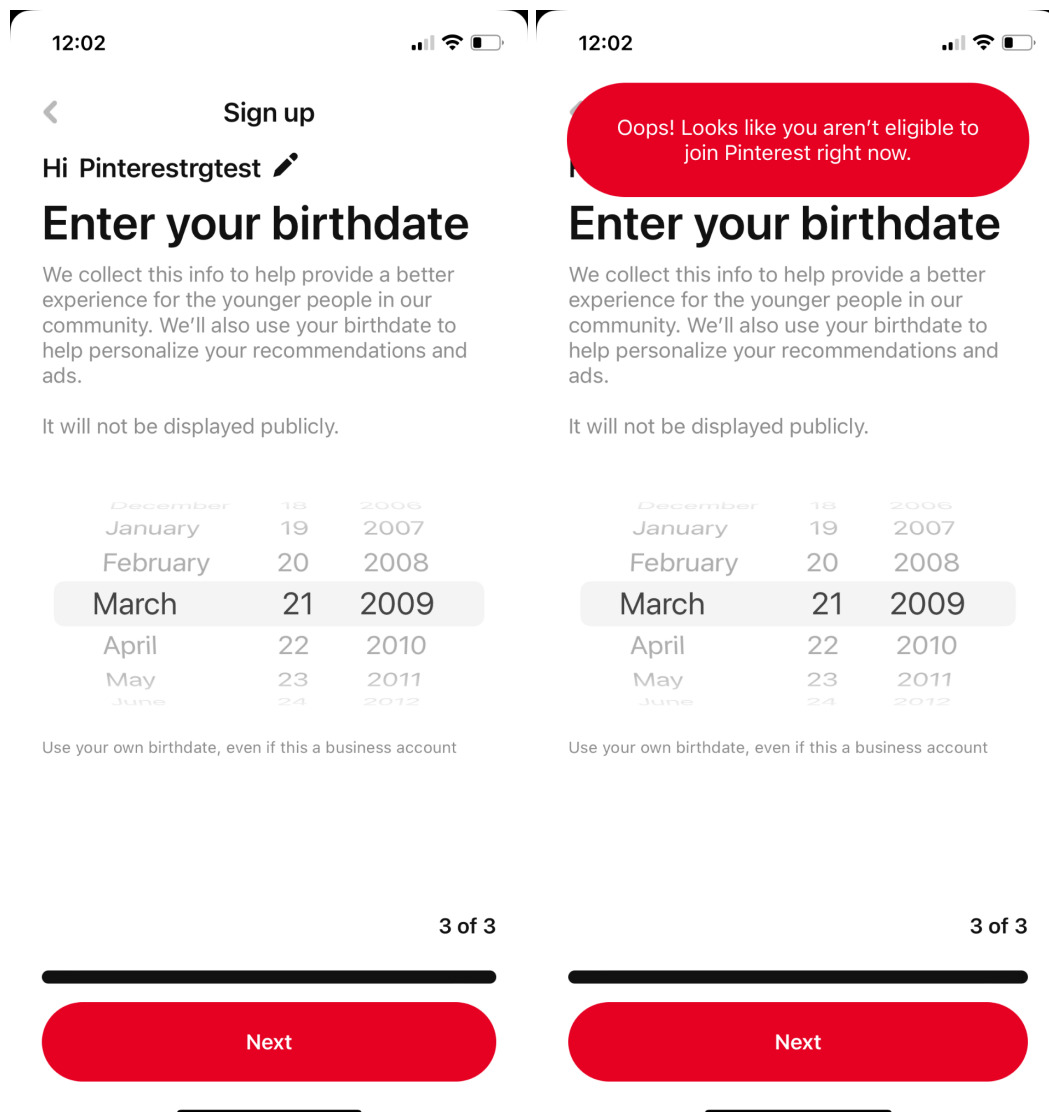
[REDACTED - CONFIDENTIAL]

We take age assurance measures seriously and continue to monitor best practices in this area. These are industry-wide challenges as technological solutions continue to evolve, and we remain committed to exploring the best ways to combat this issue. There are active debates regarding whether age assurance regimes, while they may prevent circumvention by underage users, may introduce undue burdens on an internet user's privacy by preventing them from visiting a site if they wish to withhold information from an internet platform regarding their identity. Such concerns have been articulated by, amongst others, the Information Commissioner's Office in its opinion on age assurance under the Children's Code,[1] the Commission nationale de l'informatique et des libertés (CNIL) in its guidance on age verification,[2] and the Open Rights Group's commentary on the Online Safety Bill.[3]

As these debates continue, we will continue supporting cross-industry efforts to develop technological solutions to the challenges posed by age assurance, and thereby enhance the safety of younger users on the internet.

C. High-Privacy Default Settings for Teen Users

For users between 13 and 17 in the UK, we provide several high-privacy default user settings that are designed to help safeguard their online experience.

For these users, privacy personalization options are designed to default to "off" and cannot be changed. These users will not receive personalized Pinterest recommendations based on their off-Pinterest activity, and we will not use their Pinterest activity to advertise Pinterest to them on other services.

For these same users, the "Search Privacy" setting is defaulted to "on." The "Search Privacy" setting means that Pinterest users have a tag added to their profiles which tells Google, Bing, or other search engines not to include their profile information in search results.

In addition, teen users will have their notifications defaulted to "off" (excluding routine account service messages), but can choose whether to receive additional notifications through their Privacy and Data Settings.

---

[1] https://ico.org.uk/media/about-the-ico/documents/4018659/age-assurance-opinion-202110.pdf
[2] https://www.cnil.fr/en/online-age-verification-balancing-privacy-and-protection-minors
[3]
https://www.openrightsgroup.org/app/uploads/2022/02/ORG-response-to-ICO-the-use-of-age-assurance.pdf

Finally, if a teen user attempts to add information to their profile, Pinterest presents a just-in-time notice, notifying teens in plain language that any information they add will be shared publicly and that they should not share sensitive or contact information.

D. Safe Messaging Features

We provide several features that allow Pinners to safely send and receive messages, including some that are available to all users, along with others that are particular to teens.

A teen user on Pinterest can only receive messages from users who are known to them. If the sender and recipient follow each other on Pinterest, if the sender is in the contacts that the recipient has synced from their phone or another service, if the sender and recipient are collaborating on a board together, or if the sender and recipient have previously messaged with each other, the message can be received. If the recipient teen only follows the sender's account, then the recipient needs to approve the message by accepting a contact request. Messages sent from any other user will be blocked.

[REDACTED - CONFIDENTIAL]

When any user receives a message from someone outside of their network (i.e. the sender is not someone (i) whom they follow, (ii) they have messaged before on Pinterest, (iii) they collaborate on a board with, or (iv) in their synced contacts), they receive a contact request and do not need to view the message itself before dismissing the contact request. If they decline the request, they are presented with the option to block or report that person. The contact request also warns users not to share confidential information.

We also make it easy for all users to block other users, and/or report them for sending messages that violate our content policies. This can be done either in the messaging tool on our app or website, or through our Help Center.

E. No Targeted Ads to Teen Users

Pinterest has prohibited all targeted paid advertising to users between the ages of 13 and 17 in the UK. We understand that teens' personal data merits specific protection under the UK GDPR and the ICO's Age Appropriate Design Code, and eliminating this type of advertising is one of the measures we have taken to ensure those protections.

III.   Protecting Users from Harmful Content

We work hard to identify and deactivate harmful content from Pinterest, starting with setting policies that clearly prohibit harmful content. We enforce these policies through robust content reporting and moderation tools which we constantly iterate on to make technical improvements and keep up with new behaviors and trends. We also partner with leading experts to support this work and help us to create a more positive corner of the internet for users.

A. Community guidelines

Pinterest's mission is to bring everyone the inspiration to create a life they love. To help cultivate a safe and inspired community, Pinterest develops and enforces content policies that help ensure Pinterest is a positive place where people can find real-life ideas for what to try next, cook next, wear next, or do next. Pinterest has strong Community guidelines, developed in collaboration with subject matter experts, that outline what is and is not allowed on Pinterest: https://policy.pinterest.com/en/community-guidelines

Pinterest

Community guidelines

Our mission

Pinterest's mission is to bring everyone the inspiration to create a life they love. That being said, not all content is inspiring - so we have community guidelines to outline what we do and don't allow on Pinterest. These guidelines are our acceptable use policy, so if you find content that shouldn't be on Pinterest, please report it to us . We use your reports to learn and evolve our standards, and work with subject matter experts to inform and update our guidelines. You can learn more about how we enforce these guidelines in our biannual transparency report .

Content safety

Pinterest isn't a place for antagonistic, explicit, false or misleading, harmful, hateful, or violent content or behavior. We may remove, limit, or block the distribution of such content and the accounts, individuals, groups and domains that create or spread it based on how much harm it poses.

We're committed to presenting you with clear and transparent expectations that are easy to understand and follow. If you have questions or encounter problems on Pinterest, please contact us .

All users agree to abide by the Community Guidelines as part of the account registration process, and they are communicated in simple, clear language that is accessible to all users, including teens.

Among other harmful content, our Community Guidelines broadly prohibit all of the types of content that are expected to be listed as "primary priority" and "priority" harmful content under the Online Safety Bill, including:

- Pornography and sexually explicit content;
- Content promoting self-harm;
- Content promoting disordered eating;
- Content promoting suicide;
- Abuse, bullying, and harassment;
- Health misinformation; and
- Hateful or violent content.

Our Community guidelines are applied to all users and content on Pinterest to promote a safe and positive environment for everyone, but certain areas of our policies prohibit

content posing unique harms to younger users, such as sexual exploitation of minors, and harmful pranks and challenges (see Section II.A., above).

B. Manual Content Moderation

All Pins, boards, accounts, links, messages, and comments on Pinterest can be easily reported to Pinterest's moderation team via either the reporting feature on the Pinterest app or website, or a web form that can be accessed by anyone, whether or not they are a Pinterest user.

Pinterest's in-product reporting tool is available on all versions of Pinterest (web, iOS, or Android), and can be accessed via the ellipsis ("...") menu on each Pin, in either the grid view (i.e. the home feed, or search results) or closeup view (i.e. when viewing a single Pin).

After clicking "..." the user is presented with a menu that includes "Report Pin." Next, the user selects the reason for their report.

Report Pin

○ **Spam**
Misleading or repetitive posts

○ **Nudity, pornography or sexualized content**
Sexually explicit content involving adults or nudity, non-nudity or intentional misuse involving minors

○ **Self-harm**
Eating disorders, cutting, suicide

○ **Misinformation**
Health, climate, voting misinformation or conspiracies

○ **Hateful activities**
Prejudice, stereotypes, white supremacy, slurs

○ **Dangerous goods**
Drugs, weapons, regulated products

○ **Harassment or criticism**
Insults, threats, non-consensual nude images

○ **Graphic violence**
Violent images or promotion of violence

○ **Privacy violation**
Private photos, personal information

○ **My intellectual property**
Copyright or trademark infringement

Cancel   Next

Depending on the reporting reason selected by the user, they may be prompted to select a specific type of violation, and are provided with an explanation of the types of content Pinterest does and does not remove in that category. The user then clicks "Report" to finalize their report.
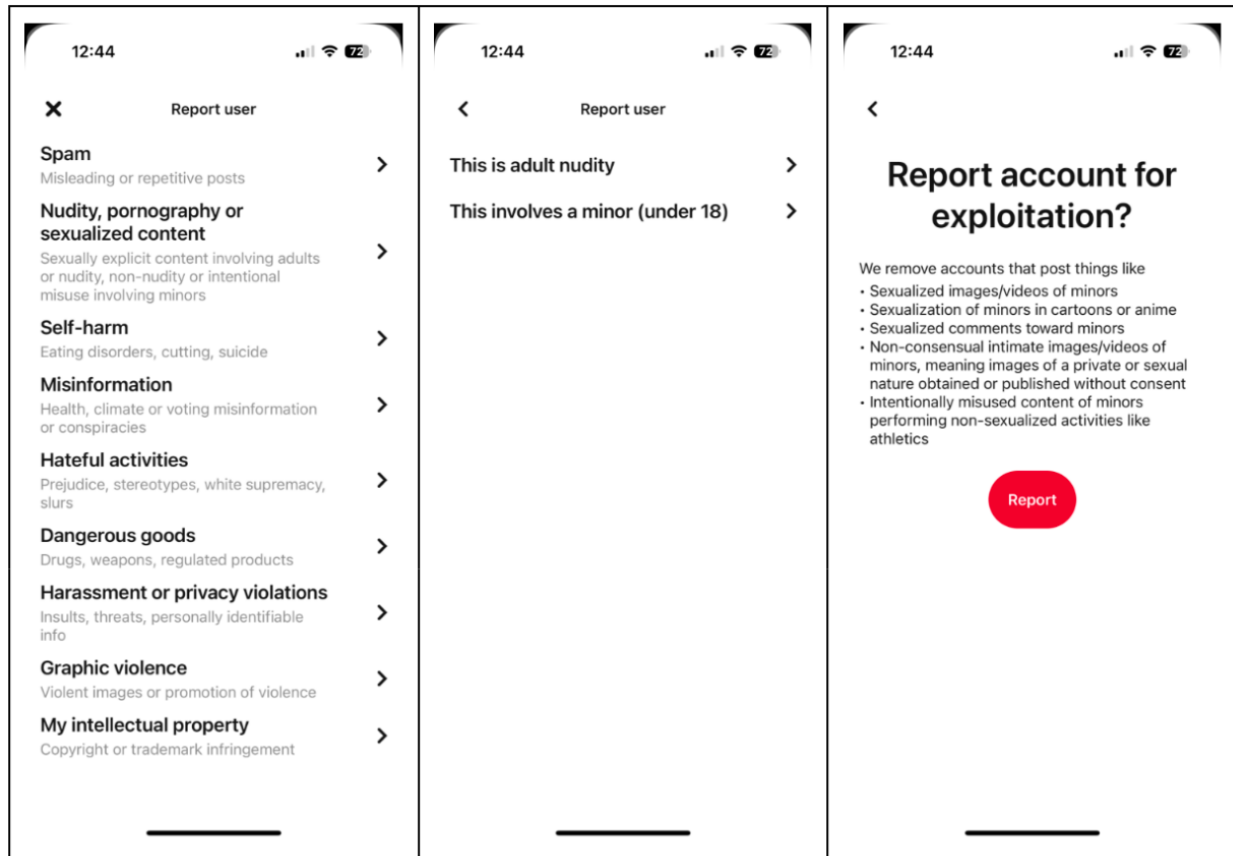
While this reporting flow is only available to users within the product, any person may report something without being logged into Pinterest by using the contact form available in the Pinterest Help Center (https://help.pinterest.com/en/contact), which is similar to the in-product reporting flow. We also review reports received by email and other means.

The majority of content reported to Pinterest via the in-product reporting flow or the Help Center is manually reviewed by a global team of agents which operates 7 days a week (as discussed in Section III.C. below, some reports are handled by automated tools). If the content is deemed to be in violation of Pinterest's Community guidelines, we will take action on the content as well as the accounts, individuals and groups that create or spread it. In the event that a user believes that their content was erroneously deactivated, they may appeal the decision via the Pinterest Help Center at https://help.pinterest.com/en/contact.

Our agents also conduct detailed manual sweeps for harmful content when we're alerted to concerning new trends with other platforms. We will also block external domains that are dedicated to topics that violate our policies, such as self-harm.

We continuously provide training for our review team members in an effort to build the expertise and knowledge critical to content moderation. *[REDACTED - CONFIDENTIAL]*

C. Hybrid and Automated Content Moderation Methods

In addition to manual review, Pinterest deactivates Pins through automated tools, as well as a hybrid approach that combines elements of both approaches, in order to identify and deactivate harmful content at scale.

We use automated tools that rely on a combination of signals to identify and moderate content that is potentially violative. Our machine learning models assign scores to each image added to the platform and using those scores, our automated tools can then apply the same enforcement decision to other Pins containing the same images.

Pinterest uses the latest modeling techniques (use of sequences of instructions for solving problems) and continuously iterates on these models by adding new data and exploring new technical breakthroughs to either maintain or improve their performance over time to effectively address unsafe content. To control the quality of a machine learning model, we evaluate the performance of that model through offline analysis and online experiments.

Hybrid deactivations include those where a human determines that a Pin violates policy, and automated systems help expand that decision to enforce against machine-identified matching Pins. We'll also have human moderators review the most sensitive and risky content identified by our automated systems, such as CSAM, to ensure that it is handled properly, including referrals to law enforcement where appropriate.

We also maintain a list of sensitive terms which is used to block search results or prevent content from appearing in recommendations where it may violate our policies, including terms associated with self-harm, suicide, drug abuse, and eating disorders. In response to searches containing these terms, we also display an advisory that connects users with resources if they or someone they know are struggling. Our sensitive terms list is continually expanded as we identify online trends.

The mechanisms used to address different potential policy violations may vary based on the state of available technology, the volume of violative content and other factors such as the complexity of evaluation. We continue to iterate and evolve our tools and expect ongoing improvements going forward.

D. <u>Content safety partnerships</u>

While Pinterest has robust tools for tackling harmful content, we also recognize that this effort is complex and always evolving. That is why Pinterest works with others in the industry as well as expert organizations, government agencies, and nonprofits to leverage their expertise, get feedback on Pinterest's approaches, and onboard their expert content to Pinterest. These partnerships also help us support expert research and development to help the industry combat harmful content and make the internet a safer place for everyone, no matter what app or website they use.

We partner with leading experts in different subject areas who have demonstrated global knowledge of trends in the online environment and have provided incredibly valuable feedback on best practices. These include, among others, the Weprotect Global Alliance and the Technology Coalition (child sexual exploitation), Beat (disordered eating), and Samaritans (self-harm and suicide). We are also part of Samaritans' Online Excellence Programme, a three-year industry-wide program to promote consistently high standards across the sector in relation to self-harm and suicide content. The program includes a research and insight program, industry guidelines to support sites and platforms in managing self-harm and suicide content online using safe and sensitive approaches, an online harms advisory service, and a hub of online safety resources.

Pinterest also partners with Graphika, a leading third-party network analysis firm. Graphika leverages AI to study the social media landscape and its clients include multiple platforms, such as Facebook, Google, and Twitter, which helps it spot harmful content across the online ecosystem. Through our partnership with Graphika, we remain informed about trends on other platforms, which allows us to mitigate risk before it even arises on Pinterest.

We also regularly engage, individually and with other online platforms, in stakeholder discussions around key developments in this area. For example, we are a member of the Digital Trust & Safety Partnership, which brings together a number of leading technology companies who are committed to developing industry best practices and providing objective and measurable third-party assessments of members' trust and

![Pinterest](Pinterest logo)

safety practices. The Partnership engages with consumer and user advocates, policymakers, law enforcement, relevant NGOs and various industry-wide experts.

IV.   Combating Suicide and Self-Harm Content

In light of the questionnaire on suicide and self-harm content promulgated by Ofcom and the Accelerated Capability Environment (ACE), as well as its relevance to younger users, we would like to offer some additional information on the topic of how Pinterest tackles suicide and self-harm content. It should first be noted that content promoting suicide or self-harm is strictly prohibited by our Community guidelines, and accordingly, it is subject to all of the content moderation measures described above. However, as detailed herein, there are certain considerations that are unique to this issue.

When discussing this topic, we would be remiss if we did not mention the recent Molly Russell inquest. Our thoughts are with the Russell family and with other families who have been affected by suicide or self-harm, and Molly's story has reinforced our commitment to create a safe and positive space online.

As stated in our 8 December 2022 response to the Coroner's prevention of future deaths report, Pinterest is undertaking several voluntary actions to accelerate our ongoing content safety improvements, including:
- Making changes to our self-harm policy to ensure stricter enforcement of certain categories of content, such as removing references to self-harm or suicide in artwork, memes, or jokes;
- Partnering with a third party content checking service with the aim of providing independent testing of our progress in our moderation efforts with respect to self-harm and suicide content on Pinterest;
- Developing and testing tools to further limit the distribution of sad or depressive content on Pinterest to teens;
- Consulting with mental health experts to ensure that we are delivering the best possible resources to Pinners who search for self-harm or suicide related content; and
- Continuing to work through the challenges of age assurance with experts, legislators and the rest of the market.

Of these measures, the changes to our self-harm policy have already been enacted, *[REDACTED - CONFIDENTIAL]*. Consultation with mental health experts and evaluating age assurance technology and regulation are ongoing efforts.

A. Distinguishing Harmful from Supportive Content

Pinterest isn't a place for content that displays, rationalizes or encourages suicide or self-injury. We'll limit the distribution of or remove content including self-harm instructions, graphic or otherwise triggering imagery or descriptions of self-harm, promotion of self-harm, mocking of people who self-harm or who have attempted or died by suicide, images of accessories used to self-harm, negative self-talk and insensitive humor about self-harming behavior, suicide pacts, challenges and hoaxes.

We recognize, however, that some content relating to suicide or self-harm has the potential to support users' journey of healing or recovery by contributing to a content ecosystem wherein users can safely explore their feelings and share their experiences with their own mental health. In these cases, removing or limiting the distribution of such "helpful" content would be over-actioning or have otherwise unintended consequences that would create a negative user experience. Accordingly, we distinguish between "harmful" suicide or self-harm content, and content that may be considered "helpful" or "support" content.

[REDACTED - CONFIDENTIAL]

In making decisions about what types of content may be considered harmful and therefore policy-violating as opposed to what may be considered helpful or supportive content, we may consider potential risks to both the users creating, saving or otherwise engaging with this content as well as risks to users who may unintentionally view this content and the larger content ecosystem of the platform itself. To help make these distinctions, we consult with experts in this space, including through our ongoing partnership with Samaritans (see Section III.D, above).

B. Providing Mental Health Support and Resources

In addition to removing harmful content, we take numerous steps to support mental health and direct users to supportive resources, including suicide and self-harm hotlines.

When users search for a blocked term related to suicide or self-harm, we show a suicide and crisis helpline advisory with no search results. For less sensitive terms, we show the same advisory and include search results, but limit the use of autocompletion and search recommendations for these terms.

Pinterest

> ## Can we help?
> Pins about this topic can be upsetting for people who view them. If you or someone you know is going through something difficult, or struggling with thoughts of suicide or self-harm, you're not alone. Helplines can provide free, confidential and immediate support.
>
> **Call or Text 988 Suicide & Crisis Lifeline**
> 988
>
> **Visit 988 Suicide & Crisis Lifeline**
> 988lifeline.org

We also show advisories to users who view a board containing "helpful" self-harm content. This advisory points them to these same external mental health resources.

When we take action on a user's content under our self-harm policy, whether that be removing it or limiting its distribution on the platform, the user receives an email notifying them of this and providing resources that they can access should they be struggling with suicidal thinking or experiences of self-harm. These resources include suicide, self-harm, and eating disorder hotlines where they can get free, confidential and immediate support.

In instances where a user searches for terms that deal with difficult emotions but aren't policy-violating and don't have self-harm or suicide context, Pinterest's compassionate search feature is shown.
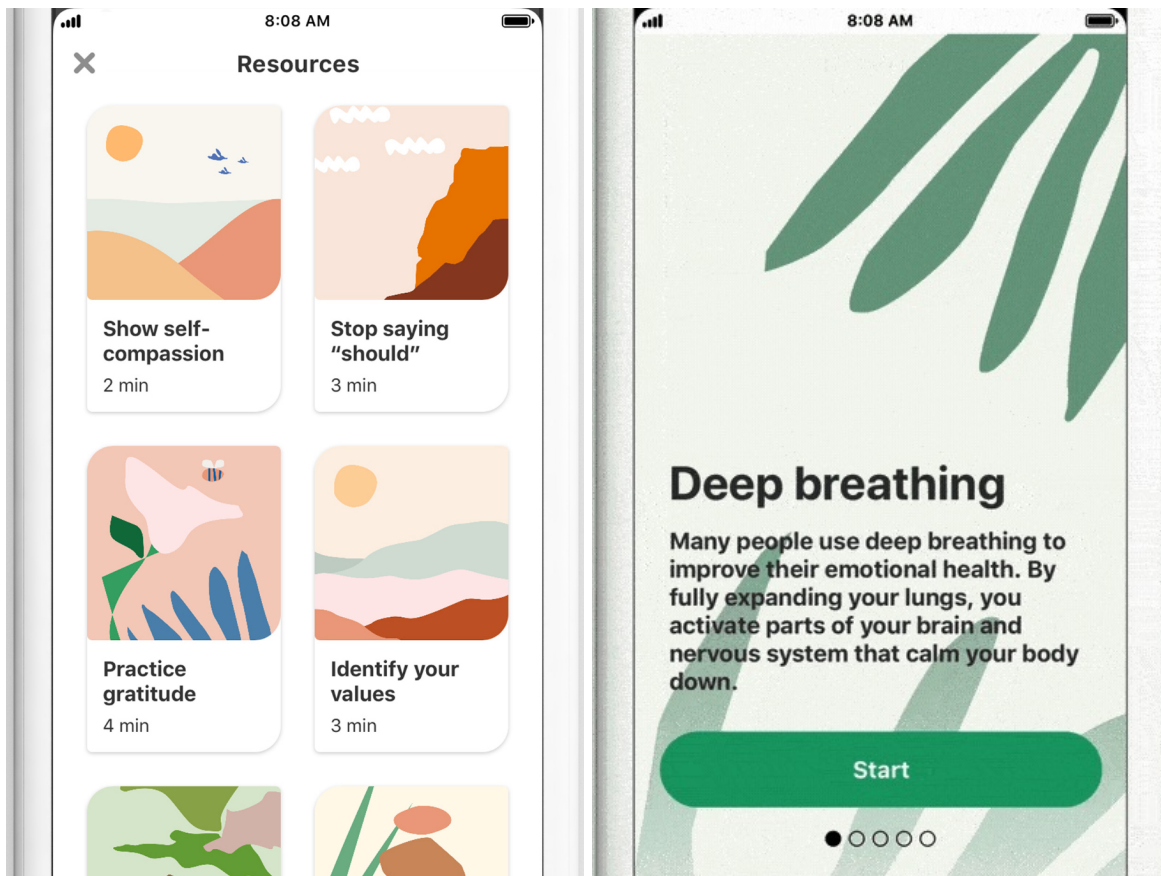
Compassionate search offers a collection of interactive activities, developed with external experts, to help Pinners safely access supportive resources to improve their mood if they are feeling stressed, anxious, sad, or trying to manage difficult emotions. For example, if Pinners search for content that indicates they may be feeling down like "sad quotes" or "work anxiety," they will see a prompt asking if they'd like to be guided through steps to practice self-compassion.

C. **Challenges to Tackling Self-Harm Content**

There are several content moderation challenges that are specific to this area. *[REDACTED - CONFIDENTIAL]*

Also, determining whether content is harmful can be a nuanced analysis, requiring reviewers to exercise judgment and take context into account. This requires time and expertise, but it is a necessary analysis to undertake so that we avoid removing content that can aid in users' recovery. Furthermore, there are related types of prohibited

content that overlap with suicide and self-harm but need to be treated differently under our policies (for example, graphic violence).

Despite these challenges, combating self-harm remains a priority for us. Our policies and moderation practices will continue to evolve to keep up with new behaviors and trends to maintain a positive and safe environment for users.

V.   Decision-making Structures and Accountability

Multiple teams across Pinterest work together on making content safety decisions. For example, our Policy, Legal, Trust & Safety Product and Engineering, and Trust & Safety Operations teams are each responsible for parts of the company's content safety efforts and work together, along with company leadership, to set priorities and implement measures to protect Pinners. In this structure, multiple overlapping stakeholders can work together to ensure that our content safety measures are simultaneously effective, scalable, and legally compliant.

We measure the effectiveness of our content moderation in multiple ways. As reported in our biannual Transparency Report, we measure the reach of deactivated content for each policy area. Reach is a metric that measures how many users saw a Pin prior to a Pin being deactivated for violating our policies. To calculate this metric, we look at each policy-violating Pin deactivated in a reporting period. Then we count the number of unique users that saw each Pin during the reporting period for at least 1 second before it was deactivated, and report the percentages of deactivated Pins in each category that were seen by 0 people, 1-9 people, 10-100 people, and >100 people. https://policy.pinterest.com/en/transparency-report

We look at a wide variety of other metrics to understand the effectiveness of our content moderation efforts. *[REDACTED - CONFIDENTIAL]*

We also look at a number of metrics to help understand the speed and accuracy of the content moderation actions we take. *[REDACTED - CONFIDENTIAL]*

[REDACTED - CONFIDENTIAL]

In addition to measuring the moderation of harmful content, we also seek to measure the positive impact that Pinterest has on users. For example, in a recent study conducted with University of California, Berkeley's Greater Good Science Center, researchers found that using Pinterest helped university students to buffer against

negative emotions — even while they were studying for stressful exams. Subjects who spent just 10 minutes per day seeking out inspirational content on Pinterest reported less burnout, less stress, and improved social connectedness. The findings were replicated in multiple countries, including the UK.
https://business.pinterest.com/en-us/blog/positivity-research-pinterest-berkeley/

VI.   Conclusion

Thank you again for the opportunity to contribute to the important work being done by Ofcom to prepare for enforcement of the Online Safety Bill. Pinterest welcomes further engagement on this issue, 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.