



2024-2025 Annual Online Safety Report for Facebook under Code of Practice for Online Safety in Singapore

Introduction

Singapore Code of Practice for Online Safety

The [Code of Practice](#)¹ (COP) contains ex-ante requirements to ensure that designated social media services have systems in place to minimise harmful content on the platform, calling out in particular measures for young users under 18; it also includes transparency reporting requirements. The COP, issued by the Infocomm Media Development Authority (IMDA), specifies outcomes that designated Social Media Services (“Service”), designated under section 45K(1) of the Broadcasting Act 1994, have to meet, to enhance online user safety, particularly for young users, and curb the spread of harmful content on the platform.

Background

This report is developed by Meta Platforms Inc. (Meta) in relation to the Facebook online platform (Facebook) in accordance with the annual online safety requirements under section 45K(1) of the Singapore Broadcasting Act 1994.

This report is limited to Facebook, the Social Media Service designated under section 45K(1) of the Broadcasting Act 1994, and does not cover any other services provided by Meta, that are outside the scope of that designation. The report contains information for a reporting period of April 1, 2024 to March 31, 2025. In some

¹ <https://www.imda.gov.sg/-/media/imda/files/regulations-and-licensing/regulations/codes-of-practice/codes-of-practice-media/code-of-practice-for-online-safety.pdf>

situations, due to process and/or product limitations, it was not possible to provide certain categories of data. Where applicable, this is noted in the additional comments in the relevant sections below.

Below we provide information per the respective sections and paragraphs of the COP. More information on our approach to content moderation and safety tools can be found in the [Meta Transparency Centre](#)² and [Meta Safety Centre](#)³ respectively.

Executive Summary

We recognise our responsibility at Meta to protect the safety of people who use our services. It is inherent and essential to our business: Singaporeans, other people around the world and businesses globally, will only continue to use our technologies if they have positive, meaningful and safe experiences. We have around 40,000 people working on safety and security issues globally at Meta, with over \$30 billion invested in teams and technology in this area over the last decade. As part of this effort, Meta has developed more than [50 tools and features](#)⁴ to help support teens and their parents in having a safe online experience on our apps.

We also recognise the importance of collaboration between the technology industry, governments and civil society to combat online harms. This Code is a positive step in facilitating this collaboration by providing greater transparency about the efforts technology companies have made towards online safety. We believe that transparency of such information will help inform public discourse and policy development.

This report gives an overview of the various policies, enforcement practices, tools, products, resources and partnerships we have developed to enhance the safety and security of our users in relation to our commitments under the Code. We also provide Singapore-specific metrics for a number of harm categories on content created in Singapore that we have taken action on. The metrics represented are for the period of April 1, 2024 to March 31, 2025, unless otherwise specified, and can be found in the transparency section of each harm category. Given our years of investment in online safety, this report contains information about our work that took place beyond the reporting period (i.e. prior to April 1, 2024), and which remain in effect today.

A few key highlights:

1. We announced during the reporting period that we will be **introducing [Teen Accounts on Facebook and Messenger](#)**, a feature that is already live on Instagram. Teen Accounts are accounts for teens 13-17 that are automatically set to more protective teen safety settings. These settings aim to help keep teens safe on Facebook by limiting unwanted contact, showing content that's right for their age and helping them manage their time on Facebook. We will share more in our next report.

² <https://transparency.fb.com/en-gb/>

³ <https://about.meta.com/actions/safety>

⁴ <https://about.fb.com/news/2023/06/parental-supervision-and-teen-time-management-on-metas-apps/>

2. We added a **new page in our Transparency Centre on Age Appropriate Content**⁵ to help parents and teens better understand how this approach to content works. The new page sets out the different layers of protection and provides examples of the types of content we remove, hide from teens only and avoid recommending.
3. We continue to take the lead in **fostering meaningful collaboration with industry to help make the ecosystem safer**:
 - a. In April 2024, we started to [share more sextortion-specific signals to Lantern](#)⁶, the platform we helped establish that allows participating tech companies to share signals about predatory accounts and behaviors. This industry cooperation is critical, because predators don't limit themselves to just one platform – and the same is true of sextortion scammers.
 - b. In September 2024, we worked with the Mental Health Coalition to establish [Thrive](#), the first signal sharing program for participating tech companies to share signals about violating suicide and self-harm content so that other companies can investigate and take action if the same or similar content is being shared on their platforms.
4. We continue to build on our partnership with EYEYAH! to **expand public education programs in support of youth safety and wellbeing**, including workshops for educators.
5. In response to the 2024 Online Safety Assessment Report findings, we are working through different measures related to improved handling of in-app user reports from Singapore. We look forward to continued partnership with the IMDA on assessing the performance of our in-app user reporting system.

2024 – 2025 Annual Online Safety Report to be submitted to IMDA under the Code of Practice for Online Safety

Instructions

- Template for designated Social Media Services' (dSMS) submission of annual online safety report;
- The annual online safety report will be published on IMDA's website. Please ensure that the report contains all the information requested in this template and adheres to the format of the template as closely as possible;
- Please ensure that information provided is for the reporting period that IMDA has agreed to with the dSMS;
- Please include Singapore-relevant data and information where possible;
- Please try to map the content categories to the harmful/inappropriate content categories under the Online Safety Code;
- The annual online safety report must be submitted to OCO@imda.gov.sg by no later than 30 June 2025;

⁵ <https://transparency.meta.com/en-gb/policies/age-appropriate-content>

⁶ <https://www.meta.com/help/policies/809291991003600/>

Code Obligations:	For dSMS's	Additional Comments and Explanatory Notes on the data provided
Section A: User Safety		
Paragraph 8: End-users must be able to use the Service in a safe manner. In this regard, the Service must put in place measures to minimise end-users' exposure to harmful content, empower end-users to manage their safety on the Service and mitigate the impact on end-users that may arise from the propagation of harmful content. <i>And</i> Paragraph 9: Children in particular, may lack the capacity or experience to deal with the information and content available online and will need more protection to ensure a safer online space for them. In this regard, the Service must therefore also have specific measures to	Please provide information on the measures in place for all End-users in Singapore: Paragraph 8: All end-users in Singapore The mission of Meta Platforms, Inc. (Meta) is to enable people to build community and bring the world closer together. But none of this is possible if people don't feel safe with our technologies. We have around 40,000 people working on safety and security issues globally at Meta, with over \$30 billion invested in teams and technology in this area over the last decade. Every quarter we publish a Community Standards Enforcement Report⁷, which tracks our progress and demonstrates our continued commitment to making Facebook safe and inclusive. Our approach to online safety consists of five components, which we elaborate on in the proceeding sections of this report: ● Policies that provide clear rules on what is allowed and not allowed on our applications. These are our Community Standards⁸, which cover 22 different violation areas spanning from frauds and scams to bullying and harassment. These standards are global and available in over 93 different languages. ● Enforcement processes, tools and technologies that help us scale and accelerate policy enforcement efforts.	We require that people who use our apps (i.e. Facebook for the purpose of this report) are aged 13 years and above. "Children" between the ages of 13-17 who use our applications will be referred to as "teens" throughout this report. The only time we refer to "minors" in this report is when it is in direct relation to our Community Standards as currently published on the Meta

⁷ <https://transparency.meta.com/reports/community-standards-enforcement/>

⁸ <https://transparency.meta.com/en-gb/policies/community-standards/>

protect children from harmful content.	● Tools, products and resources that raise awareness of online safety issues, provide access to accurate and credible information, give more context on content people see and provide people with more control over their online experience. Meta has developed more than 50 tools and features⁹ to help support teens and their parents. Since 2023, the following safety tools were launched on Facebook: ○ January 2023: We began giving teens more ways to manage the types of ads they see on Facebook with ad topic controls¹⁰. ○ February 2023: Meta and the National Centre for Missing and Exploited Children (NCMEC) launched Take It Down¹¹, a new tool to help prevent the spread of young people's intimate images. NCMEC is a nonprofit organisation whose mission is to help find missing children, reduce child sexual exploitation, and prevent child victimisation. ○ June 2023: Although out of scope for this report, it is worth noting that we began requiring people to send an invite to get their permission to connect on Messenger. We limit these message request invites to text only, so people can't send any photos, videos, or voice messages, or make calls, until the recipient has accepted the invite to chat. These changes mean people won't receive unwanted photos, videos, or other types of media from people they don't follow. ○ June 2023: We began showing teens a notification when they spend 20 minutes on Facebook, prompting them to take time away from the app and set daily time limits¹². ○ October 2023: We gave people the option to manually hide comments¹³, to give them even greater control over comments that they may find upsetting or unwelcome. ○ November 2023: We brought parental supervision¹⁴ tools to Facebook, allowing parents to oversee things like:	Transparency Centre.
---	--	-----------------------------

⁹ <https://www.meta.com/help/policies/safety/tools-support-teens-parents/>

¹⁰ <https://about.fb.com/news/2023/01/age-appropriate-ads-for-teens/>

¹¹ <https://about.fb.com/news/2023/02/helping-prevent-the-spread-of-young-peoples-intimate-images-online/>

¹² <https://about.fb.com/news/2023/06/parental-supervision-and-teen-time-management-on-metas-apps/amp/>

¹³ https://www.facebook.com/help/297845860255949?cms_platform=android-app&helpref=platform_switcher

¹⁴ <https://about.fb.com/news/2023/06/parental-supervision-and-teen-time-management-on-metas-apps/>

	■ The amount of time their teen spends on Facebook ■ The ability to schedule breaks for their teens and access expert resources on managing their teen’s time online ○ January 2024: We began hiding more types of age-inappropriate content¹⁵ for teens on Facebook. ○ January 2024: We announced stricter default message settings for teens¹⁶ under 16, meaning only people they follow or people they’re already connected to can message them or add them to group chat. ○ April 2024: We started to share more sextortion-specific signals to Lantern¹⁷, the platform we helped establish that allows participating tech companies to share signals about predatory accounts and behaviors. This industry cooperation is critical, because predators don’t limit themselves to just one platform – and the same is true of sextortion scammers. ○ March 2025: We have introduced a new friends tab¹⁸. The friends tab on Facebook is designed to show content only from Facebook friends, without any recommended content. This allows the user to focus on updates and posts from people they know, making it easier for the user to stay connected with friends. ● Partnerships that provide on-the-ground knowledge and expertise and enhance digital literacy. In Singapore, some of our notable partnerships include working with local community organisations like EYEYAH!, an award winning educational company that uses eye-catching artworks and illustrations to create engaging learning materials for children and teens, TOUCH Community Services, SG Her Empowerment and more. Meta also consults with experts on the Media Literacy Council (MLC) as well as government partners such as the Ministry of Digital Development & Information’s (MDDI) Digital Readiness and Learning Division, and the InfoComm Media Development Authority’s (IMDA) Digital Literacy and Wellness / DfL Learning Design department. Please see our response to paragraphs 13 and 21 below for details of programs we run with these partners.	
--	--	--

¹⁵ <https://about.fb.com/news/2024/01/teen-protections-age-appropriate-experiences-on-our-apps/>

¹⁶ <https://about.fb.com/news/2024/01/introducing-stricter-message-settings-for-teens-on-instagram-and-facebook/>

¹⁷ <https://www.meta.com/help/policies/809291991003600/>

¹⁸ <https://about.fb.com/news/2025/03/bringing-magic-of-friends-back-to-facebook/>

	● Transparency of our efforts for the public to scrutinise and hold us accountable by sharing our policies, enforcement and transparency reports published on our Transparency Centre¹⁹. Please provide information on any differentiated or additional measures in place for Children: Paragraph 9: Children We require that people who use our apps are aged 13 years and above. We want teens to have safe, age-appropriate experiences on our apps. We regularly consult with experts in adolescent development, psychology and mental health to help make our applications safe and age-appropriate for teens, including improving our understanding of which types of content may be less appropriate for teens. Meta has developed more than 50 tools and features²⁰ to help support teens and their parents, and we have spent over a decade developing policies and technology to address content and behaviour that breaks our rules. In particular, we have recently launched a package of new measures for teens that involves the following, including those mentioned in our response to paragraph 8: ● Making it more difficult for adults to find, follow and message unconnected teens ● Limiting advertisers' ability to reach teens ● Further limiting teens' ability to see sensitive content so as to ensure they have more age-appropriate content experiences	
--	---	--

¹⁹ <https://transparency.meta.com/>

²⁰ <https://about.fb.com/news/2023/06/parental-supervision-and-teen-time-management-on-metas-apps/>

- Messaging defaults for teens that prevent them from being contacted by other accounts they are not connected to
- [Parental supervisory tools](#)²¹ for teens across Facebook and Messenger, that allow parents and guardians to set daily time limits, blocking hours when teens are unable to use the apps, as well as approve or deny changes teens may make to some of their settings.
- **Teens Accounts on Facebook:** Meta rolled out Instagram Teen Accounts during the reporting period and announced that soon we'll begin making Teen Accounts available on Facebook and Messenger thereafter. Teen Accounts on Facebook and Messenger will offer similar, automatic protections to limit inappropriate content and unwanted contact, as well as ways to ensure teens' time is well spent.

We elaborate on these measures in the proceeding sections.

A note on Age Assurance

We understand that age assurance is out of scope for this report; however, as it is related to the safety of online safety and providing an age-appropriate experience, we are sharing the approach taken by Facebook:

Understanding age is an industry-wide challenge because people may misrepresent how old they are to access apps and experiences that weren't designed for them. We've invested in a combination of technologies to estimate the age of a person using our services and have taken a multilayered approach:

- We require everyone on Facebook to share their age with us in order to have an account.
- We require everyone on Facebook who attempts to change their date of birth to go from an age under 18 to 18 or over to verify their age by uploading their ID or recording a video selfie.
- We use artificial intelligence (AI) to help better understand whether or not someone is a teen or an adult on Facebook in certain use cases (i.e. U18/O18)

²¹ <https://familycenter.meta.com/>

	○ People can report underage accounts, and we train our content reviewers to flag accounts that may be used by those who are underage. As age assurance technologies evolve across industry with companies working to find more effective, user-friendly and privacy protective ways to facilitate this, Meta continues to globally advocate for age assurance at the application store level with age signals passed throughout app ecosystems. Understanding a user's real age is key to all of the efforts by policymakers and app providers to promote a more age appropriate experience online. However, an app-by-app age assurance approach has a number of issues because parents and young people will face considerable time, privacy and security impositions in order to engage with the age assurance approach adopted by each app. This is why we have been advocating globally for a 'whole-of-ecosystem' approach that requires app store and OS providers to take the lead in verifying age and share age bands APIs with app providers.	
--	---	--

Paragraph 11: End-users' exposure to harmful content must be minimised via reasonable and proportionate measures. These measures include, but are not limited to, a set of community guidelines and standards, and content moderation measures that are put in place and affected by the Service. The Service's community guidelines and standards must address the categories of harmful content in paragraph 4 and must be published.	Please provide information on the community guidelines and standards in place and how they address the categories of harmful content in paragraph 4. Meta maintains a set of globally applicable policies and standards, including our Community Standards²² and Advertising Standards²³ that defines what is and isn't allowed on Facebook. These standards apply to content worldwide and are integral to helping our users have positive experiences on Facebook. They define what we consider to be unacceptable use of our services and provide users with transparency on those rules. Our detection and enforcement mechanisms enforce these standards using comprehensive guidelines to best enable consistency across decisions, and we provide information about how this works in our Transparency Centre. The goal of our Community Standards is to create a place for expression and give people a voice whilst also protecting users from abuse. We make our policies and standards accessible by translating our Community Standards into 93 languages. To try to ensure that everyone's voice is valued, we take great care to include people from all communities, particularly marginalised communities, in the evolution of our standards. We also engage externally with safety advocates, human and civil rights experts, mental health experts, educators, think tank researchers, policy writers, users and parents to inform policy updates. Each section of our Community Standards starts with a "Policy Rationale" that sets out the aims of the policy followed by specific policy lines that outline: i) content that's not allowed; and ii) content that requires additional information or context to enforce on, content that is allowed with a warning screen or content that is allowed but can only be viewed by adults aged 18 and older. Violating our policies may result in deleted content, disabled accounts or other restrictions. The Meta policy management process requires the Community Standards to be routinely reviewed to validate they are up-to-date with evolving trends across cultures. This process has a robust feedback loop to identify and prioritise needed policy changes, update required roadmaps and retrain our models and human reviewers to implement the policy change(s). Additionally, some policy teams coordinate to enable integration across problem areas and services and to identify risk trends surfacing through process and system monitoring.	
--	---	--

Additionally, Meta has internal policies, frameworks and procedures in place to enforce the Community Standards and manage the operational activities surrounding the identification, enforcement and management of violating content on our online services. Please see below for a table setting out how the Community Standards address each of the categories of harmful content in paragraph 4 of the Code of Practice:

Code of Practice Category	Community Standards
Sexual Content	Adult Nudity and Sexual Activity ²⁴ Child Sexual Exploitation, Abuse, and Nudity ²⁵
Violent Content	Violent and Graphic Content ²⁶ Violence and Incitement ²⁷
Suicide and Self Harm content	Suicide and Self Injury ²⁸
Cyberbullying	Bullying and Harassment ²⁹ Hateful Conduct ³⁰
Content Endangering Public Health	Restricted Goods and Services ³¹
Content Facilitating Vice	Restricted Goods and Services ³² Dangerous Organisations and Individuals ³³

More information about how the Community Standards apply specifically to teens can be found in our response to paragraph 17.

Please provide information on the content moderation measures in place and how they are enforced or effected. Please include screenshots or evidence where possible.

Every day, we remove millions of violating pieces of content and accounts on Facebook. In most cases, this happens automatically, with technology to detect, restrict, and remove content and accounts that may go against our Community Standards and other policies. In other cases, our technology selects content for human review. Our review teams review a blend of user reports and

	content surfaced by our technology. Our technology also supports the review teams by prioritising the most critical content to be reviewed, based on severity, virality, and likelihood of a violation. Our review systems use technology to prioritise high-severity content with the potential for offline harm (e.g. posts related to terrorism and suicide) and viral content that is spreading quickly and has the potential to reach a large audience, in order to prevent as much harm as possible. Our content review team operates globally and reviews content 24/7, including the four official languages in Singapore. The team includes reviewers with language expertise to enforce our policies in cases where certain words or content require additional contextual understanding. We also have a global pool of language-agnostic reviewers for content that does not require language expertise, such as nudity and sexual activity. Of the violating content we take action on, our technology detects the vast majority of it before anyone reports it and we publish how effective we proactively detect and remove violating content in our quarterly Community Standards Enforcement Report³⁴ in our Transparency Centre. On January 7, 2025, we announced that we are going to change how we enforce our policies to reduce mistakes, beginning in the United States. We have been using automated systems to scan for all policy violations, and we’re going to continue to focus these systems on tackling illegal and high-severity violations, like terrorism, child sexual exploitation, drugs, fraud and scams. For less severe policy violations, we’re going to rely on someone reporting an issue before we take any action. As conveyed to the IMDA on March 21, 2025, we are still working through future changes to enforcement where for “less severe policy violations”, we will rely on someone reporting an issue before we take any action. Additional details on how our content moderation policies and processes work can be found on the enforcement section³⁵ on the Transparency Centre. Restricting Accounts We also take action against accounts that post content that violate our Community Standards, Terms of Service or our Content Distribution Guidelines on multiple, separate occasions in the recent past. Actions we take are based on a strike system and vary depending on the severity of the violation and	
--	--	--

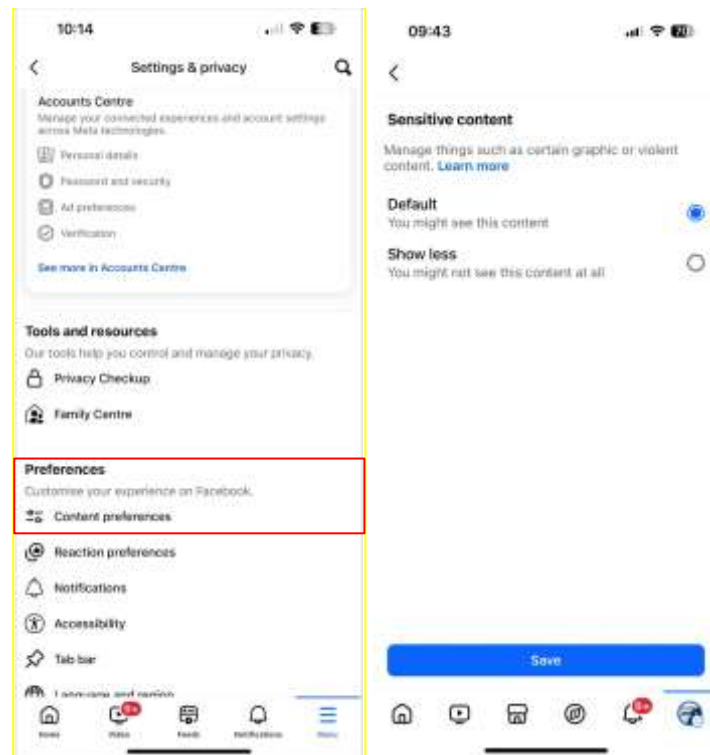
	the number of violations. These restrictions generally only apply to accounts, but they may also be extended to Pages that represent an individual, such as a celebrity or political figure. If content posted goes against our more severe policies, such as our policies on Dangerous Organisations and Individuals, the account may receive additional, longer restrictions from certain features, on top of the standard restrictions above. For example, the account may be restricted from creating ads and using Facebook Live for a set period of time, starting on the first strike. For most violations, if the user continues to post content that goes against our Community Standards after repeated warnings and restrictions, we will disable the account.	
--	--	--

Paragraph 12: End-users must have access to tools that enable them to manage their own safety and effectively minimise their exposure to, and mitigate the impact of, harmful content and unwanted interactions on the Service. Such tools may include: (a): Tools to restrict visibility of harmful content and/or unwanted comments (b): Tools to limit visibility of the end-user's account, including profile and content, as well as contact and/or interactions with other end-users (c): Tools to limit location sharing	Please provide information on the following, and include screenshots or evidence where possible. All end-users in Singapore: At Meta, we take a comprehensive approach to making our technologies a safe place for everyone. Our goals for developing safety tools are simple. We want users to: ● Stay safe on our technologies ● Keep their accounts secure ● Know how to protect their information ● Have more control over their online experience On Facebook, users are in charge and in control of what they choose to share, the content others see and experience from them, who can contact them. Users can share with us their content preferences. Taking full advantage of our safety tools enables an enriched experience for the users. The Meta Safety Centre³⁶ provides information to users about our tools and policies to stay safe, secure their accounts and protect their information. Paragraph 12(a): Tools to restrict visibility of harmful content and/or unwanted comments ● Unfollow and unfriend: When a user unfollows someone, they won't see that person's posts in their feed. On Facebook, they'll still be friends with them when they unfollow. If a user chooses to unfriend someone, Facebook won't notify the person, but the person will be removed from the user's friends list. ● Block: Users can also block someone to unfriend them. This will prevent them from starting conversations with the user on Facebook or seeing things that the user posts on their profile. In addition, people that user blocks can no longer tag them or invite them to events or groups. Blocking is reciprocal, so the user also won't be able to see things the blocked person posts. When a user blocks someone, we don't notify them. In October 2022, we updated our tools such that when a user blocks someone³⁷, they now also have the option to block other accounts that person may have or create, making it more difficult for that person to contact	
---	--	--

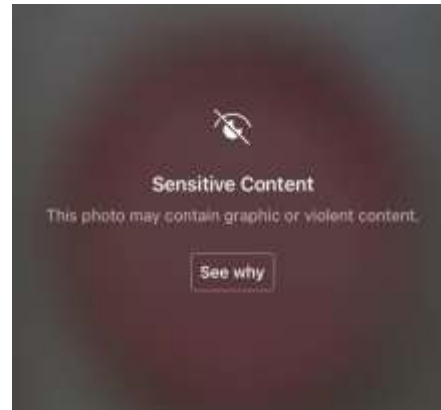
³⁶ <https://www.facebook.com/safety>

³⁷ https://www.facebook.com/help/1000976436606344?helpref=about_content

	the user. Users also have the option to “mute” and “snooze” Facebook friends as a way to limit seeing content from specific connections. Content preferences: At Facebook, we believe in giving users control over their experience. Our content preferences option allows users to customise their feed to see more or less of certain types of content, such as posts from friends, family, or specific Pages, by indicating if they are “interested” or “not interested.” There is a Favourite option which helps users manage their "Favourites" to select people and Pages that they want to prioritise, meaning those posts will be shown higher in their feed. Additionally, users can select specific friends or Pages to "Unfollow" or “Snooze” to temporarily stop seeing their posts. If users have previously unfollowed a Page or person, they can reconnect with them from the “reconnect” option to start seeing their posts in their feed. Users can also manage the level of sensitive content by adjusting their preferences to see less sensitive content which will limit their exposure to content that may be considered inappropriate for them. We provide clear instructions on how to access and use these controls, ensuring users can tailor their experience to suit their preferences.	
--	---	--

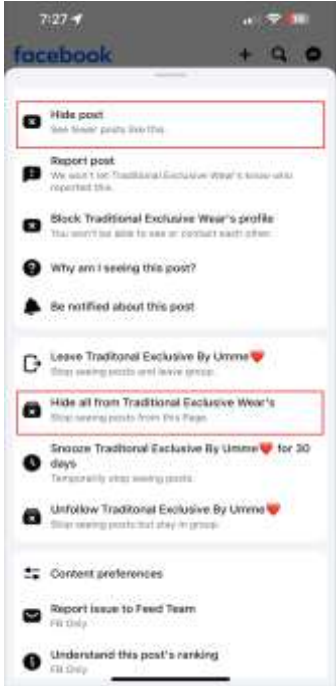


- Warning screens:** We include a warning screen over potentially sensitive content on Facebook, such as violent or graphic imagery, posts that contain descriptions of bullying or harassment, if shared to raise awareness (e.g. news reporting), some forms of nudity (e.g. in educational context) and posts related to euthanasia and Suicide & Self-Injury (SSI) recovery. To ensure the safe browsing experience for users, these labels and notices are used to restrict accidental exposure to disturbing and inappropriate material. We cover certain content in Feed and other surfaces, so people can choose whether to see it. We restrict users under 18 from seeing this content.

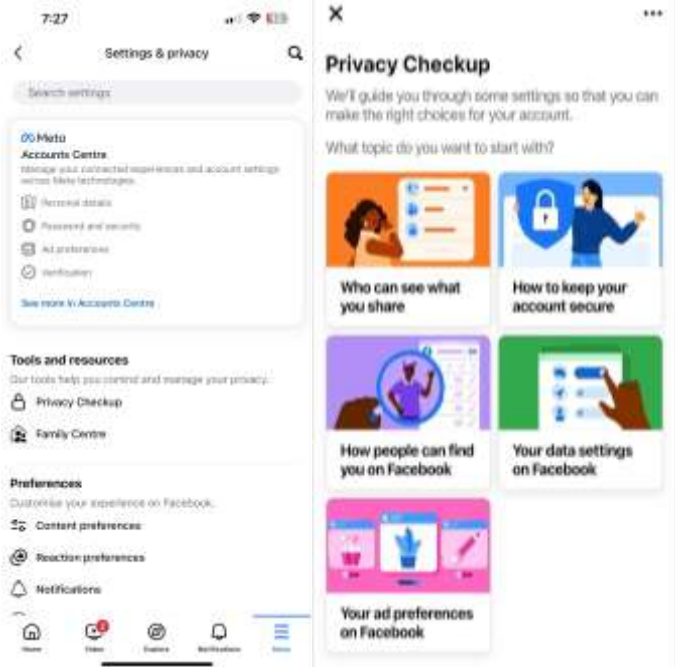


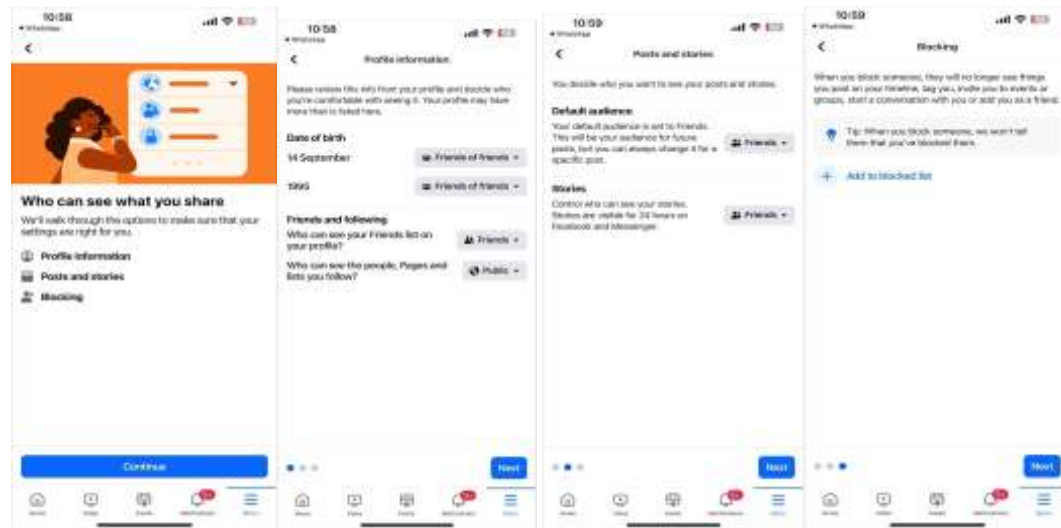
- **Managing comments:** Using privacy and safety tools users can control comments on their posts. Users can adjust who is able to comment on their posts: the public, just friends or only certain profiles and Pages. To promote safe, meaningful conversations with the community and prevent harassment, users can hide comments containing certain words. Profanity filter on the Page helps users to hide the most commonly reported words and phrases marked offensive by the community.
 - Hide comments on profile: The comment will still be visible to the person who wrote it and their friends. This can help prevent more unwanted comments. The comment will be hidden for everyone else that user chose to share it with (e.g. friends or public). Replies to the comment will also be hidden. If a user wants to hide specific replies to a comment, they can select and hide each reply separately. If they only hide replies to a comment, the original comment will still be visible to everyone else that you chose to share it with.
- **Moderation Assist:** Moderation Assist helps reduce the amount of time spent on moderation by allowing Page admins to set criteria that automatically hides incoming comments that are not wanted on Page posts. With Moderation Assist, Page admins have the ability to select from a list of criteria to automatically hide certain types of comments on posts, such as comments with links, from profiles with no profile picture, from new followers and more. The Moderation Assist activity Log lets admins review all actions taken based on selected criteria, undo individual actions, or edit or remove criteria at any time.

	The following criteria can help to proactively moderate new comments on Page posts that may be likelier to be less authentic, low quality and spammy: 1. Criteria by author: (a) No profile picture: author does not have a profile picture; (b) New account: author has had their Facebook account for less than one week; (c) Friends: Author does not have any friends or followers 2. Criteria by comment: (a) Link in comment: comment has a link; (b) Links to specific sites: comment has links to specific sites; (c) Image in comment: comment has an image; (d) Video in comment: comment has a video; (e) Keywords in comment: Comment has certain keywords. Variations of the keyword also trigger the criteria. For example if a user added “tree” to the keyword list, Moderation Assist hides keywords with TREE, t.r.e.e., tr33, treee and #tree. ● Hide Post: Hide post feature allows users to remove specific pieces of content from their news feed without unfollowing or blocking the person or Page that shared it. This option helps users to control their feed as per their preferences. Users can even hide all posts from the same person or Page in one go by clicking on Hide all posts option.	
--	--	--

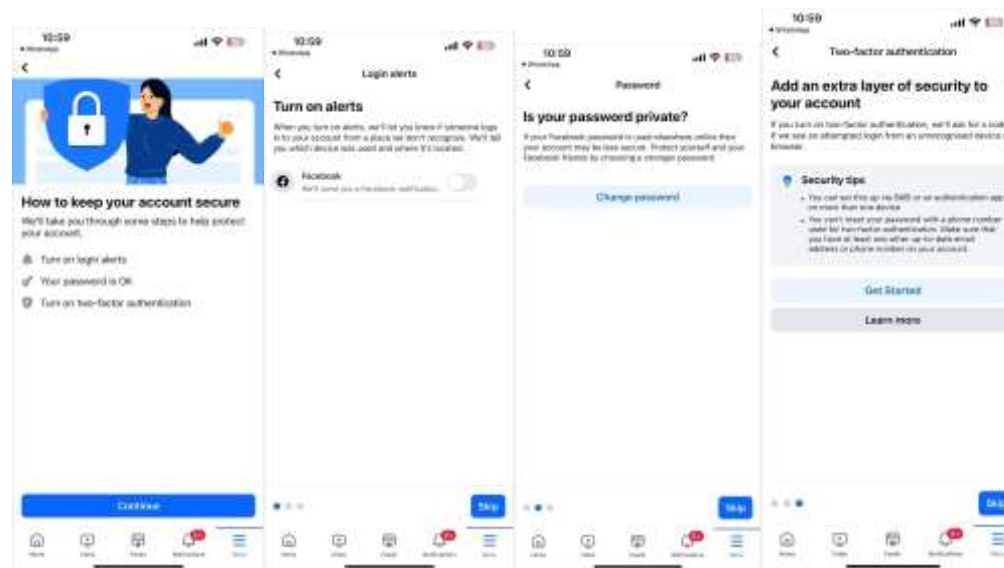
	 The screenshot shows the Facebook mobile app interface. At the top, the status bar shows 7:27. The Facebook logo is at the top left. Below the header, a menu of options for interacting with a post is displayed. The options are: 'Hide post' (highlighted with a red box), 'Report post', 'Block Traditional Exclusive Wear's profile', 'Why am I seeing this post?', 'Be notified about this post', 'Leave Traditional Exclusive By Umme' (with a heart icon), 'Hide all from Traditional Exclusive Wear's' (highlighted with a red box), 'Snooze Traditional Exclusive By Umme' (with a heart icon and 'for 30 days'), 'Unfollow Traditional Exclusive By Umme' (with a heart icon), 'Content preferences', 'Report issue to Feed Team', and 'Understand this post's ranking'. Each option has a small icon to its left.	
	All end-users in Singapore: Paragraph 12(b): Tools to limit visibility of the end-user's account, including profile and content, as well as contact and/ or interactions with other end-users We care about the safety of our global community, which is why we provide tools to help keep users' accounts secure and protect users' privacy. • Privacy Checkup: Users can use the Privacy Checkup³⁸ tool to review and adjust settings to make sure that they are sharing their profile, posts and other things with who they want. Facebook provides various options to limit visibility of users' content and protect privacy.	

³⁸ https://www.facebook.com/help/443357099140264?helpref=about_content

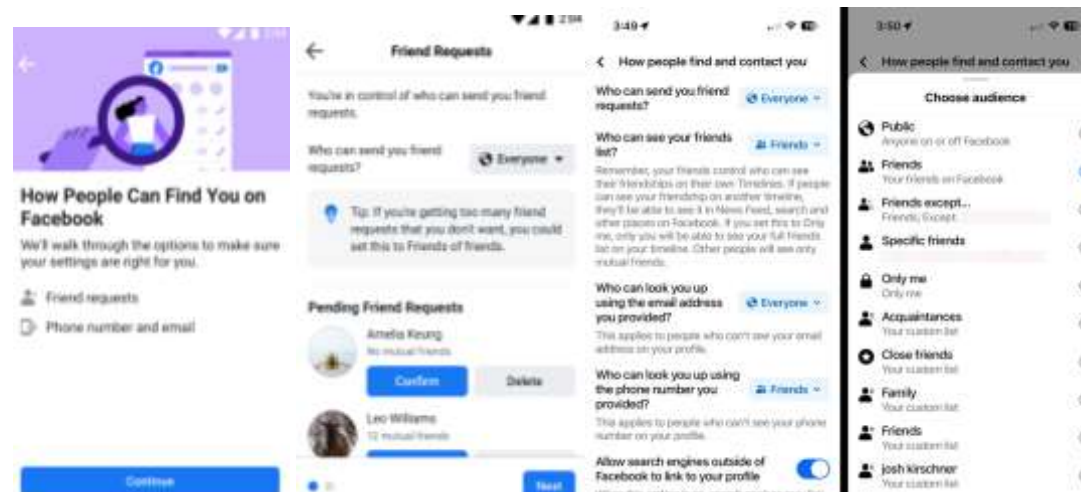
	 The image shows two screenshots from a mobile app. The left screenshot is the 'Settings & privacy' menu, which includes sections for 'Meta Accounts Centre', 'Tools and resources', and 'Preferences'. The right screenshot is the 'Privacy Checkup' screen, which guides users through various privacy settings with icons and text labels. ○ Who can see what you share: This privacy setting option gives guidance on deciding who should see a user's content, including past and future posts that the user shares. It also helps to control others' interactions with the user's profile information, such as phone number, email address, date of birth and relationship status. Users can also review their block list on Facebook.	
--	--	--



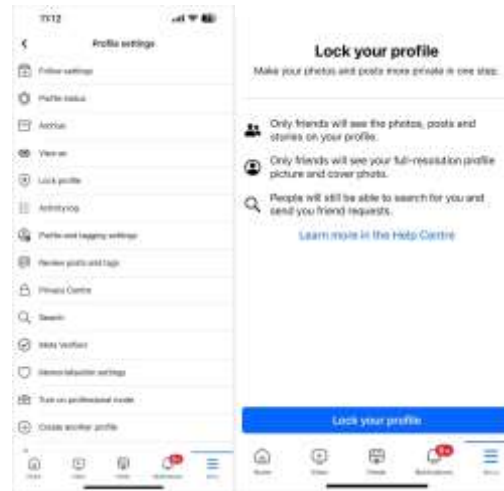
- **How to keep your account secure:** This section guides users to protect their account from unauthorised access, account misuse and account hijacking by mitigating risks. It encourages users to apply an additional layer of security by enabling login alerts, having strong passwords and by turning on two-factor authentication to maintain account integrity.



- **How people can find you on Facebook:** This section lets users choose how other end-users can connect with them. Users can decide who can send friend requests or look them up on Facebook using their phone number or email address.



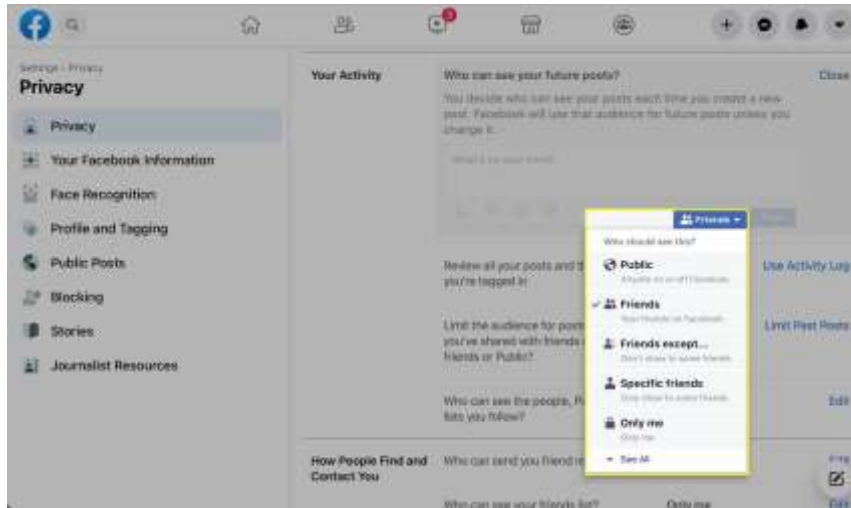
- **Lock your profile:** Lock your profile is a tool that was developed specifically with teens and women's safety in mind. When a user's profile is locked, people who aren't their friends can't download, enlarge or share the user's profile photo. They also can't see any posts or other photos on the user's timeline, regardless of when they may have posted it. Users can lock their profile to protect their posts and photos from people who aren't their friends on Facebook.



- **Limit past posts:** Users can change who can see their past posts.



- **View as:** Users can use this tool to view their profile as someone who isn't their friend, to see what they see.
- **Audience selector:** Users can choose a specific audience when posting and even set it in privacy settings as defaults. Their options may include:
 - **Public:** When users share something with Public that means anyone, including people who do not have a Facebook account, can see it.
 - **Friends:** This option lets users post stuff to their friends on Facebook.
 - **Only Me:** This option allows users to post stuff to their timeline that is visible only to them. Posts with the audience of Only Me will appear in users' feeds but not on their friends' feeds.
 - **Custom:** When users choose Custom, they can selectively share something with specific people, or hide it from specific people.



- **Tag and mention controls:** Users can choose whether they want everyone, only people they follow or no one to be able to tag or mention them in a comment, caption or story.
- **Unfollow and unfriend:** When a user unfollows someone, they won't see that person's posts in their feed. On Facebook, they'll still be friends with them when they unfollow. If a user chooses to unfriend someone, Facebook won't notify the person, but the person will be removed from the user's friends list.
- **Block:** Users can also block someone to unfriend them. This will prevent them from starting conversations with the user on Facebook or Messenger or seeing things that the user posts on their profile. In addition, people that user blocks can no longer tag them or invite them to events or groups. Blocking is reciprocal, so the user also won't be able to see things the blocked person posts or start conversations with the blocked person. When a user blocks someone, we don't notify them. In October 2022, we updated our tools such that when a user [blocks someone](https://www.facebook.com/help/1000976436606344?helpref=about_content)³⁹, they now also have the option to block other accounts that person may have or create, making it more difficult for them to contact the user.

³⁹ https://www.facebook.com/help/1000976436606344?helpref=about_content

- **Friend lists:** Users can organise their friends and limit or custom-share what certain friends see using [Friend lists](#)⁴⁰. To help users get started, we already make available the following friend lists:
 - Restricted list may be used to limit certain friends to only see public information (e.g. posts and profiles info that the user chose to make public) and posts the friend has been tagged in. This list may be used for people that have been added as a friend but the user may want to share all the posts with (e.g. the user's manager or co-workers).
 - Close friends list may be used to share certain content exclusively with close family and friends.
 - Acquaintances list allows users to share less by excluding certain friends from posts.
 Users can also create custom friend lists. When a user posts something, they can choose in the [audience selector](#)⁴¹ which list of friends they want to share with.



- **Take a Break (from profile):** This feature allows users to take a break from a profile that they're friends with on Facebook. The user who implemented take a break, they'll be able to:
 - See less of someone's profile: If they choose to see that person less, the person's posts and posts they're tagged in won't appear in their Feed. They also won't be prompted to message that person or tag them in photos. To see that person's posts again, they can follow them.

⁴⁰ https://www.facebook.com/help/200538509990389?helpref=faq_content

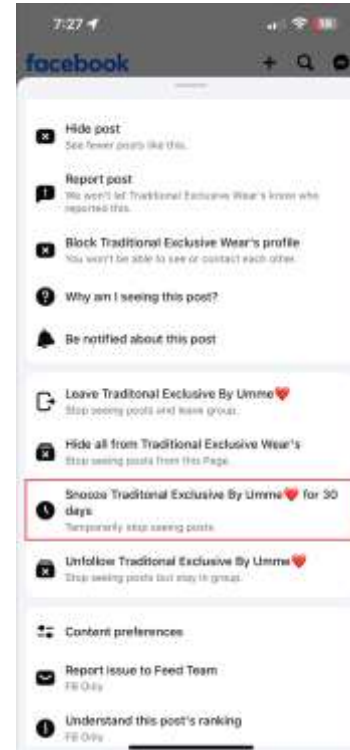
⁴¹ https://www.facebook.com/help/120939471321735?helpref=faq_content

- Limit a profile's ability to see posts and posts that user is tagged in: Limiting Someone's Ability to See Posts: Users can hide their posts from the profile they want to take a break from. That person will be added to their Restricted list and will only see posts the user tags them in or shares publicly.
- Edit Who Can See Past Posts: Users can limit who can see their past posts, including photos, status updates, videos, and more. They can limit posts they're tagged in or posts they've tagged someone else in by changing their privacy settings.

These posts can only be seen by the profiles that are tagged in them. These posts won't appear on users' timeline any more, but will still appear on other profiles' timelines and search results unless the profile who originally posted removes the content. Users will also be removed as a contributor from any shared albums with the profile that they want to take a break from.



- **Snooze:** Since 2017, users can temporarily hide posts from a person, Page, or group for 30 days by enabling the Snooze feature on Facebook. The profile, Page or group that users have snoozed won't know that they've been snoozed. Users can only snooze someone from a post in their Feed.

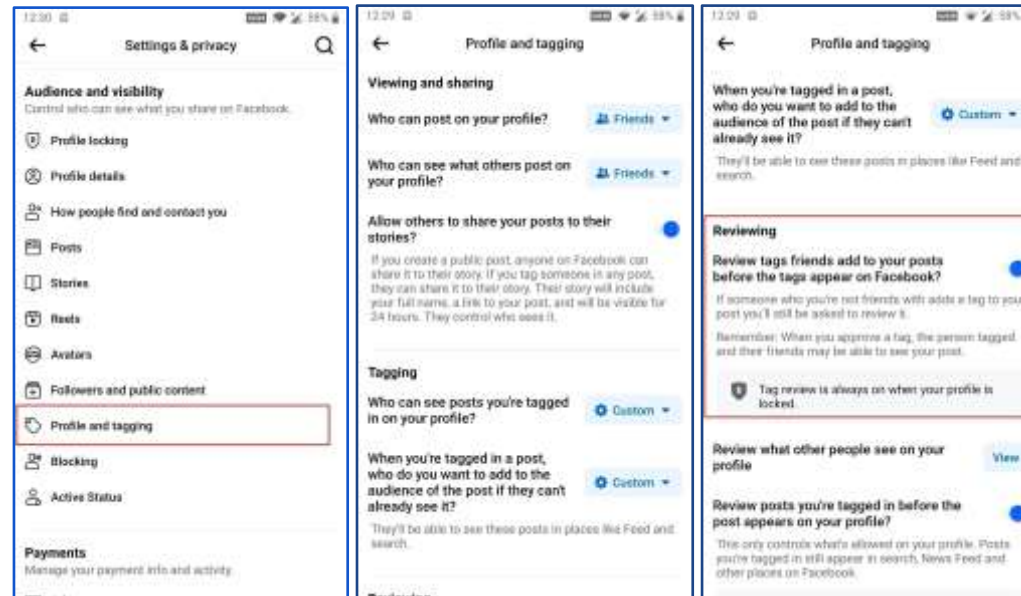


All end-users in Singapore:

Paragraph 12(c): Tools to limit location sharing

As it's important for teens in particular to think before they share their location, location sharing is off for them by default. When either an adult or teen turns on location sharing, we include a consistent indicator as a reminder that they're sharing their location.

Anyone can tag a user in their posts, including posts with a location tag. Users will see a notification if they have been tagged. Tag and timeline review features on Facebook allows users to control how they are represented on the application by reviewing and approving or rejecting tags and posts shared on their timeline. These settings can be adjusted to require approval for all tags or posts, or to automatically approve tags or posts from friends. Depending on the user's settings, if someone they aren't friends with tags them, users will receive a request to approve the tag before it appears on their timeline.



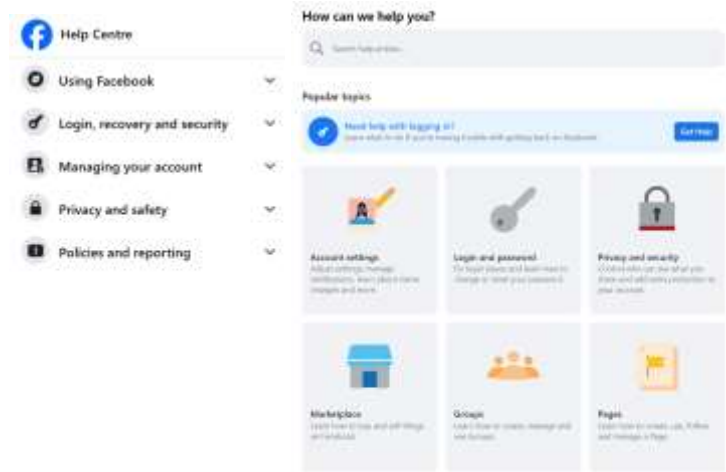
Paragraph 13: End-users must be able to easily access information related to online safety on the Service. Such information must be easy to understand and must include the availability of tools and local information, including Singapore-based safety resources or support centres, if available. The service should seek to implement, support and/ or maintain programmes and initiatives to educate and raise awareness of such information. <i>And</i> Paragraph 21: Children must be able to easily access information related to online safety on the Service. Such information must be easily understood by children and must include information on tools available to protect children harmful and/or inappropriate content and unwanted interactions, as well as local information, including Singapore-based safety resources or support centres, if available. The Service should seek to implement, support and/ or maintain	Please provide information on the measures, and include screenshots or evidence where possible. All end-users in Singapore: Paragraph 13: We offer a number of resources to educate users on how to protect themselves online, including how to find and use our safety tools. We offer many tools, products and resources to users to address different areas of safety risks and harms, including: • A dedicated Meta Safety Centre⁴² with targeted themes and audiences, including, but not limited to, youth safety, women’s safety and LGBTQ+ safety. The Safety Centre has a dedicated page for women’s safety⁴³ that offers information about tools and resources that can help women feel safe online. The Safety Centre is accessible within the Facebook app. (Please see our response to paragraph 24 below for more details.) • A dedicated Facebook Help Centre that offers a wide range of resources where people can find detailed FAQs, step-by-step guides and guidance on privacy, safety and security on	
--	--	--

⁴² <https://www.facebook.com/safety>

⁴³ <https://about.meta.com/actions/safety/audiences/women>

programmes and initiatives to educate and raise awareness of such information.

Facebook. The Help Centre provides users with resources for managing privacy, staying safe online and keeping their accounts secure from spam and scams. The resources and support information includes links to Singapore-based support institutions and helpline numbers. (Please see below for the list of local helpline numbers.) The Help Centre is accessible within the Facebook app. (Please see our response to paragraph 24 below for more details.)



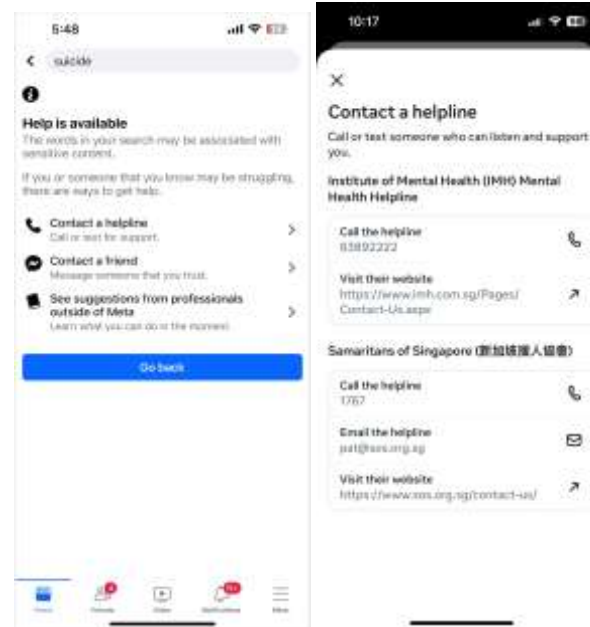
The Singapore based institutes and helplines listed below are available on the [Help Centre](#)⁴⁴ to help users in difficult times:

- [Mental Health Helpline](#)⁴⁵ from Institute of Mental Health
- [Samaritans of Singapore](#)⁴⁶

⁴⁴ https://www.facebook.com/help/1553737468262661/?helpref=hc_fnav

⁴⁵ <https://www.imh.com.sg/Pages/Contact-Us.aspx>

⁴⁶ <https://www.sos.org.sg/our-services/#tab-one>



- Transparency Centre:** We publish information on our policies and the enforcement of our policies in the Transparency Centre, which in turn helps educate users and allows the public to scrutinise our content moderation practices. For years, we have been publishing regular [transparency reports](https://transparency.meta.com/reports/)⁴⁷ to give our community visibility into how effectively we enforce our policies, respond to data requests and protect intellectual property. Please see below graph from our Community Standards Enforcement Report depicting the content actioned by us globally for violating our policy on Adult Nudity and Sexual Activity on Facebook for illustration.

⁴⁷ <https://transparency.meta.com/reports/>

We keep people safe and let people hold us accountable by sharing our policies, enforcement and transparency reports.



Our policies

Our policies are designed to keep people safe and let people hold us accountable by sharing our policies, enforcement and transparency reports.

[Policies](#)



Our enforcement

We use a combination of automated tools and human reviewers to enforce our policies and keep people safe.

[How Reports/Enforcement](#)



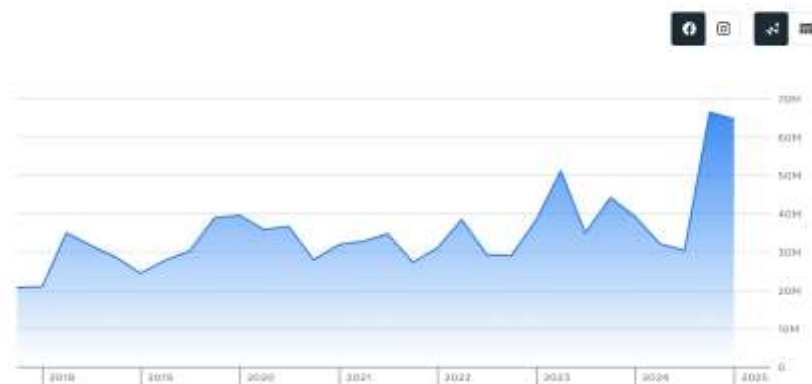
Our transparency reports

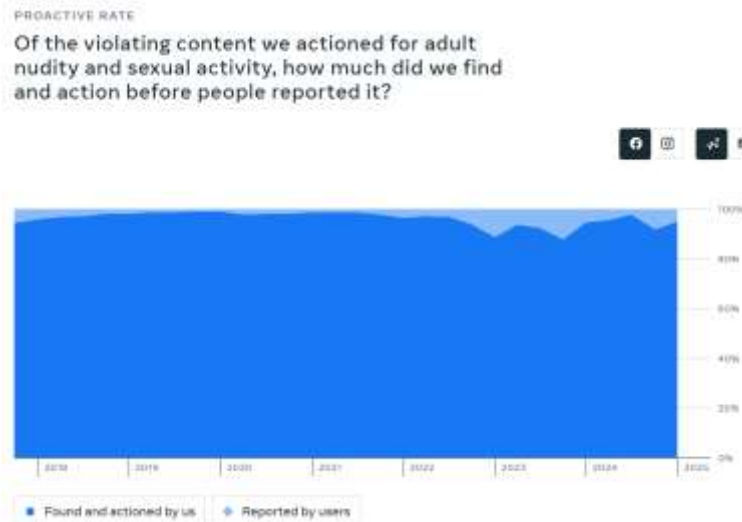
We share our transparency reports to help people understand how we enforce our policies and keep people safe.

[Transparency reports](#)

CONTENT ACTIONED

How much adult nudity and sexual activity content did we take action on?





- Last year on Transparency centre, we have published a new page named Age Appropriate Content⁴⁸ which describes the multi layered approach that we take to prevent teens from seeing content that's sensitive or mature in three main ways. We remove content completely when it breaks our rules, hide sensitive or mature content from teens, and avoid recommending an even broader set of content. Here's an example of how we apply these different layers of protection as it relates to our policies related to nudity or sexual activity.

✗ We remove images and videos containing nudity or explicit sexual activity, including when generated by AI. We make exceptions in some cases for medical, educational, artistic and public interest content.

⚠ For teens, we go further, hiding images and videos that don't contain explicit nudity or sexual activity but could be considered sexually suggestive because of a sexual pose or see-through clothing. Teens can't see this content even when posted by someone they know.

⁴⁸ <https://transparency.meta.com/en-gb/policies/age-appropriate-content>

🔵 We also avoid recommending that content to adults, and there are certain types of content – such as some implicitly sexual content – that we also don't recommend to teens.

These layers combine to help make sure that teens are seeing content that's appropriate for them.

Meta has implemented and supported various programmes and initiatives to educate and raise awareness of online safety in the community in Singapore. Following are some key initiatives we implemented during the reporting period:

- **Workshops for educators in Singapore:** On 20 and 22 May 2025, our partner, EYEAH!, with support from Meta, conducted two workshops for educators in Singapore. The workshops focused on using visual literacy to promote critical thinking, creativity, and cyber wellness through the "CyberGrime" module, a curated collection of artworks that explores cyber wellness for youths. Participants learned how to use the illustrations to explore themes of digital well-being, such as anxiety, addiction and cyber bullying, equipping them with new tools and strategies to enrich student learning on these topics.



- **SG Her Empowerment (SHE):** On February 24, 2025, Meta presented at SHE's annual symposium on AI, Misogyny and Online Harms on our approach to online safety. SHE has been our Trusted Partner since 2022. As a Trusted Partner, SHE reports platform violations to us and emerging trends in online harms for the Singapore context.
- **NCPC Crime Prevention Roadshow 2025:** At the NCPC Roadshow 2025, held from 22–23 February, EYEHYAH! and Meta presented an interactive scams game aimed at educating the public about online safety. The game recorded over 1,000 plays over the two days. EYEHYAH!'s illustrations were also used across various booths giving the festival a unique eye-catching look. The event included a visit from then-Minister of State Sun Xueling, who presented EYEHYAH! and Meta with a token of appreciation for their efforts in public education.



- **Digital for Life Festival 2024 – EYEHYAH! and Meta Booth Showcase:** At the Digital for Life Festival 2024 Showcase, held on 2–3 November at Heartbeat @ Bedok and 9–10 November at VivoCity, the EYEHYAH! booth in partnership with Meta featured a gamified scams kiosk, offering interactive games, educator conversation cards, and vibrant animations to raise awareness about online scams. Visitors explored observational games and received an EYEHYAH! SCAMS issue after participating in booth activities. We also gave out one-page tips on safety features and tools on Facebook and Instagram, including a preview of Teen Accounts. With footfall of 66,000 people, the game recorded approximately 2,100 plays over the two days, and over 1,000 magazines were given away. The pop-up was honoured by visits from Mr. Tan Kiat How, Senior Minister of State, and Dr. Wan Rizal, Member of Parliament for Jalan Besar GRC.





Please provide information on the measures, and include screenshots or evidence where possible.

Paragraph 21: Children

At Meta, child protection is always a top priority. We want to support parents and caregivers as they guide their teens' online experiences. We've worked with experts to build a collection of resources that foster conversations between caregivers and teens to help teens navigate and develop healthy online safety habits.

We offer a number of informative resources to parents, teens and educators to help increase awareness and online safety education among teens. Specifically:

- The [Family Centre](https://familycenter.meta.com/in/)⁴⁹ provides parents with tools and resources to help support their teens' online experience. Parents may oversee their teens' accounts within Meta technologies, set up and use supervision tools and access resources from experts and review helpful articles, videos and tips on topics like how to talk to teens about social media. These are accessible from within the Facebook app.



- The [Meta Safety Centre](https://about.meta.com/actions/safety/)⁵⁰ has dedicated sections for youth safety, encompassing multiple areas related to online safety concerns. Here, teens can find advice and tools that can help them balance their time and ensure they are prioritising connections that make them feel good, seeking support when needed and helping others.
 - The [Youth Safety](https://about.meta.com/actions/safety/audiences/youth/)⁵¹ section guides youth to safely navigate digital spaces, as it is critical for protecting their emotional health and well-being. This section contains

⁴⁹ <https://familycenter.meta.com/in/>

⁵⁰ <https://about.meta.com/actions/safety/>

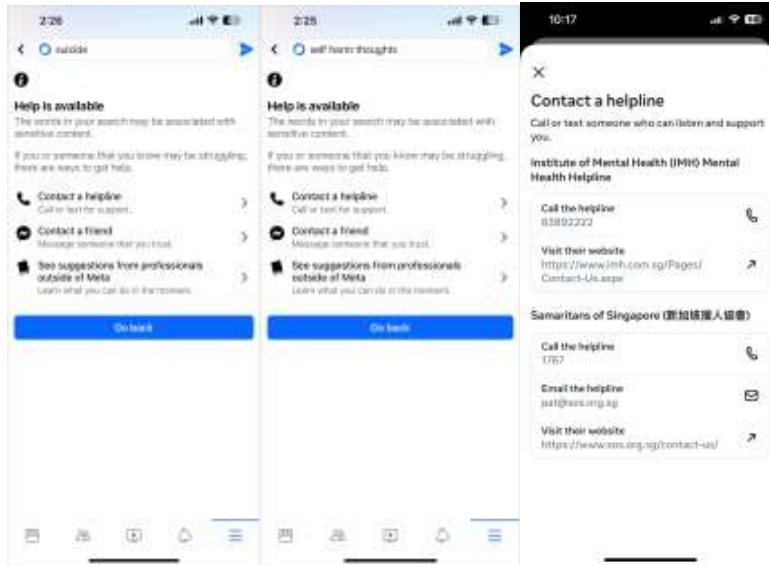
⁵¹ <https://about.meta.com/actions/safety/audiences/youth/>

	various safety tips, tools and resources to educate and enable youth to stay safe, prioritise their well-being, create positive change and prevent or respond to those struggling with bullying or suicidal ideation. ○ The Bullying and Harassment⁵² hub on the Meta Safety Centre is a resource for teenagers, parents and educators seeking support for issues related to bullying and other conflicts. At Meta, we take bullying and harassment situations seriously. Bullying and harassment is a unique challenge and one of the most complex issues to address because context is critical. We work hard to enforce against this content while also equipping our community with tools to protect themselves in ways that work best for them. We're always creating new resources and tools, and revising policies with expert help to continually tackle the issue. This section offers step-by-step guidance, including information on how to start important conversations about bullying, intimate image abuse and sextortion. It also suggests steps on how to report such situations and provides references of partners in action to prevent bullying. ● We Think Digital: Singapore was the first country in which we launched our digital literacy program We Think Digital⁵³ in 2019. Working with experts across the Asia Pacific, this program aims to equip young people to develop the knowledge and skills needed to thrive as digital citizens. These resources, developed in partnership with the Berkman Klein Centre for Internet and Society at Harvard University, are designed to be used by educators and families both in the classroom and at home with lesson plans and activities to help build core competencies skills teens need to navigate the digital world in safe ways. ● Get Digital Hub: We have collaborated with experts to develop Get Digital⁵⁴, a digital citizenship and wellbeing program that provides schools and families with lesson plans and activities to help build the core competencies and skills young people need to navigate the digital world in safe ways and become wise, competent and responsible digital citizens who	
--	--	--

⁵² <https://about.meta.com/actions/safety/topics/bullying-harassment>

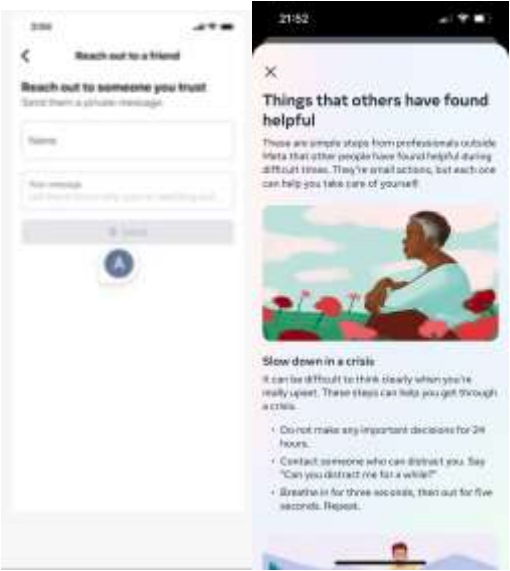
⁵³ <https://wethinkdigital.fb.com/sg/en-sg/>

⁵⁴ Meta, Get Digital Hub, <https://www.facebook.com/fbgetdigital>

	creatively, compassionately and safely use, control and build technology to improve individual and community prosperity and well-being.	
Paragraph 14: End-users who use high-risk search terms such as, but not limited to, terms relating to self-harm and suicide on the Service must be actively offered relevant safety information (stated in paragraph 13) such as, but not limited to, local suicide prevention hotlines, if available. <i>And</i> Paragraph 22: Children who use high-risk search terms, such as, but not limited to, terms relating to self-harm and suicide, on the Service must be actively offered relevant safety information (stated in paragraph 21) such as, but not limited to, local suicide prevention hotlines, if available.	Please provide information on the measures, and include screenshots or evidence where possible. Paragraph 14: All end-users in Singapore When end-users try to search terms or content related to suicide or self-harm that violate our Suicide and Self-Injury⁵⁵ policy, we will show a prominent full-screen message which directs them to local organisations that can provide support, such as Samaritans of Singapore (SOS) and the Mental Health Helpline from the Institute of Mental Health (IMH), and curated help resources from professionals outside of Meta that we collaborate with 	

⁵⁵ <https://transparency.meta.com/policies/community-standards/suicide-self-injury/>

In addition to directly contacting a helpline, end-users are able to contact a friend or see suggestions from professionals outside of Meta. Such suggestions include how to “slow down in a crisis”, “change your surroundings”, or “take care of yourself”. If our Community Operations team becomes aware of people at risk of immediate harm, either through a report from our community or our proactive detection technology, we will contact local emergency services to get them help.



Meta does allow people to share content discussing their own struggles with suicide, self-harm or eating disorders; however, the policy is not to recommend this content. We regularly consult with experts in suicide and self-injury, as well as body image and eating disorder experts, to help inform our policies and enforcement, and we work with organisations around the world to provide assistance to people in distress. When users search for terms related to suicide or self-harm, these related search results will be hidden, and the end-user will be directed to the expert resources mentioned above.

	Our approach, guided by these experts and organisations, aims to strike a delicate balance between preventing people seeing sensitive or upsetting content, while giving people space to talk about their own experiences, which experts say is important. In September 2024, Meta worked with the Mental Health Coalition to establish Thrive, the first signal sharing program to share signals about violating suicide and self-harm content. Through Thrive, participating tech companies will be able to share signals about violating suicide or self-harm content so that other companies can investigate and take action if the same or similar content is being shared on their platforms. Thrive builds on the work we already do at Meta to remove harmful content that shows graphic imagery or encourages suicide or self-harm, while still giving space for users to talk about their own experiences.	
	Please provide information on the measures, and include screenshots or evidence where possible. Paragraph 22: Children Please refer to our response to paragraph 17(c), to understand how teens who use high-risk search terms, such as, but not limited to, terms relating to self-harm and suicide, on Facebook are actively offered relevant safety information.	
Paragraph 15: End-users' exposure to child sexual exploitation and abuse material and terrorism content on the Service must be minimised through the use of technologies and processes. These technologies and processes must proactively detect and swiftly remove child sexual exploitation and abuse material and terrorism content as technically feasible, such that the extent and length of time to	Please provide information on the measures, and include screenshots or evidence where possible. All end-users in Singapore: <u>Child Sexual Abuse Material (CSAM):</u> Our response here on CSAM addresses paragraph 15 and 16. We do not tolerate any behaviour or content that exploits children online, and to avoid any potential for abuse, we take action on nonsexual content as well, such seemingly benign photos of children in the bath. We have a detailed policy in place to address Child Sexual Exploitation, Abuse and Nudity	

which such content is available on the Service is minimised. <i>And</i> Paragraph 16: End-users must be protected from preparatory child sexual exploitation and abuse activity and terrorism activity on the Service through reasonable and proportionate steps taken by the Service to proactively detect and swiftly remove preparatory child sexual exploitation and abuse activity (such as online grooming for child sexual abuse) and terrorism activity (such as glorifying or endorsing terrorist activities and recruitment).	(CSEAN)⁵⁶ activity or content on our applications. Below is additional information on how we minimise any exposure to child sexual exploitation and abuse material, including preparatory activity, through proactive detection and removal of such content. Proactive Detection of CSAM Meta uses industry-leading technology to find and remove violating CSAM content so it doesn't spread on our applications. While messaging is out of scope for this report, we would add that we use a combination of technology and behavioural signals to detect and prevent grooming, or potentially inappropriate interactions between a teen and an adult. Specifically, Meta uses: • Photo- and video-matching technologies and PhotoDNA to detect identical or near-identical photos and videos of known content (something that has already been reviewed and determined to be CSAM), so we can quickly remove it. • Artificial intelligence and machine learning to proactively detect child nudity and previously unknown child exploitative content. We use this to find accounts that engage in potentially inappropriate interactions with minors across Meta technologies so that we can remove them and prevent additional harm. These automatic detection tools may help us remove this content even before anyone reports it. We also remove accounts that promote this type of content and any CSAM content reported to us by members of our community. • Meta runs PhotoDNA on links from other internet sites and their associated content to detect known child exploitation housed elsewhere on the internet. Not only does this help keep our technologies safer but it also helps keep the broader internet safer as all violating content is shared with the NCMEC⁵⁷ which maintains a database of known CSAM hashes that is then shared with participating industry players. • Meta expanded its work to detect and remove networks that violate our child exploitation policies and updated our child safety policies to clarify that we will remove Facebook profiles, Pages and groups that are dedicated to sharing otherwise innocent images of children with captions, hashtags or comments containing inappropriate signs of affection or commentary about the children depicted in the image.	
--	--	--

⁵⁶ <https://transparency.meta.com/en-gb/policies/community-standards/child-sexual-exploitation-abuse-nudity/>

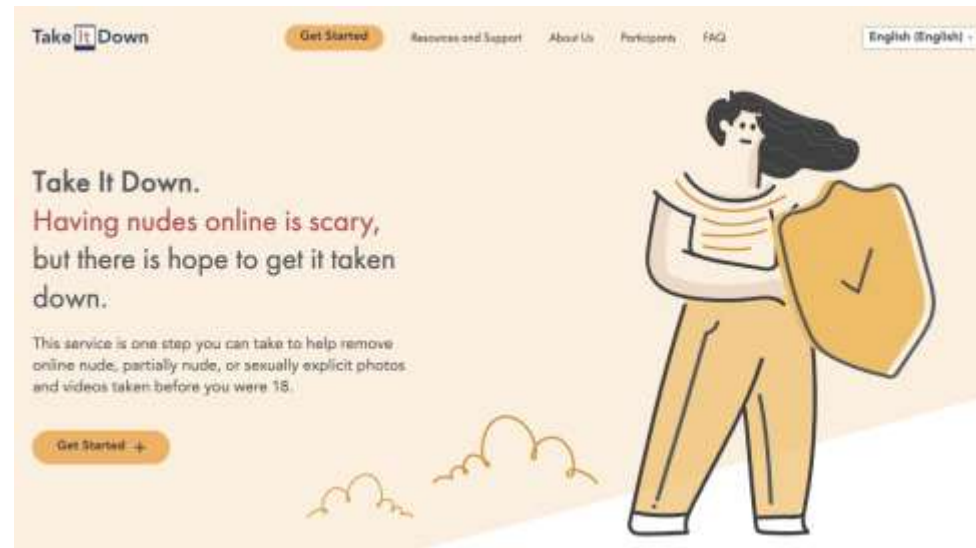
⁵⁷ <http://www.missingkids.com/home>

	● Restricting Potential Predators and Removing Their Networks: Since 2023, we've developed technology that identifies potentially suspicious adults, and we review more than 60 different signals to find these adults, such as if a teen blocks or reports an adult, or if someone repeatedly searches for terms that may suggest suspicious behavior. On Facebook, we're using this technology to better find and address certain Groups, Pages and Profiles. For example, Facebook Groups with a certain percentage of members that exhibit potentially suspicious behavior will not be suggested to others in places like Groups You Should Join. Additionally, Groups whose membership overlaps with other Groups that were removed for violating our child safety policies will not be shown in Search. We also hire specialists with backgrounds in law enforcement and online child safety to find predatory networks and remove them. These specialists monitor evolving behaviors exhibited by these networks – such as new coded language – to not only remove them, but to inform the technology we use to proactively find them. All apparent instances of child exploitation identified on our site globally are reported to NCMEC, which coordinates with law enforcement authorities from around the world. To strengthen our work on cross-industry prevention of any exposure to child sexual exploitation on our applications, Meta has also invested in the following: ● Technology Coalition's Project Protect: In 2020, Meta (then Facebook) joined Google, Microsoft and 15 other tech companies to announce the formation of Project Protect⁵⁸, a plan to combat online child sexual abuse, a renewed commitment and investment from the Technology Coalition expanding its scope and impact to protect children online and guide its work for the next decade. ● In February 2023, Facebook was announced as one of the founding members of Take It Down⁵⁹, a new platform by NCMEC to find child exploitative content and help prevent young people's intimate images from being posted online in the future. Take It Down assigns a unique hash value, a numerical code, to their image or video privately and directly from their own device. Once they submit the hash to NCMEC, companies like ours can use those hashes	
--	---	--

⁵⁸ <https://about.fb.com/news/2020/06/fighting-child-exploitation-online/>

⁵⁹ <https://takeitdown.ncmec.org/>

to find any copies of the image, take them down and prevent the content from being posted on our apps in the future. Other companies can join to share hashes.

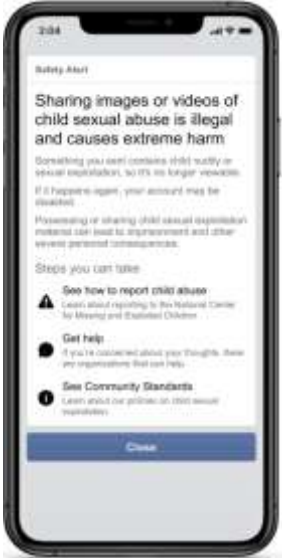


- In November, 2023, we [announced](#)⁶⁰ we were founding members of [Lantern](#)⁶¹, a program run by the Tech Coalition that enables technology companies to share signals about accounts and behaviors that violate their child safety policies. This industry cooperation is critical, because predators don't limit themselves to just one platform.
- We are also a part of the WePROTECT Global Alliance industry committee. WeProtect brings together experts from government, the private sector and civil society to protect children from sexual exploitation and abuse online.
- Meta collaborated with Thorn, a nonprofit organisation that builds technology to defend children from sexual abuse, to develop [stopsextortion.com](#)⁶², which we translated onto our Safety Centre to expand its reach and provide language coverage all over the globe. In addition, we worked with Thorn and its youth program, NoFiltr, to create [educational](#)

⁶⁰ <https://about.fb.com/news/2023/11/lantern-program-protecting-children-online/>

⁶¹ <https://www.technologycoalition.org/newsroom/announcing-lantern>

⁶² <http://stopsextortion.com>

	materials⁶³ that reduces the shame and stigma surrounding the sharing of intimate images. Our work together is designed to empower teens to seek help and take back control when they've experienced harmful interactions online. • Safety alerts: We've implemented a safety tool since 2021, which we're including in this report to provide transparency on our ongoing efforts to protect users and prevent the spread of exploitative content. This tool informs users who have shared viral or meme content that exploits children about the harm it can cause, warning them that it's against our policies and there are legal consequences for sharing such material. We share this safety alert in addition to removing the content, banking it, and reporting it to NCMEC. Accounts promoting this content will be removed. We're also using insights from this safety alert to identify behavioral signals of those at risk of sharing this material, so we can educate them on its harm and encourage them not to share it on any surface, public or private.  Removal of CSAM	
--	--	--

⁶³ <https://nofiltr.org/quizzes/https://nofiltr.org/quizzes/>

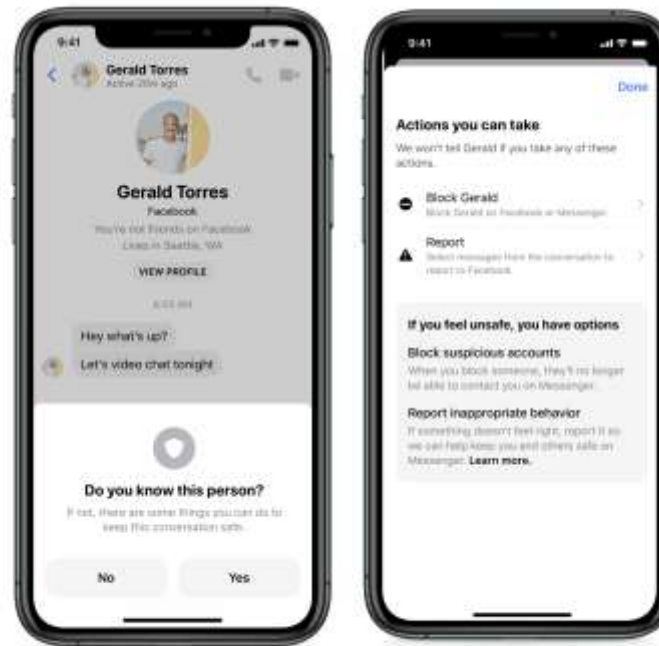
	Meta removes CSAM content under its Child Sexual Exploitation, Abuse, and Nudity⁶⁴ policy in the Meta Transparency Centre. We remove content in the following categories: • Child sexual exploitation: Content, activity or interactions that threaten, depict, praise, support, provide instructions for, make statements of intent, admit participation in or share links of the sexual exploitation of children (including real minors, toddlers or babies, or non-real depictions with a human likeness, such as in art, AI-generated content, fictional characters, dolls etc.). • Solicitation: Content that solicits sexual content or activity depicting or involving children • With respect to Paragraph 16 specifically, inappropriate interactions with children: Content that constitutes or facilitates inappropriate interactions with children. • With respect to Paragraph 16 specifically, exploitative intimate imagery and sextortion: Content that attempts to exploit real children by coercing money, favours or intimate imagery with threats to expose intimate imagery or information or sharing, threatening or stating an intent to share private sexual conversations or intimate imagery. • Sexualisation of children: Content (including photos, videos, real-world art, digital content and verbal depictions) that sexualises real or non-real children or groups, Pages, and profiles dedicated to sexualising real or non-real children. • Child nudity: Content that depicts real or non-real child nudity. • Non-sexual child abuse: Videos or photos that depict real or non-real non-sexual child abuse regardless of sharing intent or content that praises, supports, promotes, advocates for, provides instructions for or encourages participation in non-sexual child abuse. <u>Additionally, with respect to Paragraph 16:</u> On Preventing Unwanted Interactions Between Adults and Teens: We know there are bad actors on the internet trying to engage with teens. That’s why a significant focus of our work is to help keep teens safe by stopping unwanted contact between teens and adults they don’t know or don’t want to hear from.	
--	---	--

⁶⁴ <https://transparency.meta.com/en-gb/policies/community-standards/child-sexual-exploitation-abuse-nudity/>

	To protect teens from unwanted contact from adults, we've introduced several features that prevent adults from interacting with people under 18 who don't follow them. We also have robust mechanisms to identify potentially suspicious adult accounts. A "suspicious" account is one that belongs to an adult that may have recently been blocked or reported by a young person, for example. We monitor for dozens of signals of potentially suspicious behaviors, and when we identify accounts exhibiting a number of these behaviors, we disable them. • Our proactive technology uses dozens of signals to send Facebook accounts exhibiting potentially suspicious behavior to our content reviewers who perform a holistic review of the account and its content and can disable these accounts if appropriate. • We also implemented a new system to find coded terms used by potential predators in Facebook profile names, Pages and more. These accounts may be automatically removed or sent for further analysis. • We have a new mechanism to automatically disable Facebook accounts if they exhibit a number of signals we monitor for potentially suspicious behavior. We also take additional measures to protect teens from interacting with potentially suspicious adults. For example: • We restrict adults from starting private chats with teens they're not connected to on Facebook. • For example, when an adult tries to message a teen who isn't connected to them and doesn't follow them, they receive a notification that DM'ing them isn't an option. • We also show Safety Notices if a potentially suspicious adult who a teen is connected to tries to DM them. • We also restrict teens from messaging unconnected teens, by default. Users under 16 need parental approval to adjust the setting to start receiving messaging requests from other teens they are not connected to. • We also limit the message requests people can send to someone who doesn't follow or isn't already connected to them, to text-only messages. People can't send images, videos, or voice notes in message requests until the recipient accepts the request to chat.	
--	---	--

	● Messenger also gives you the option to ignore a conversation and automatically move it out of your inbox, without having to block the sender. Messenger's new default settings for teens under 16 (or under 18 in certain countries, including UK and EU) restrict teens to only receiving messages from Facebook friends, or people they're connected to through phone contacts, for example. On restricting adults from discovering teens: ● We don't show young people's accounts in Explore, Reels, 'People You May Know' or 'Accounts Suggested For You' to these adults. ● If they find young people's accounts by searching for their usernames, they can't follow them. ● They also can't see comments from young people on other people's posts, nor can they leave comments on young people's posts. ● We restrict teens from starting 1:1 messaging with potentially suspicious adults they aren't connected to. ● We don't show potentially suspicious adults in teens' 'People You May Know' recommendations and vice versa. ● As an extra layer of protection, we remove the message button on teens' Facebook accounts when they're viewed by potentially suspicious adults altogether. ● We're prompting teens to report accounts to us after they block someone. On Educating Teens: We aim to prevent harm by educating young people through education and safety notices. We notify teens when we have signals of a potentially suspicious adult account attempting to follow or message them, and prompt young users to be more cautious about interactions in their private messages. ● We remind teens to only accept friend requests from people they know: If a minor receives a friend request from an adult who is not already a connected "friend" (a possible stranger), we will provide a warning message to them that says, "Only accept friend requests from people you really know."	
--	---	--

- If someone blocks someone or deletes a friend request we now check to see if they want to report them too. Safety notices notify young people when an adult who has been exhibiting potentially suspicious behavior is interacting with them in DMs. For example, if an adult is sending a large amount of friend requests to people under 18, we'll alert the recipients within their messages and give them an option to end the conversation, or block, report, or restrict the adult.



Safety notice

Options to block or report conversation

To help make sure teens are regularly checking and updating their safety and privacy settings on Facebook, we're also sending teens new notifications encouraging them to update their settings with a single tap. With these new settings, teens under 16 will receive messages from Facebook friends, or people they're connected to through phone contacts, for example.

	● We've also introduced new privacy, safety and control features like delivery controls that let people choose who can message them, alongside existing features like report, block, and message requests. ● We use a variety of signals to detect unwanted message requests and route them to a separate, less visible folder. ● On Messenger, people who users don't follow can only send one text-only message until the user has accepted their chat request. This means teens can't receive multiple messages, videos, or images from other teens they don't follow. <u>Terrorism content:</u> There is no place for terrorism across Meta technologies. We remove terrorists and posts that support terrorism whenever we become aware of them. Meta has one of the most comprehensive policies in the industry targeting terrorist organisations, as well as hate groups and large-scale criminal organisations, like cartels. We define terrorist organisations and individuals as: "any non-state actor that: ● Engages in, advocates or lends substantial support to purposive and planned acts of violence, ● Which causes or attempts to cause death, injury or serious harm to civilians, or any other person not taking direct part in the hostilities in a situation of armed conflict, and/or significant damage to property linked to death, serious injury or serious harm to civilians ● With the intent to coerce, intimidate and/or influence a civilian population, government or international organisation ● In order to achieve a political, religious or ideological aim." Our definition is agnostic to the ideology or political goals of a group, which means it includes everything from religious extremists and violent separatists to white supremacists and militant environmental groups. It's about whether they use violence to pursue those goals. This definition and policy are applied globally. Under the Community Standards, we have a set of policies that prohibit terrorist content on Facebook, including our policies on Dangerous Organisations and Individuals⁶⁵, Violence and	
--	---	--

⁶⁵ <https://transparency.fb.com/policies/community-standards/dangerous-individuals-organizations/>

	Incitement⁶⁶, and Coordinating Harm and Promoting Crime⁶⁷. Taken together, these policies are congruent with the regulation’s definition of “terrorist content”. Detection of Terrorism content Measures to identify and remove or disable access to terrorist content, individuals, or organisations include: • We continue to designate dangerous organisations and individuals based on their behaviour both online and offline – most significantly, their ties to violence. As part of the designation process, we identify signals that indicate a banned organisation has a presence, and then we use technology to “fan out” and proactively investigate associated accounts, Pages, and groups, before removing this “cluster” all at once. • We use media-matching technology to detect when someone is trying to upload content that is identical or near-identical to photos, videos, text, and even audio that we have already removed for violating our policies around terrorism. This means that if we previously removed a propaganda video from ISIS, we can work to prevent other accounts from uploading the same video to our site. In many cases, this means that terrorist content intended for upload to Facebook simply never reaches the app. • We use machine learning classifiers to help us understand whether content uploaded to our technologies is advocating for terrorism, based on previous content we have removed for praising or supporting terrorist organisations. • We’ve increased our capability to rapidly respond to live streams that may breach our Community Standards, including by reviewing all live streams in an area that may involve footage of an attack and increasing our 24/7 capacity to respond to live stream reports. As well as our efforts around rapid response, we’ve also introduced additional restrictions on who can use Facebook Live. We apply a ‘one strike’ policy to Live in connection with a broader range of offences. Anyone who violates our most strict policies are restricted from using Live for set periods of time – for example 30 days – starting on their first offence. For instance, someone who shares a link to a statement from a terrorist group with no context is immediately blocked from using Live for a set period of time.	
--	---	--

⁶⁶ <https://transparency.meta.com/en-gb/policies/community-standards/violence-incitement/>

⁶⁷ <https://transparency.meta.com/en-gb/policies/community-standards/coordinating-harm-publicizing-crime/>

	• We have a team of specialists who exclusively or primarily focus on countering dangerous organisations as their core responsibility. This group includes former academics who are experts on counterterrorism, former prosecutors and law enforcement agents, investigators and analysts, and engineers. This team continues to study trends in terrorism, organised hate, and other dangerous organisations and works with partners to better understand these organisations as they evolve. • We also use the hash (unique digital fingerprints) sharing database of known violent terrorism imagery or propaganda, provided by the Global Internet Forum to Counter Terrorism (GIFCT), to help quickly detect and remove content involving a hash. The database houses hundreds of thousands of unique hashes. Again, in many cases, this means that terrorist content intended for upload never reaches our technologies and applications. • To help other platforms that may not have the resources and the technology, we have developed and made available a free open source software tool called Hasher-Matcher-Actioner (HMA) that identifies copies of images or videos and takes action against them en masse. HMA builds on previous open source image and video matching software from Meta, and it can be used for any type of violating content. We hope the tool will be adopted by a range of companies to help them stop the spread of terrorist content on their platforms, and that it will be especially useful for smaller companies who don't have the same resources as bigger companies.	
--	--	--



Meta collaborates with governments, intergovernmental agencies, and expert organizations to enhance cross-industry efforts against terrorism and violent extremism. To strengthen our work on cross-industry prevention of any DOI on our applications, Meta has also invested in the following:

- GIFT: We joined forces with YouTube, Microsoft and Twitter in 2017 to create the Global Internet Forum to Counter Terrorism (GIFT), an organisation built with the goal to prevent terrorists and violent extremists from exploiting digital platforms. GIFT became an independent NGO in 2019 and coordinates crisis response across the industry in response to attacks.
- Law enforcement: When we become aware of a specific, imminent and credible threat to human life, we do not hesitate to notify law enforcement. Meta also works with law enforcement around violent attacks. More details about our work with law enforcement can be found [here](https://transparency.meta.com/reports/government-data-requests/)⁶⁸.
- Independent research: We commission independent research from think-tanks, academics and NGOs on various topics of violent extremist and terrorist use of the Internet in order to help our industry understand and make progress on these important issues. In 2022, Meta announced a research partnership with the Middlebury Institute of International Studies'

⁶⁸ <https://transparency.meta.com/reports/government-data-requests/>

	Centre on Terrorism, Extremism and Counterterrorism to analyse evolving trends in violent extremism and effective tools that help communities combat it. We are dedicated to maintaining the safety and integrity of our platforms. One of the ways we achieve this is through upstream prevention, which involves taking proactive measures to prevent the spread of DOIs before they can take root in individuals or communities. This includes using targeted interventions and promoting positive speech. We believe that constructive dialogue is essential, and we strive to create an environment that encourages such exchanges. Our approach to countering extremism is multi-faceted and proactive, involving a combination of prevention, collaboration and promotion of positive speech. • Safer searches: We also, when appropriate, point people to resources when they search for terms associated with dangerous organisations and individuals. When users search for DOI terms on Facebook they are instead provided with resources, helping to deter searches for DOI content and provide educational resources to users. • Providing resources to community organisations: We also run a safety ads programme, which works to support CSO partners to proactively prevent harm by providing them with advertising credits and strategic support when they are running counter-speech programmes. By working together with these organisations, we aim to grow our partnerships to combat DOIs. Removal of Terrorism content Meta removes terrorism content on Facebook under its Dangerous Organisations and Individuals⁶⁹ policy, Violence and Incitement⁷⁰ policy, and Coordinating Harm and Promoting Crime⁷¹ policy in the Meta Transparency Centre. We remove the following type of content: • Under our Dangerous Organisations and Individuals policy, we designate and ban individuals and organisations tied to terrorism, organised hate and large-scale criminal activity, attempted multiple-victim violence, multiple victim violence, serial murders, and violent	
--	---	--

⁶⁹ <https://transparency.fb.com/policies/community-standards/dangerous-individuals-organizations/>

⁷⁰ <https://transparency.meta.com/en-gb/policies/community-standards/violence-incitement/>

⁷¹ <https://transparency.meta.com/en-gb/policies/community-standards/coordinating-harm-publicizing-crime/>

	events. We also remove all glorification, substantive support, and representation of these individuals and organisations when we become aware of it. We also designate and ban militarised social movements, violent non-state actors, and violence-inducing conspiracy networks like QAnon, and we do not allow them to have a presence on our technologies. Accounts, Pages, or groups representing these organisations or individuals are removed when we're made aware of them. Our Dangerous Organisations and Individuals policy prohibits any organisation or individual that proclaims a violent mission or is engaged in violence from having a presence on Facebook. Specifically, we do not allow on our applications: ○ Terrorist organisations and individuals ○ Hate entities ○ Criminal organisations ○ Multiple-victim violence and serial murderers (including attempted multiple victim violence) ○ Hateful ideologies ○ Violent non-state actors and Violence-Inducing Entities ○ Militarised social movements ○ Violence-inducing conspiracy networks ○ Hate banned entities. In an effort to prevent and disrupt real-world harm, we do not allow organisations or individuals that proclaim a violent mission or are engaged in violence to have a presence on our technologies. We assess these entities based on their behaviour both online and offline, including, most significantly, their ties to violence. Under the Dangerous Organisations and Individuals policy, we designate individuals, organisations, and networks of people. These designations are divided into two tiers that indicate the level of content enforcement, with Tier 1 resulting in the most extensive enforcement because we believe these entities have the most direct ties to offline harm. ○ Tier 1 focuses on entities that engage in serious offline harms, including organising or advocating for violence against civilians, repeatedly dehumanising or advocating for harm against people based on protected characteristics, or engaging in systematic criminal operations. We remove glorification, support, and representation of Tier 1	
--	--	--

	entities, their leaders, founders or prominent members, as well as unclear references to them. ○ Tier 2 includes violent non-state actors that engage in violence against state or military actors in an armed conflict but do not intentionally target civilians. We remove glorification, material support, and representation of these entities, their leaders, founders or prominent members. ● Our Violence and Incitement policy prohibits content that incites or facilitates violence and constitutes a credible threat to public or personal safety. We remove content, disable accounts, and work with law enforcement when we believe there is a genuine risk of physical harm or direct threats to public safety. ● Our Coordinating Harm and Promoting Crime policy prohibits users from facilitating, organising, promoting, or admitting to certain criminal or harmful activities targeted at people, businesses, property, or animals. We allow people to debate and advocate for the legality of criminal and harmful activities, as well as draw attention to harmful or criminal activity that they may witness or experience as long as they do not advocate for or coordinate harm. We detect dangerous organisations and terrorist content via a playbook and a series of automated techniques, which were initially developed to detect content related to widely known terrorist organisations like ISIS, al Qaeda, and their affiliates. We have since expanded our playbook and techniques.	
Paragraph 16: End-users must be protected from preparatory child sexual exploitation and abuse activity and terrorism activity on the Service through reasonable and proportionate steps taken by the Service to proactively detect and swiftly remove preparatory child sexual exploitation and abuse activity (such as online grooming for child sexual abuse) and terrorism activity (such as glorifying or	Please provide information on the measures, and include screenshots or evidence where possible. All end-users in Singapore: <u>CSAM</u>: Please see our response to paragraph 15. <u>Terrorism content</u>: We do not allow content that glorifies, supports, or represents events that Meta designates as violating violent events, including terrorist attacks, hate events, multiple-victim violence or attempted multiple-victim violence, serial murders, or hate crimes. Nor do we allow (1) glorification, support, or representation of the perpetrator(s) of such attacks; (2) perpetrator-generated content	

endorsing terrorist activities and recruitment).	relating to such attacks; or (3) third-party imagery depicting the moment of such attacks on visible victims. We also remove content that glorifies, supports or represents ideologies that promote hate, such as nazism and white supremacy. We remove unclear references to these designated events or ideologies. We also detect and identify the violence inducing entities that are engaged in preparing or advocating for future violence but have not necessarily engaged in violence to date. These are also entities that may repeatedly engage in violations of our Hateful Conduct or Dangerous Organisations and Individuals policies on or off the application. We remove glorification, material support, and representation of these entities, their leaders, founders or prominent members. These concepts apply to the organisations themselves, their activities and their members. Proactive Detection of preparatory Terrorism content Measures to proactively detect and remove preparatory terrorism activity (such as glorifying or endorsing terrorist activities and recruitment) include: • The DOI (Dangerous Organisations and Individuals) search intercept is an intervention deployed when users search for DOI terms on Facebook. These interventions serve three main purposes: deterring users from accessing DOI content, referring users at risk of radicalization to support services, and providing educational resources to users. The interventions supplement discoverability reduction initiatives by deterring users from accessing content that may violate DOI policies. • Strategic network disruptions: We use Strategic Network Disruption (SND) to combat terrorism and organized hate on our platform. SND removes entire networks of banned entities at once, making it harder for them to reconstitute and sending a clear message that they're not welcome. This approach also helps us study their tactics and improve detection. By proactively addressing these threats, we aim to create a safer environment for our users. Ongoing SND efforts are crucial to staying ahead of evolving threats. In June of 2023, the independent academic journal Proceedings of the National Academy of Sciences (PNAS) published a paper written by Meta researchers Daniel Robert Thomas and Laila A. Wahedi studying the efficacy of our SND strategy on disrupting hate groups.	
---	---	--

	Expanding our Dangerous Organisations and Individuals⁷² policy to address violence inducing social movements: • We have expanded our Dangerous Organisations and Individuals policy to address organisations and movements that have demonstrated significant risks to public safety but do not meet the rigorous criteria to be designated as a dangerous organisation and banned from having any presence on Facebook. While we will allow people to post content that supports these movements and groups, so long as they do not otherwise violate our content policies, we will restrict their ability to organise on our application. • Under this policy expansion, we have imposed restrictions to limit the spread of content from Facebook Pages, groups and accounts. We also remove Facebook Pages and accounts where we identify discussions of potential violence, including when they use veiled language and symbols particular to the movement to do so. We take the following actions on Facebook accounts, Pages and groups associated with these movements, including: ○ Removal when they discuss potential violence. ○ Limiting recommendations making them ineligible to be recommended to people. ○ Reduce ranking in Feed. ○ Reduce in Search meaning they will not be suggested through our Search Typeahead function and will be ranked lower in search typeahead results. ○ Reviewing related hashtags on Facebook: We have temporarily removed the Related Hashtags feature on Facebook, which allows people to find hashtags similar to those they are interacting with. ○ Prohibit use of ads, commerce surfaces and monetization tools. ○ Prohibit fundraising. • The proactive detection measures described in our response to paragraph 15 above applies to this section. Removal of Terrorism content:	
--	--	--

⁷² <https://transparency.meta.com/en-gb/policies/community-standards/dangerous-individuals-organizations/>

	Meta removes content that glorifies, supports or represents ideologies that promote terrorism or hate on Facebook under its Dangerous Organisations and Individuals⁷³ policy, Violence and Incitement⁷⁴ policy, and Coordinating Harm and Promoting Crime⁷⁵ policy in the Meta Transparency Centre. • We remove glorification, defined as any of the below, of Tier 1 and Tier 2 entities (mentioned in paragraph 15) as well as designated events like: ○ Legitimising or defending the violent or hateful acts of a designated entity by claiming that those acts have a moral, political, logical or other justification that makes them acceptable or reasonable; e.g. "Hitler did nothing wrong." ○ Characterising or celebrating the violence or hate of a designated entity as an achievement or accomplishment; e.g. "Hizbul Mujahideen is winning the war for a free and independent Kashmir." ○ An aspirational statement of membership or statement that you would like to be a designated entity or the perpetrator of a violating violent event; e.g. "I wish I could join ISIS and be part of the Khilafah." • We remove support, defined as any of the below, of Tier 1 entities and material support of Tier 2 entities. Material support involves: ○ Any act which improves the financial status of a designated entity – including funnelling money towards or away from a designated entity; e.g. "Donate to the KKK!" ○ Any act which provides material aid to a designated entity or event; e.g. "If you want to send care packages to the Sinaloa Cartel, use this address:" ○ Recruiting on behalf of a designated entity or event; e.g. "If you want to fight for the Caliphate, DM me." Other support involves:	
--	--	--

⁷³ <https://transparency.fb.com/policies/community-standards/dangerous-individuals-organizations/>

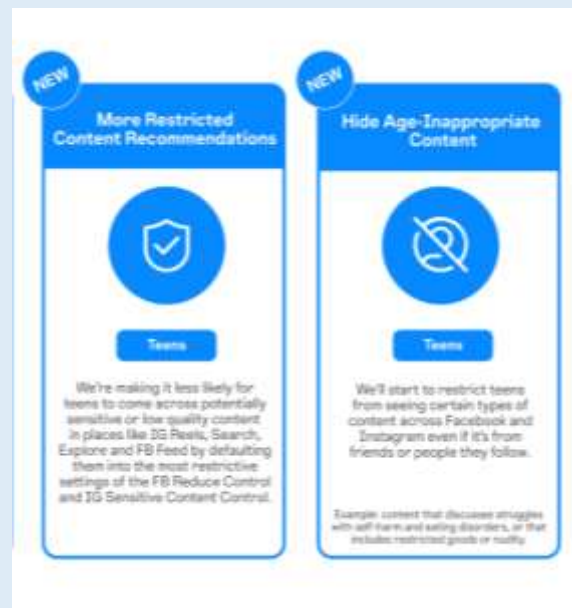
⁷⁴ <https://transparency.meta.com/en-gb/policies/community-standards/violence-incitement/>

⁷⁵ <https://transparency.meta.com/en-gb/policies/community-standards/coordinating-harm-publicizing-crime/>

	○ Channelling information or resources, including official communications, on behalf of a designated entity or event, e.g. Directly quoting a designated entity without caption that condemns, neutrally discusses or is a part of news reporting. ○ Putting out a call to action on behalf of a designated entity or event; e.g. "Contact the Atomwaffen Division – (XXX) XXX-XXXX." ● We remove representation, defined as any of the below, of Tier 1 and 2 designated organisations and designated events. ○ Stating that you are a member of a designated entity, or are a designated entity; e.g. "I am a grand dragon of the KKK." ○ Creating a Page, profile, event, group or other Facebook entity that is or purports to be owned by a designated entity or run on their behalf, or is or purports to be a designated event; e.g. a Page named "American Nazi Party."	
Paragraph 17: Besides harmful content, children's exposure to inappropriate content must also be minimised through reasonable and proportionate measures. These measures include, but are not limited to, a set of community guidelines and standards and content moderation measures put in place and effected by the Service that are appropriate for children. These community guidelines and standards must minimally address the following categories of content, and must be published: (a): Sexual content (b): Violent content	Please provide information on the measures, and include screenshots or evidence where possible. Children: We want teens to have safe, age-appropriate experiences on Facebook. We regularly consult with experts in adolescent development, psychology and mental health to help make our technologies safe and age-appropriate for teens, including improving our understanding of which types of content may be less appropriate for teens. Our approach to ensuring that the content teens see are appropriate is comprised of: ● Removing Violating Content: We work to remove content that goes against our policies, such as content that exploits children or content that incites violence (see paragraph 11 on how we remove violating content and paragraphs 17(a)(b)(c)(d) on related policies). ● Hiding Age-Inappropriate Content: We restrict teens from seeing certain types of content even if it's from friends or people they follow. For example, mature content is restricted to adults; anyone below the age threshold who is logged in to Facebook won't be able to view or play a restricted video. Or, for content that is identified as 'disturbing' or 'sensitive' but does not violate our policies, we apply a warning screen to alert people, which only users 18 and older may view. More examples of this can be found in paragraphs 17(a)(b)(c)(d) below.	

(c): Suicide and self-harm content
(d): Cyberbullying content

- **Limiting Content Recommendations:** We refrain from recommending content that goes against our [Recommendation Guidelines](https://about.fb.com/news/2020/08/recommendation-guidelines/)⁷⁶ or suggested posts in people's Feeds. We automatically place teens into the most restrictive content control setting to make it more difficult for teens to come across potentially sensitive content or accounts.
- **More Restricted Content Recommendations:** We've made it less likely for teens to come across potentially sensitive or low quality content by defaulting them to the most restrictive settings. For example, we hide results for certain topics such as suicide and self-harm.
- **Prohibiting ads about restricted topics:** Our [Advertising Standards](https://transparency.meta.com/policies/ad-standards)⁷⁷ prohibit ads about restricted topics, like alcohol, financial products and weight loss products and services, to be shown to teens. We default ad preference settings for teens to "See Less" for restricted topics to prevent teens from opting into ads that may not be age-appropriate. We have also removed the ability for advertisers to target teens based on interest, activities and gender.



⁷⁶ <https://about.fb.com/news/2020/08/recommendation-guidelines/>

⁷⁷ <https://transparency.meta.com/policies/ad-standards>

- Updating the Transparency Centre with our approach to age-appropriate experiences: Last year, we published a new page in our [Transparency Centre](https://transparency.meta.com/en-gb/policies/age-appropriate-content/)⁷⁸ to help parents and teens better understand how this approach to content works. The new page sets out the different layers of protection and provides examples of the types of content we remove, hide from teens only and avoid recommending. We recognize parents are concerned that their teens might see mature or inappropriate content online, which is why we have stricter rules around the kinds of content teens can see on our apps. In addition to removing content completely when it breaks our rules, we take steps to help prevent teens from seeing content that may be sensitive or mature. We avoid recommending it, and in some cases we hide it from teens altogether, even if it's shared by someone they follow.

Children:

Paragraph 17(a): Sexual content

Aside from our Child Sexual Exploitation, Abuse and Nudity (CSEAN) policy, we have two other policies that address sexual content on Facebook.

Adult Nudity and Sexual Activity

We prohibit depictions of adult nudity, adult sexual activity and implicit sexual activity, as well as restrict certain categories of permitted adult content (such as real-world art) from being viewed by people under 18, under our [Adult Nudity and Sexual Activity](https://transparency.meta.com/en-gb/policies/community-standards/adult-nudity-sexual-activity/)⁷⁹ policy. We default to removing sexual imagery to prevent the sharing of non-consensual or underage content. The policy focuses on photorealistic imagery, which we define as an image or video that looks like a photograph or video of a real person (or event). Where it is unclear if an image or video is photorealistic, we presume that it is.

⁷⁸ <https://transparency.meta.com/en-gb/policies/age-appropriate-content/>

⁷⁹ <https://transparency.meta.com/en-gb/policies/community-standards/adult-nudity-sexual-activity/>

	In addition to prohibiting depictions of adult nudity, adult sexual activity and implicit sexual activity for all users, we restrict people under 18 from viewing the categories of content listed below: • Real-world art, where: ○ Imagery depicts implicit, explicit, or other sexual activity or stimulation ○ Imagery depicts bestiality, provided it is shared neutrally or in condemnation and the people or animals depicted are not real • Implicit adult sexual activity in advertisements, recognized fictional images or with indicators of fiction. • Digital imagery and real world art of adult sexual activity, where: ○ The sexual activity or stimulation isn't explicit and is not a fetish as specified above ○ The content was posted in a satirical or humorous context ○ Only body shapes or contours are visible Adult Sexual Exploitation We prohibit content attempting to exploit people and further restrict visibility for people under 18 for some permitted sexual content that is shared to raise awareness. We recognise that people use Facebook to discuss and draw attention to sexual violence and exploitation. We recognize the importance of and allow for this discussion. We also allow for the discussion of sex worker rights advocacy and sex work regulation. We draw the line, however, when content facilitates sexual encounters or commercial sexual services between adults. We do this to avoid facilitating transactions that may involve trafficking, coercion and non-consensual sexual acts. We also restrict sexually-explicit language that may lead to sexual solicitation. Our policy on Adult Sexual Exploitation⁸⁰ prohibits content depicting, advocating for, or mocking non-consensual sexual touching. We also prohibit content that attempts to exploit people, including sextortion, sharing, threatening, stating an intent to share, offering or asking for non-consensual intimate imagery (NCII), sharing, threatening to share or stating an intent to share private sexual conversations and secretly taking non-commercial imagery of a person's commonly sexualized body parts. These restrictions apply to all users.	
--	---	--

⁸⁰ <https://transparency.meta.com/en-gb/policies/community-standards/sexual-exploitation-adults/>

	In addition to prohibiting content attempting to exploit people, we may provide warning labels and restrict people under 18 from viewing certain content including: • Content depicting non-consensual sexual touching when shared to raise awareness (without entertainment or sensational context), the survivor is not identifiable, and the content does not involve nudity • Content depicting fictional non-consensual sexual touching (movie trailers, etc.) when shared by trusted partners to raise awareness and without sensational context In addition to our at-scale policy of removing content that threatens or advocates rape or other non-consensual sexual touching, we may also disable the posting account and work with law enforcement, in addition to removing the content. We may also enforce on content shared by a third party that identifies survivors of sexual assault when reported by an authorized representative or trusted partner.	
	Children: Paragraph 17(b): Violent content We have two policies that address violent content. Under our Violent and Graphic Content policy, there are different actions that we take based on the sensitivity of the content. Our Violence and Incitement policy focuses on content that potentially incites offline harm. Violent and Graphic Content There are different levels of actions that we take under our Violent and Graphic Content⁸¹ policy. We either remove content or apply warning screens to alert people that the content may be sensitive or disturbing. For users under 18, we restrict their ability to view such labelled content. We prohibit and remove content that is particularly violent or graphic, such as videos depicting dismemberment, visible innards or charred bodies. We also remove content that contains sadistic remarks towards imagery depicting the suffering of humans and animals.	

⁸¹ <https://transparency.meta.com/en-gb/policies/community-standards/violent-graphic-content/>

	We allow some graphic content (with limitations) that may have a public interest value such as to help people to condemn and raise awareness of important issues such as human rights abuses, armed conflicts or acts of terrorism. We also allow some graphic content that may have a social, medical or other value. We display warning screens and labels for such content and restrict users under 18 from seeing this content. Violence and Incitement Under our Violence and Incitement⁸² policy, we aim to prevent potential offline harm that may be related to content on Facebook. While we understand that people commonly express disdain or disagreement by threatening or calling for violence in non-serious ways, we remove language that incites or facilitates serious violence. We remove content, disable accounts and work with law enforcement when we believe there is a genuine risk of physical harm or direct threats to public safety. We also try to consider the language and context in order to distinguish casual statements from content that constitutes a credible threat to public or personal safety. In determining whether a threat is credible, we may also consider additional information such as a person's public visibility and the risks to their physical safety. We remove threats of violence against various targets, such as threats against public figures that could lead to serious injury or threats targeted at children, private adults, and groups of people on the basis of their protected characteristic(s). We also remove content that may help facilitate violence, such as content that asks for, offers or admits to offering certain services of violence (for example, hitmen, mercenaries, assassins, female genital mutilation) or advocates for the use of these services or glorification of gender-based violence. Children: Paragraph 17(c): Suicide and self-harm content Suicide, Self-Injury and Eating disorders	
--	--	--

⁸² <https://transparency.meta.com/en-gb/policies/community-standards/violence-incitement/>

	In addition to suicidal and self-harm content, our Suicide and Self-Injury⁸³ policy covers content related to eating disorders. We regularly consult with experts in suicide and self-injury or eating disorders to help inform our policies and enforcement, and we work with organisations around the world to provide assistance to users in distress. Our approach, guided by these experts and organisations, aims to strike a delicate balance between preventing users seeing sensitive or upsetting content, while giving users a space to talk about their own experiences, which experts say is important. We do not allow users to intentionally or unintentionally celebrate or promote suicide, self-injury or eating disorders, but we do allow users to discuss these topics because we want Facebook to be a space where users can share their experiences, raise awareness about these issues and seek support from one another. We remove any content that encourages suicide, self-injury or eating disorders, including fictional content such as memes or illustrations and any self-injury content which is graphic, regardless of context. We also remove content that mocks victims or survivors of suicide, self-injury, or eating disorders, as well as real time depictions. Content about recovery from suicide, self-injury or eating disorders that is allowed, but may contain imagery that could be upsetting, such as a healed scars, is placed behind a sensitivity screen. When users post or search for suicide or self-injury related content, we will direct them to local organisations that can provide support, and if our Community Operations team detects immediate harm, we will engage local emergency services to get them help. As we regularly consult experts in adolescent development, psychology and mental health to help make our technologies safe and age-appropriate for teens, in January 2024, we started to remove borderline suicide, self-harm or eating disorders content. For example, if a user is posting about their ongoing struggle with thoughts on self-harm, we understand that it could destigmatize the issue; however, as it is a complex topic, it may not be suitable for all teens.	
--	--	--

⁸³ <https://transparency.meta.com/en-gb/policies/community-standards/suicide-self-injury/>

	In addition to prohibiting content that promotes, encourages, coordinates, or provides instructions for suicide, self-injury, or eating disorders, we also remove the below content for users under 18, which is otherwise displayed with a sensitive content warning screen for adults: • Photos or videos depicting a person who engaged in euthanasia/assisted suicide in a medical setting. We remove this type of content, as well as other types of age-inappropriate content, from teens' experiences on Facebook. We already aim not to recommend this type of content to teens in places like Search and Explore, but since January 2024, such content is no longer shown to teens in Feed, even if it's shared by someone they follow. Children: Paragraph 17(d): Cyberbullying content Bullying and harassment At Meta, we take bullying and harassment situations seriously. Bullying and harassment is a unique challenge and one of the most complex issues to address, because context is critical. We work hard to enforce against this content and behaviour while also equipping our community with tools to protect themselves in ways that work best for them. We continuously consult with experts to help update our policy and work with global organisations to maintain our bullying prevention resources and provide guidance to parents, caregivers and educators who have a teen involved in bullying. The organisations include International Bullying Prevention Organization, The Diana Award Anti-Bullying Ambassador Programme and the Yale Centre for Emotional Intelligence. Bullying and harassment happen in many places and come in many different forms, from making threats and releasing personally identifiable information to sending threatening messages and making unwanted potentially violating contact. We do not tolerate this kind of behaviour because it prevents people from feeling safe and respected on Facebook.	
--	--	--

	Our Bullying and Harassment⁸⁴ policy is tiered. In addition to content restricted under universal protection for everyone, we have additional protections for teens to protect them against specific types of content. For private individuals, we remove content that's meant to degrade or shame, including, for example, claims about someone's sexual activity. For users under 18, we provide heightened protections because we recognise that bullying and harassment can have more of an emotional impact on teens. Additional protection for teens: Content that we take action on for all users under 18 include: ● Claims about sexual activity, except in the context of criminal allegations against adults (non-consensual sexual touching). ● Content sexualising another adult (sexualisation of minors is covered in the Child Sexual Exploitation, Abuse and Nudity policy). ● Dehumanising comparisons (in written or visual form) to or about animals and insects, including subhuman creatures, that are culturally perceived as inferior; bacteria, viruses, microbes and diseases; inanimate objects, including rubbish, filth, faeces. ● Content manipulated to highlight, circle or otherwise negatively draw attention to specific physical characteristics (nose, ear and so on). ● Content that ranks them based on physical appearance or character traits. ● Content that degrades individuals who are depicted being physically bullied (except in self-defence and fight-sport contexts). Content that we take action on for private minors, private adults and minor involuntary public figures (such as children of public figures): ● Targeted cursing. ● Claims about romantic involvement, sexual orientation or gender identity. ● Calls for action, statements of intent, aspirational or conditional statements, or statements advocating or supporting exclusion. ● Negative character or ability claims, except in the context of criminal allegations and business reviews against adults.	
--	--	--

⁸⁴ <https://transparency.meta.com/en-gb/policies/community-standards/bullying-harassment/>

	• Expressions of contempt or disgust, or content rejecting the existence of an individual, except in the context of criminal allegations against adults. Self-reported content from private minors, private adults and minor involuntary public figures that we take action on: • First-person voice bullying. • Unwanted manipulated imagery. • Comparison to other public, fictional or private individuals on the basis of physical appearance. • Claims about religious identity or blasphemy • Comparisons to animals or insects that are not culturally perceived as intellectually or physically inferior ("tiger", "lion"). • Neutral or positive physical descriptions. • Non-negative character or ability claims. • Attacks through derogatory terms related to a lack of sexual activity. Content specific to private minors that we take action on: • Allegations about criminal or illegal behaviour. • Videos of physical bullying against minors, shared in any context. • Derogatory terms related to female gendered cursing. Hateful Conduct At Meta, we define hateful conduct⁸⁵ as direct attacks against people – rather than concepts or institutions – on the basis of what we call protected characteristics (PCs): race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease. Additionally, we consider age a protected characteristic when referenced along with another protected characteristic. We also protect refugees, migrants, immigrants and asylum seekers from the most severe attacks, though we do allow commentary on and criticism of immigration policies. Similarly, we provide some protections for non-protected characteristics, such as occupation, when they are referenced along with a protected characteristic. Sometimes, based on	
--	---	--

⁸⁵ <https://transparency.meta.com/en-gb/policies/community-standards/hate-speech/>

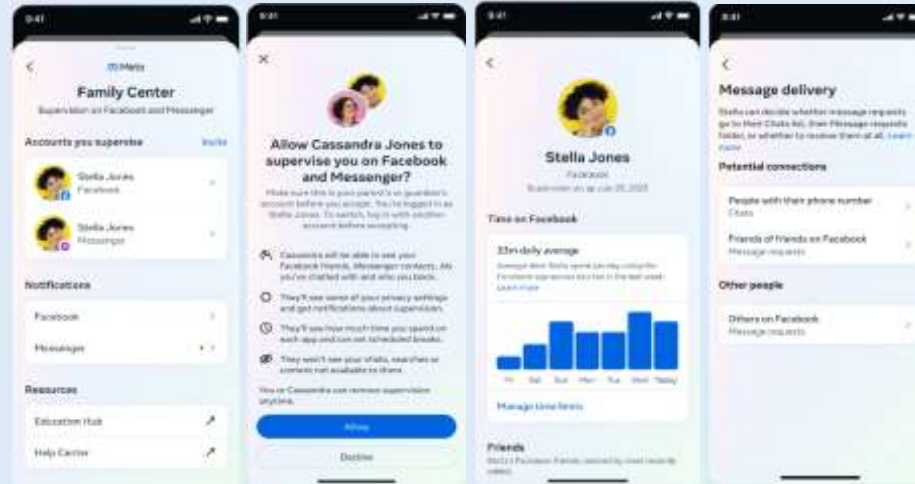
	local nuance, we consider certain words or phrases as frequently used proxies for protected characteristics. We remove dehumanising speech, allegations of serious immorality or criminality, and slurs. We also remove harmful stereotypes, which we define as dehumanising comparisons that have historically been used to attack, intimidate or exclude specific groups, and that are often linked with offline violence. Finally, we remove serious insults, expressions of contempt or disgust, swearing and calls for exclusion or segregation when targeting people based on protected characteristics.	
Paragraph 18: Children must not be targeted to receive content that the Service is reasonably aware to be detrimental to their physical or mental well-being. Such content includes the categories of harmful and/or inappropriate content in paragraphs 4 and 17. In this regard, content targeting refers, but is not limited to, advertisements, promoted content and content recommendations.	Please provide information on the measures, and include screenshots or evidence where possible. Children: At Meta, child protection is always a top priority. We do not target teens to receive content (such as advertisements, promoted or boosted content, recommended content) in relation to harmful or inappropriate content categories below: a. Sexual content b. Violent content c. Suicide and self-harm content d. Cyberbullying content e. Content endangering public health f. Content facilitating vice and organised crime <u>Content recommendations:</u> Please refer to our responses at paragraphs 14 and 17(a)(b)(c)(d) for the types of content we would not recommend to teens. <u>Advertisements and Promoted Content:</u> The content categories above are listed as unacceptable per our Advertising Standards⁸⁶ and ads related to such content are prohibited on Facebook. We have taken various steps to create age-appropriate ad experience for teens. Our Advertising Standards already prohibit ads about restricted	

⁸⁶ <https://transparency.meta.com/policies/ad-standards>

	topics, like alcohol, financial products and weight loss products and services, to be shown to teens (people under 18). The topics we already restrict in our policies will be defaulted to see less, so that teens can't choose to opt into content that may not be age-appropriate.	
Paragraph 19: Children or their parents/ guardians must have access to tools that enable them to manage children's safety, and effectively minimise children's exposure to, and mitigate the impact of, harmful and/ or inappropriate content and unwanted interactions on the Service. These tools may include the following: (a): Tools to effectively manage the content that children see and/or their experiences. (b): Tools to: Limit the public visibility of children's accounts, including their profile and content;	Please provide information on the measures, and include screenshots or evidence where possible. Parents and Children: Paragraph 19(a): Tools to effectively manage the content that children see and/or their experiences. Parental supervision on Facebook <u>Parental supervision</u>⁸⁷ is a set of tools and insights that parents and guardians can use to help support 13 to 17 year olds on Facebook with more positive and age appropriate experiences. Starting in May 2024, we've made it simpler for parents to see and manage all the teen accounts they supervise across different apps - Facebook, Messenger, Instagram and Meta Quest - in one place with the Family Centre. When supervision has been set up, a parent can: ● See how much time their teen has spent on Facebook each day for the last week, and their average daily time spent for the week ● Set scheduled breaks for their teens ● See their teen's Facebook friends ● See some of their teen's privacy settings and content preferences ● See the people and Pages that their teen has blocked	

⁸⁷ <https://about.fb.com/news/2023/06/parental-supervision-and-teen-time-management-on-metas-apps/>

- ii. Limit who can contact and/or interact with children's accounts; and
- iii. Limit location sharing



- Parents using supervision will be prompted to approve or deny their teens' (under 16) requests to change their default safety and privacy settings to a less strict state – rather than just being notified of the change. Our content recommendation controls – known as “Sensitive content control” on Facebook – makes it more difficult for teens to come across potentially sensitive content or accounts in places like Search and Explore.

Children:

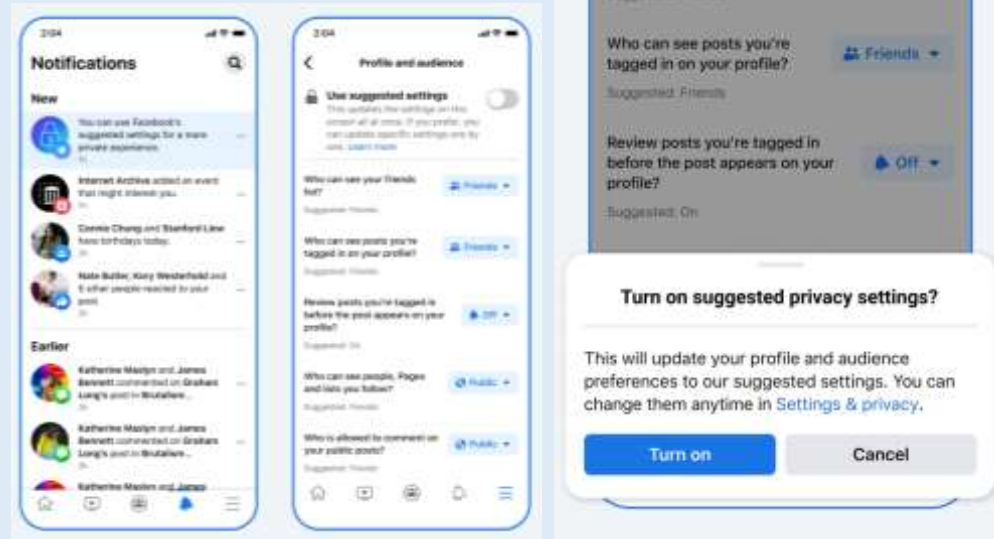
Paragraph 19(b): Tools to:

- i. Limit the public visibility of children's accounts, including their profile and content;
- ii. Limit who can contact and/or interact with children's accounts; and
- iii. Limit location sharing;

The measures listed in our responses to paragraph 12(a), (b) and (c) are available to teens. The measures that we outline below are additional protections that we provide to teens.

We work hard to help keep people on Facebook safe. Facebook requires everyone to be at least 13 years old before they can create an account. Our privacy and visibility settings for teens are generally

	more restrictive than the setting for adults. We have zero tolerance towards child exploitation and take proactive and aggressive measures to protect our young users from predators. Parents using the Parental Supervision tool (see paragraph 19(a)) will be prompted to approve or deny their teens' (under 16) requests to change their default safety and privacy settings to a less strict state – rather than just being notified of the change. (i) Limit the public visibility of children's accounts, including their profile and content: Everyone who is under the age of 16 will be defaulted into more private settings when they join Facebook, including: • Who can see their friends list • Who can see the people, Pages and lists that they follow • Who can see posts that they're tagged in on their profile • Reviewing posts that they're tagged in before the post appears on their profile • Who is allowed to comment on their public posts • We protect certain information, such as minors' contact info, school and date of birth, from appearing in search to a public audience • We provide teens with specific education about what it means to post publicly. Additionally, we take steps to remind teens that they should only accept friend requests from people they know	
--	---	--



(ii) Limit who can contact and/or interact with children's accounts:

How people can find you on Facebook: This option in the Privacy Centre on Facebook will let teens or their parents or guardians review ways in which people can look teens up on Facebook and who can send them friend requests. Parents or guardians can use the parental supervision tools, as mentioned in response to paragraph 19(a), to limit visibility of their teen's account and protect them from unwanted interaction.

We restrict adults over 18 from starting private chats with teens they're not connected to on Facebook. We also limit the message requests people can send to someone who doesn't follow or isn't already connected to them, to text-only messages. People can't send images, videos, or voice notes in message requests until the recipient accepts the request to chat.

During the reporting period we announced that we'll soon begin making Teen Accounts available on Facebook and Messenger which will be applicable for the next reporting period. Teen Accounts on Facebook and Messenger will offer similar, automatic protections to limit inappropriate content and

	unwanted contact, as well as ways to ensure teens' time is well spent, as Teen Accounts on Instagram (see Instagram report for more details). (iii) Limit location sharing: Location sharing is off for teens by default. When either an adult or teen turns on location sharing, we include a consistent indicator as a reminder that they're sharing their location. Please refer to our response to paragraph 12(c) for details on tools to limit location sharing.	
Paragraph 20: Unless the Service restricts access by children, children must be provided differentiated accounts whereby the settings for the tools to minimise exposure and mitigate impact of harmful and/or inappropriate content and unwanted interactions are robust and set to more restrictive levels that are age appropriate by default. Children or their parent/ guardians must be provided clear warnings of implications if they opt out of the default settings.	Please provide information on the measures, and include screenshots or evidence where possible. Children: Please refer to our responses to paragraphs 17, 18, 19(a) and (b). Additional Facebook tools Meta rolled out Instagram Teen Accounts during the reporting period and announced that soon we'll begin making Teen Accounts available on Facebook and Messenger thereafter. Teen Accounts on Facebook and Messenger will offer similar, automatic protections to limit inappropriate content and unwanted contact, as well as ways to ensure teens' time is well spent (see Instagram report for more details).	

Section B: User Reporting and Resolution

Paragraph 23: Any individual must be able to report concerning content or unwanted interactions to the Service in relation to the categories of harmful and/or inappropriate content in paragraphs 4 and 17. In this regard, the reporting and resolution mechanism provided to end-users must be effective, transparent, easy to access, and easy to use.

Please provide information on the measures, and include screenshots or evidence where possible.

All end-users in Singapore:

In response to the 2024 Online Safety Assessment Report findings, we are working through different measures related to improved handling of in-app user reports from Singapore. We look forward to continued partnership with the IMDA on assessing the performance of our in-app user reporting system.

Meta continuously works to enhance the effectiveness of its user reporting systems on Facebook, in line with its commitments to user safety and platform integrity and regulatory obligations. The roll-out of different measures is iterative and based on technical readiness. The combined application of these measures is aimed at improving the effectiveness and efficiency of Meta's enforcement and ensuring more timely responses to user reports on Facebook.

Enhancing the user reporting experience to improve reporting quality

Meta has expanded the use of its reporting flow interface to make it easier for users to accurately describe the type of harm they are encountering, and to empower users to report issues more precisely and intuitively. Meta is also improving users' ability to report impersonation, simplifying the process to both locate the reporting function and identify the victim of the impersonation. Meta continues to invest in the stability and dependability of its reporting systems to ensure a reliable user experience.

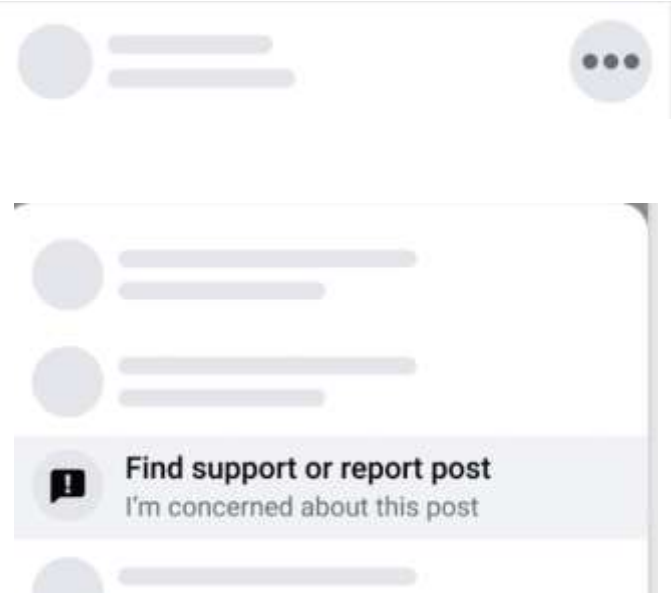
Improving enforcement

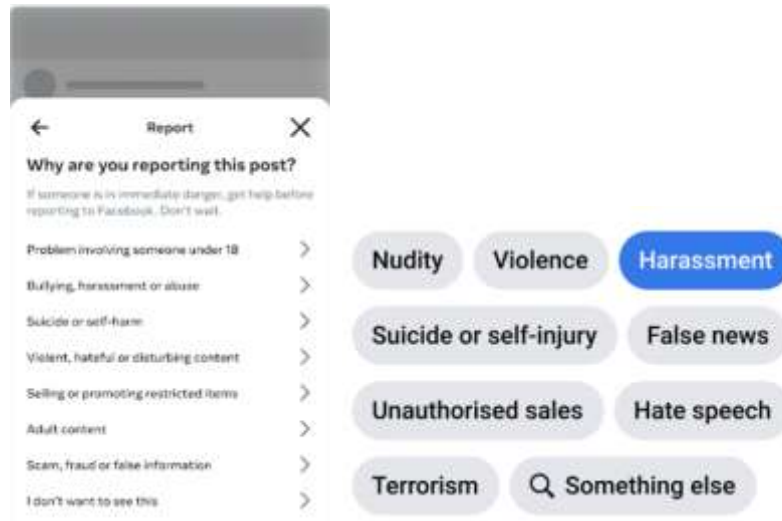
Meta continues to take measures to improve the effectiveness and efficiency of enforcement on user reports. These measures include launching high-precision AI models that automatically remove entities posting violating content, increasing both the speed and accuracy of enforcement. Meta continues to optimize human review processes by prioritizing the most harmful content and applying

	advanced AI ranking models across surfaces and entity types. This approach ensures that reports about the most harmful content receive prompt attention, enhancing the effectiveness of Meta’s systems and enforcement. Additionally, Meta has made various tooling enhancements to improve the effectiveness and efficiency of its human review teams. Improving communication This year, Meta also made improvements to its systems to reduce the time it takes to communicate report outcomes to users. Meta’s systems now check for review outcomes more frequently, which means that users will receive a faster response from Meta regarding the outcome of their report without a compromise in review quality. Reporting on Facebook Meta offers a variety of reporting tools and channels on our technologies, both in-app or via separate reporting channels, for users and non-users to report content they believe violates our policies or infringes on their rights. We also have help articles that inform users of all the available reporting channels and how to report⁸⁸. When something gets reported on Facebook, we review it and remove anything that goes against our Community Standards⁸⁹ and other applicable policies. How users can report content on Facebook Users always have an option to report, whether it's a photo, profile, ad, post, hashtag, comment, story, message or something else. Users can report content using the Report button near the content itself. When a user wants to report an account, group or Page rather than a piece of content, the Report button is located underneath the respective profile, group or Page name.	
--	---	--

⁸⁸ https://www.facebook.com/help/1753719584844061/?helpref=hc_fnav

⁸⁹ <https://www.facebook.com/communitystandards>

	 The screenshot shows a mobile app interface. At the top, there's a header bar with a circular profile picture placeholder on the left and a three-dot menu icon on the right. Below the header, there's a post area. The post includes a circular profile picture placeholder, a text line, and another text line. Below the text, there's a light blue button with a speech bubble icon containing an exclamation mark. The button text reads "Find support or report post" in bold, and "I'm concerned about this post" in a smaller font. Below the button, there's another circular profile picture placeholder and a text line. The entire interface is shown within a rounded rectangle, suggesting it's a mobile screen. Users are then directed to follow on-screen instructions and select an option from a dropdown menu that best describes how this post goes against the Community Standards. This helps us send the report to the correct review team.	
--	---	--



After these steps, we submit the report. We also lay out what people should expect next.

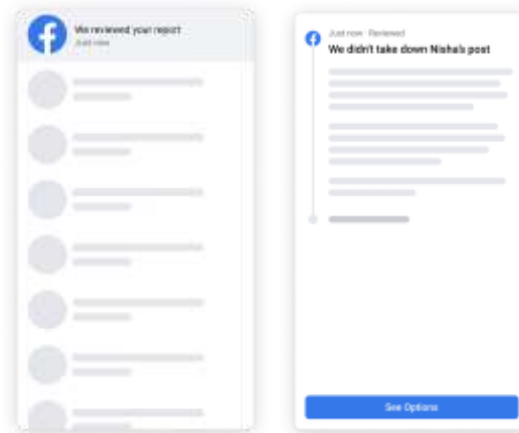


	Our Help Centre⁹⁰ also provides further details to users on how to report violating content or profiles. All end-users in Singapore: Paragraph 23(a): End-users' reports must be assessed, and appropriate action(s) must be taken by the Service in a timely and diligent manner that is proportionate to the severity or imminence of the potential harm. In particular, timelines must be expedited for content and activity related to terrorism. Appropriate action(s) may include: i. Swiftly removing the reported content or restricting access to the reported content; and ii. Warning, suspending, or banning the account(s) that generated, uploaded or shared the reported content; Review processes When something gets reported on Facebook, we review it and remove anything that goes against our Community Standards. As part of our content moderation efforts, we employ a combination of human review, technology and data. Every day, we remove millions of violating pieces of content and accounts on Facebook. In most cases, this happens automatically, with technology to detect, restrict, and remove content and accounts that may go against our Community Standards. In other cases, our technology selects content for human review. Our review teams review a blend of user reports and content surfaced by our technology. Our technology also supports the review teams by prioritising the most critical content to be reviewed, based on severity, virality, and likelihood of a violation. When we receive a report, we first review it against our Community Standards and other applicable policies. If we determine that the content goes against our policies, we remove it. Instead of simply looking at reported content in chronological order, we prioritise the most critical content to be reviewed (e.g. posts related to child sexual exploitation, terrorism and suicide). We prioritise the content that is most harmful to users based on multiple factors such as virality, severity of harm and likelihood of violation. In an instance where our systems are near-certain that content is breaking our	
--	--	--

⁹⁰ https://www.facebook.com/help/1380418588640631/?helpref=search&query=report%20content&search_session_id=d3dcf92016d8fc7a3acbbe9cf9266f02&sr=0

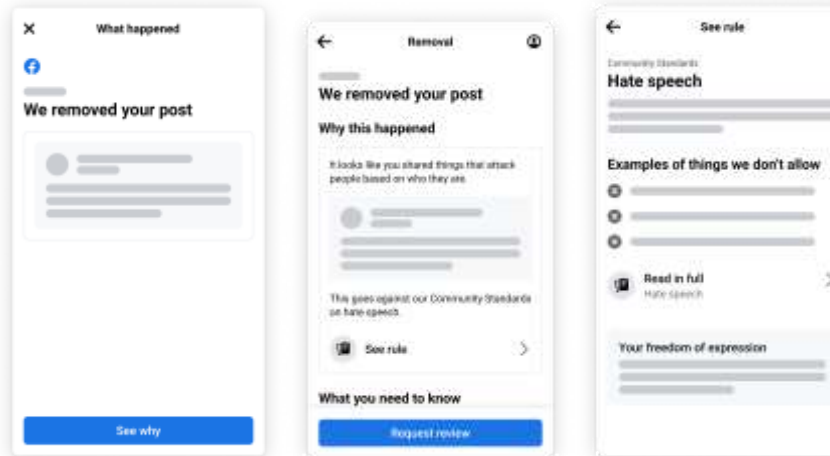
	rules, it may remove it. Where there is less certainty, it will prioritise the content for teams to review and action in a timely manner based on severity. We've increased our capability to rapidly respond to live streams that may breach our Community Standards, including by reviewing all live streams in an area that may involve footage of an attack and increasing our 24/7 capacity to respond to live stream reports. Use of Artificial Intelligence With appropriate oversight, we also use artificial intelligence to augment and scale our human review capacity. As mentioned above, we use artificial intelligence to select the content for human review on the basis of severity, virality, and likelihood of a violation. As with matching technology, artificial intelligence operates on URLs, text, images, audio, and videos. Unlike technologies that can only match violations they've seen before, artificial intelligence has the potential to identify certain violations it has never seen before. Enforcement If users post content that goes against our Community Standards, we'll remove it and may then apply a strike to the user's Facebook account. Meta informs users when something they posted goes against our Community Standards. Usually, this notice appears in Feed on logging into Facebook. Users can also find this notification in their Support Inbox. (See our response to paragraph 23(b) below for more details on the communication we show to the user.)	
--	--	--

	Please provide information on the measures, and include screenshots or evidence where possible. All end-users in Singapore: Paragraph 23(b): Where Meta receives a report that is not frivolous or vexatious: - The end-user who submitted the report must be informed of decision and action taken by Meta with respect to that report without undue delay; - Should Meta decide to take action against the report content or account(s), the end-user holding the account(s) that generated, uploaded, or shared the reported content must be informed of decision and action taken by Meta without undue delay; Post report communication When a user reports a piece of content (e.g. a photo, comment, post) to us, we'll take action to restrict access to the content or remove the content from Facebook if we find that it goes against our Community Standards or other applicable policies. When we've reviewed the content, we'll let the individual or entity know immediately whether it did or did not go against our policies in the user's Support Inbox. This notification is instantaneous, and users can see the current status of their report, and action taken. If the individual or entity reports content but we find that the content does not go against our Community Standards or other applicable policies, we will let the individual or entity know in the user's Support Inbox. At that time, if the individual or entity disagrees with our decision to leave the content up, we offer the individual or entity the opportunity to request another review. With the Support Inbox, reporting users can: - Click on any report to learn more about our policies. - Cancel a report by clicking on the report that they want to cancel and select the Delete report option. - See when Meta took action on the report and the decision we made. - Request a review of our decision.	
--	--	--



Take down experience for creators of reported content

- When a user posts something that doesn't follow our rules, we'll inform them.
- We'll also address common misperceptions and explain why we made the decision to enforce.
- We'll give users easy-to-understand explanations about the relevant rule.
- If the user disagrees with the decision, they can ask for another review and provide more information.
- We set expectations about what will happen after the review has been submitted.



Please provide information on the measures, and include screenshots or evidence where possible.

All end-users in Singapore:

Paragraph 23(c): The end-users referred to in sub-paragraphs (b)(i) and (b)(ii) must be allowed to submit requests to the Service for a review of the decision and action taken.

Meta has complaints and appeal mechanisms which have long been in place and made available to reporters of content and users who are affected by decisions in relation to content and accounts.

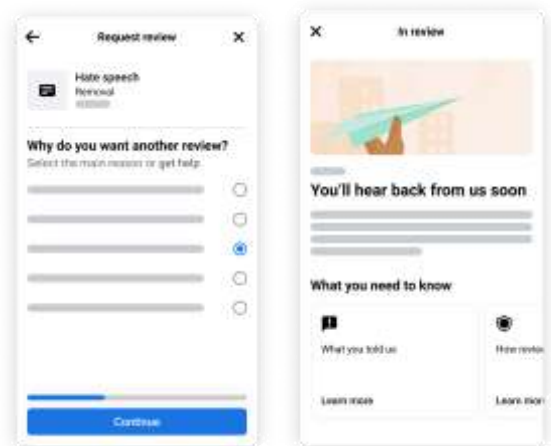
These appeal mechanisms have been expanded even further and are available as it relates to the following circumstances:

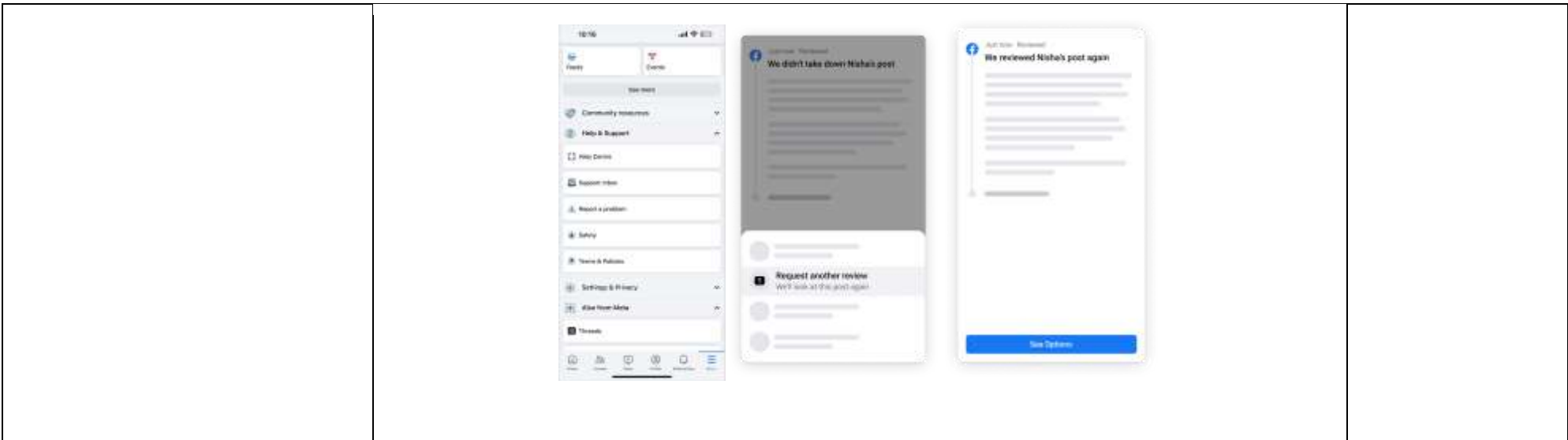
- When we remove or restrict content for violating the Community Standards or other policies
- When we restrict content in response to a legal report
- When we limit monetization features
- When we apply a feature limit to an account; and
- When we suspend or terminate an account

When a reporter or an individual who posted content that was removed requests an appeal, a different review team will review the decision. If a reviewer accepts the original reviewer’s decision, the content remains off Facebook. However, if the reviewer disagrees with the initial review and decides it should not have been removed, the content will go to another reviewer. This reviewer's decision will determine whether the content should be on Facebook or not. The content is not visible to other people on Facebook while we review it again.

As shown in the below screenshots, the users whose content we removed can appeal for another review and provide more information. We set expectations about what will happen after the review has been submitted.

They can check for updates in their Support Inbox, as mentioned in the response for sub-paragraph 23(b).





Section C: Accountability - Mandatory Information and Metrics

Paragraph 24: End-users must have access to clear and easily comprehensible information that enable them to assess the level of safety and related safety measures afforded by the Service and make informed choices.

Please provide information on the measures, and include screenshots or evidence where possible.

All end-users in Singapore:

At Meta, we give people control over what they share, who they share it with, the content they see and experience and who can contact them and share with us their content preferences. Our library of tools and resources for improved online safety supports and reflects our Community Standards and enables users to stay informed about safety in general and tools that they can use to assess the level of safety and protect themselves. We care about the safety of our global community.

Below are some examples of our [safety tools](#)⁹¹ that help to keep our users' accounts secure and protect their privacy, which we detail in the preceding sections:

- Audience selection for controlled sharing
- Friending and following
- Unfriending and unfollowing
- Blocking
- Restrict accounts
- Hiding posts
- Manage comments
- Report abuse

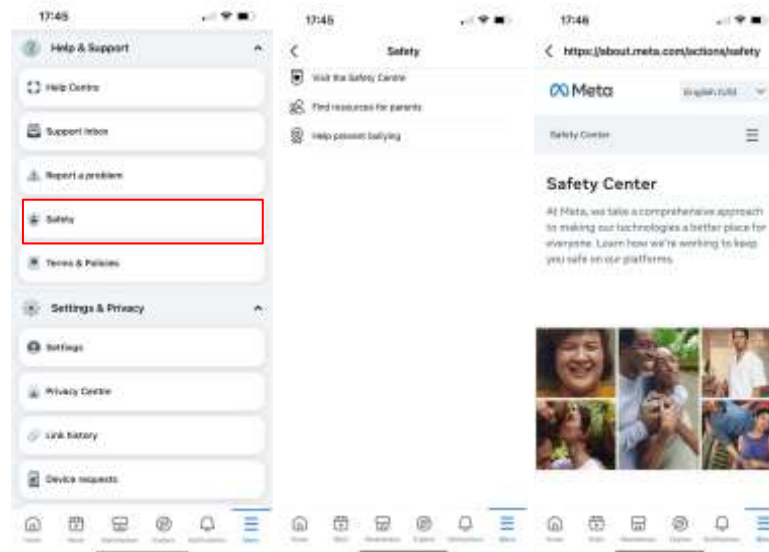
The [Meta Safety Centre](#)⁹² (also accessible in-app under Help and Support; see screenshots below) provides information to users and people more broadly about safety tools available to users of the service and information about our approach to online child protection, online well-being, intimate image abuse and bullying and harassment. It also provides tailored guidance to specific user groups, such as women, teens and parents. Meta created many of these resources through collaborations with experts in a range of issues.

⁹¹ <https://about.meta.com/actions/safety/topics/safety-basics/tools/stay-safe>

⁹² <https://about.meta.com/actions/safety>

We know parents and guardians often think about what their children are doing online and how they can help them navigate social media safely, including how they can manage their children's access and exposure to illegal and restricted online content. This is why we created [safety resources for parents](https://www.facebook.com/help/1079477105456277)⁹³ on Facebook (accessible in-app under Help and Support; see screenshots below). These resources were created in partnership with safety, youth and parenting organisations, and experts around the world and cover managing privacy, interactions, time and security on our applications. They also explain the basics of our app and a description of our safety tools and settings.

Safety Centre (accessible in-app):



⁹³ <https://www.facebook.com/help/1079477105456277>

Resources for Parents in the Safety Centre (accessible in-app):



More information for parents and guardians can be found in [Family Centre](#)⁹⁴ and additional resources on supportive online experiences can be found in the [education hub](#)⁹⁵. Family Centre is accessible via the Settings under Tools and resources which can be accessed in-app and online.

Meta has a dedicated [Privacy Centre](#)⁹⁶ (accessible in-app under Settings and Privacy; see screenshots below) where users can find information about different privacy settings available and share information about increasing privacy, including account security, audience selection and location sharing.

⁹⁴ <https://familycenter.meta.com/>
⁹⁵ <https://familycenter.meta.com/in/education/>
⁹⁶ <https://www.facebook.com/privacy/center/>



Meta also publishes regular reports on the [Transparency Centre](https://transparency.meta.com/reports/)⁹⁷ to give our community visibility into how we enforce our policies, and includes data on how much content we remove for violating our policies that we publish on a quarterly basis.

Children:

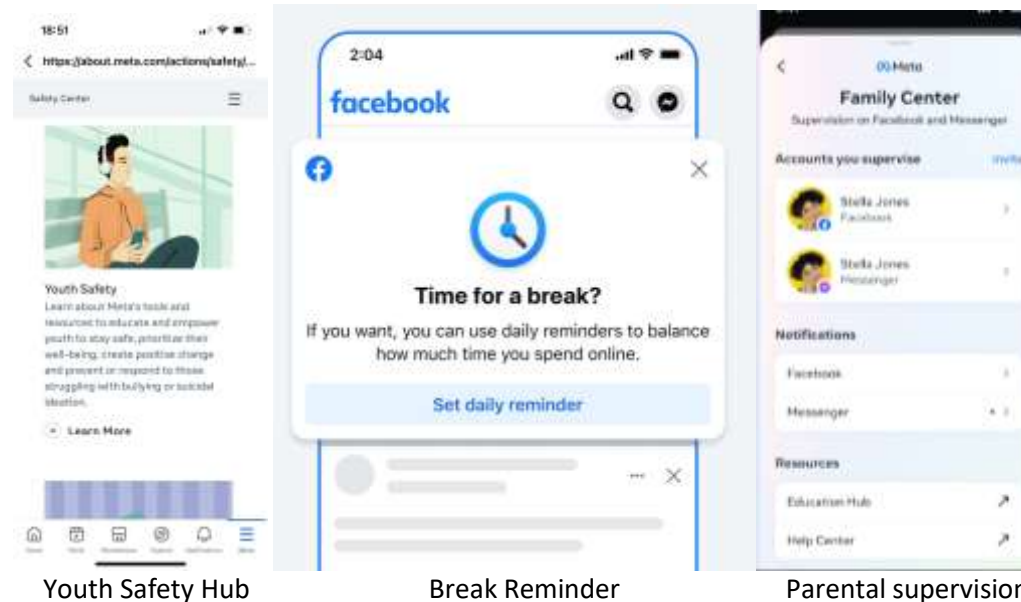
We've built more than 50 tools, features and resources that help teens have safe, positive experiences and give parents simple ways to set boundaries for their teens, several of which we have detailed in the preceding sections:

- Ad topic controls to manage ads
- Launched “Take It Down” with NCMEC to prevent spread of intimate images of teens
- Parental Supervision Tools
- Private messaging restrictions for adults above 18 with teens
- Nudging teens to manage their time on Facebook

⁹⁷ <https://transparency.meta.com/reports/>

- Hiding age-inappropriate content
- Stricter default private messaging settings
- Published a page in Transparency centre to show our approach to age-appropriate content

A comprehensive list of tools and features to support teens can be found [here](https://www.meta.com/en-gb/help/policies/safety/tools-support-teens-parents/)⁹⁸. Information about these tools are also accessible via the [Meta Safety Centre](https://about.meta.com/actions/safety)⁹⁹ (also accessible in-app under Help and Support; see screenshots below), which also provides tailored guidance to specific user groups, including teens.



⁹⁸ <https://www.meta.com/en-gb/help/policies/safety/tools-support-teens-parents/>

⁹⁹ <https://about.meta.com/actions/safety>

Paragraph 25(b): How much and what types of harmful or inappropriate content end-users in Singapore encounter on the Service	Please provide information and data metrics for: All end-users in Singapore: Please refer to our response to paragraph 26(c) below for the various types of content actioned for end-users in Singapore for violating our Community Standards. Children: Please refer to our response to paragraph 26(c) below.	
---	---	--

Section C: Accountability - Additional Information and metrics

Paragraph 26(a): [Please insert data metric(s) that IMDA has agreed to] The Service may propose suitable information and metrics to be included in its annual reports subject to agreement by IMDA. These may include but are not limited to: a. The number and types of end-user reports received from end-users in Singapore, and the number and types of harmful and inappropriate content removed as a result of end-user reports;		All end-users in Singapore: From April 1, 2024 to March 31, 2025, there were over 3.2 million pieces of content reported to us by users from Singapore on Facebook. We took action on over 130.1 thousand pieces of that content for violating our Community Standards.	We have made substantial investments to build systems and tools for logging data needed to provide accurate and sound metrics that meaningfully informs stakeholders on the efficacy of our efforts. We have been publishing Community Standards Enforcement Report¹⁰⁰ on a quarterly basis in our Transparency Centre¹⁰¹ documenting the efficacy of our efforts since 2018. Our focus has been on logging data on how well we are able to proactively detect and remove content
		Children: As a global social technology company with technologies where digital communications are borderless, our approach to measuring the efficacy of our efforts is to start with global enforcement metrics aimed at answering the question of how well Meta is at taking action on content and accounts that violate our Community Standards and other policies globally. Of the 3.2 million number reported above for “All end-users”, we are not able to provide a breakdown of how many of those were reported by users who were below 18 years old (if any). By extension, of the 130.1 thousand number reported above for “All end-users”, we are not able to provide a breakdown of the proportion of that content that may have been reported by users who were below 18 years old (if any).	
	Sexual Content	All end-users in Singapore: Please refer to our response to paragraph 26(c).	
		Children: Not available. Please refer to our explanation above.	
	Violent Content	All end-users in Singapore: Please refer to our response to paragraph 26(c).	

¹⁰⁰ <https://transparency.meta.com/reports/community-standards-enforcement/>

¹⁰¹ <https://transparency.meta.com/>

		Children: Not available. Please refer to our explanation above.	that violates our Community Standards, before a user reports it to us. The metrics we have invested in reporting have therefore centred on the enforcement and proactive rates for specific violation types in our Transparency Centre. As such, we are unable to provide a breakdown per violation type arising from reports by end-users in Singapore alone.
	Suicide and self-harm content	All end-users in Singapore: Please refer to our response to paragraph 26(c).	
		Children: Not available. Please refer to our explanation above.	
	Cyberbullying Content	All end-users in Singapore: Please refer to our response to paragraph 26(c).	
		Children: Not available. Please refer to our explanation above.	
	Content endangering public health	All end-users in Singapore: Please refer to our response to paragraph 26(c).	
	Content facilitating vice and organised crime	All end-users in Singapore: Please refer to our response to paragraph 26(c).	

Paragraph 26(b): [Please insert data metric(s) that IMDA has agreed to] The Service may propose suitable information and metrics to be included in its annual reports subject to agreement by IMDA. These may include but are not limited to: b. The time between the Service receiving end-user reports from end-users in Singapore on harmful and inappropriate content and taking action (if any) as an aggregate;		All end-users in Singapore: Please refer to our response at 26(a).	
		Children: Please refer to our response at 26(a).	
	Sexual Content	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
		Children: Not available. Please refer to the additional comments for explanation.	
	Violent Content	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
		Children: Not available. Please refer to the additional comments for explanation.	
	Suicide and self-harm content	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
		Children: Not available. Please refer to the additional comments for explanation.	
	Cyberbullying Content	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
		All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	

		Children: Not available. Please refer to the additional comments for explanation.	
	Content endangering public health	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
	Content facilitating vice and organised crime	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
Paragraph 26(c): <i>[Please insert data metric(s) that IMDA has agreed to]</i> <i>The Service may propose suitable information and metrics to be included in its annual reports subject to agreement by IMDA. These may include but are not limited to:</i>	Sexual Content	All end-users in Singapore: - From April 1, 2024 to March 31, 2025, we took action on over 193.9 million pieces of content on Facebook globally, including over 215.6 thousand pieces of content created in Singapore, for violating our Community Standards on Adult Nudity and Sexual Activity¹⁰². - Globally, we proactively detected about 94.5 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 5.5 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. We calculate the proactive rate as the number of pieces of content we took action on that we found and flagged before people using Facebook reported them, divided by the total number of pieces of content that we took action on. We use this metric as an indicator of how effectively we are proactively detecting violations. - For content created in Singapore, we proactively detected about 92.9 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 7.1 percent of content that was not proactively detected reflects content	

¹⁰² <https://transparency.meta.com/en-gb/policies/community-standards/adult-nudity-sexual-activity/>

<i>c. The number and types of harmful or inappropriate content proactively removed by the Service that are:</i> <i>i. Accessible by end-users in Singapore; and</i> <i>ii. Originated from Singapore.</i>		that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. - From April 1, 2024 to March 31, 2025, we took action on over 6.4 million pieces of content on Facebook globally, including over 12.3 thousand pieces of content created in Singapore, for violating our Community Standards on Child Endangerment: Nudity and Physical Abuse¹⁰³. - Globally, we proactively detected about 98.3 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 1.7 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. - For content created in Singapore, we proactively detected about 98.5 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 1.5 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. - From April 1, 2024 to March 31, 2025, we took action on over 27.6 million pieces of content on Facebook globally, including over 43.5 thousand pieces of content created in Singapore, for violating our Community Standards on Child Endangerment: Sexual Exploitation¹⁰⁴. - Globally, we proactively detected about 96.8 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 3.2 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. - For content created in Singapore, we proactively detected about 98.6 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 1.4 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports.	
--	--	--	--

¹⁰³ <https://transparency.meta.com/en-gb/policies/community-standards/child-sexual-exploitation-abuse-nudity/>

¹⁰⁴ <https://transparency.meta.com/en-gb/policies/community-standards/child-sexual-exploitation-abuse-nudity/>

		Children: As a global social technology company with technologies where digital communications are borderless, our approach to measuring the efficacy of our efforts is to start with global enforcement metrics aimed at answering the question of how well Meta is at taking action on content and accounts that violate our Community Standards globally. We are not able to provide a breakdown of the proportion of content listed above that may have been accessible to users who were below 18 years old (if any).	
	Violent Content	All end-users in Singapore: - From April 1, 2024 to March 31, 2025, we took action on over 50.5 million pieces of content on Facebook globally, including over 89.2 thousand pieces of content created in Singapore, for violating our Community Standards on Violent and Graphic Content¹⁰⁵. - Globally, we proactively detected about 98.7 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 1.3 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. - For content created in Singapore, we proactively detected about 98.8 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 1.2 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. - From April 1, 2024 to March 31, 2025, we took action on over 22.3 million pieces of content on Facebook globally, including over 48.4 thousand pieces of content created in Singapore, for violating our Community Standards on Violence and Incitement¹⁰⁶.	

¹⁰⁵ <https://transparency.meta.com/en-gb/policies/community-standards/violent-graphic-content/>

¹⁰⁶ <https://transparency.meta.com/en-gb/policies/community-standards/violence-incitement/>

		○ Globally, we proactively detected about 97.5 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 2.5 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. ○ For content created in Singapore, we proactively detected about 97.5 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 2.5 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports.	
		Children: As a global social technology company with technologies where digital communications are borderless, our approach to measuring the efficacy of our efforts is to start with global enforcement metrics aimed at answering the question of how well Meta is at taking action on content and accounts that violate our Community Standards globally. We are not able to provide a breakdown of the proportion of content listed above that may have been accessible to users who were below 18 years old (if any).	
	Suicide and self-harm content	All end-users in Singapore: ● From April 1, 2024 to March 31, 2025, we took action on over 25.6 million pieces of content on Facebook globally, including over 13.9 thousand pieces of content created in Singapore, for violating our Community Standards on Suicide and Self-Injury¹⁰⁷. ○ Globally, we proactively detected about 99.1 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 0.9 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports.	

¹⁰⁷ <https://transparency.meta.com/en-gb/policies/community-standards/suicide-self-injury/>

		○ For content created in Singapore, we proactively detected about 96.9 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 3.1 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports.	
		Children: As a global social technology company with technologies where digital communications are borderless, our approach to measuring the efficacy of our efforts is to start with global enforcement metrics aimed at answering the question of how well Meta is at taking action on content and accounts that violate our Community Standards globally. We are not able to provide a breakdown of the proportion of content listed above that may have been accessible to users who were below 18 years old (if any).	
	Cyberbullying Content	All end-users in Singapore: Bullying and harassment is a unique challenge and one of the most complex issues to address because context is critical. One reason for this is that these behaviours often involve subtle or indirect language, which can make it challenging to identify and remove. Another challenge is that bullying and harassment can take many different forms, and can be directed towards individuals or groups based on a wide range of characteristics. Despite these challenges, Meta takes bullying and harassment seriously and has implemented a range of policies (mentioned in our response to paragraph 17(d)) and tools to help prevent and address these behaviours in its applications. ● From April 1, 2024 to March 31, 2025, we took action on over 27.2 million pieces of content on Facebook globally, including over 79.3 thousand pieces of content created in Singapore, for violating our Community Standards on Bullying and Harassment¹⁰⁸. ○ Globally, we proactively detected about 84 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 16 percent of content that was not proactively detected reflects content that we took action on	

¹⁰⁸ <https://transparency.meta.com/en-gb/policies/community-standards/bullying-harassment/>

		after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. ○ For content created in Singapore, we proactively detected about 83.1 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 16.9 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. ● From April 1, 2024 to March 31, 2025, we took action on over 22.9 million pieces of content on Facebook globally, including over 26.3 thousand pieces of content created in Singapore, for violating our Community Standards on Hateful Conduct¹⁰⁹. ○ Globally, we proactively detected about 94.5 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 5.5 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. ○ For content created in Singapore, we proactively detected about 90.6 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 9.4 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports.	
		Children: As a global social technology company with technologies where digital communications are borderless, our approach to measuring the efficacy of our efforts is to start with global enforcement metrics aimed at answering the question of how well Meta is at taking action on content and accounts that violate our Community Standards globally. We are not able to provide a breakdown of the proportion of content listed above that may have been accessible to users who were below 18 years old (if any).	

¹⁰⁹ <https://transparency.meta.com/en-gb/policies/community-standards/hate-speech/>

	Content endangering public health	All end-users in Singapore: - From April 1, 2024 to March 31, 2025, we took action on over 8.2 million pieces of content on Facebook globally, including over 5.8 thousand pieces of content created in Singapore, for violating our Community Standards on Regulated Goods (Drugs)¹¹⁰. - Globally, we proactively detected about 96.8 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 3.2 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. - For content created in Singapore, we proactively detected about 97.3 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 2.7 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports.	
	Content facilitating vice and organised crime	All end-users in Singapore: - From April 1, 2024 to March 31, 2025, we took action on over 6.2 million pieces of content on Facebook globally, including over 6.3 thousand pieces of content created in Singapore, for violating our Community Standards on Regulated Goods (Firearms)¹¹¹. - Globally, we proactively detected about 98 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 2 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. - For content created in Singapore, we proactively detected about 96 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 4 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports.	

¹¹⁰ <https://transparency.meta.com/en-gb/policies/community-standards/regulated-goods/>

¹¹¹ <https://transparency.meta.com/en-gb/policies/community-standards/regulated-goods/>

		● From April 1, 2024 to March 31, 2025, we took action on over 4.4 million pieces of content on Facebook globally, including over 3.5 thousand pieces of content created in Singapore, for violating our Community Standards on Dangerous Organisations and Individuals (Organized Hate)¹¹². ○ Globally, we proactively detected about 96.1 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 3.9 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. ○ For content created in Singapore, we proactively detected about 93.4 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 6.6 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. ● From April 1, 2024 to March 31, 2025, we took action on over 40.5 million pieces of content on Facebook globally, including over 16.9 thousand pieces of content created in Singapore, for violating our Community Standards on Dangerous Organisations and Individuals (Terrorism)¹¹³ activity. ○ Globally, we proactively detected about 99.4 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 0.6 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports. ○ For content created in Singapore, we proactively detected about 96.1 percent of this content before it was reported to us by users. To respond to paragraph 26(a), the remaining 3.9 percent of content that was not proactively detected reflects content that we took action on after user reports were received, but it does not necessarily indicate that we took action only because of those user reports.	
--	--	--	--

¹¹² <https://transparency.meta.com/en-gb/policies/community-standards/dangerous-individuals-organizations/>

¹¹³ <https://transparency.meta.com/en-gb/policies/community-standards/dangerous-individuals-organizations/>

		Year-on-Year Changes We made changes in our measurement methodologies to more accurately capture user location for country-specific metrics, which resulted in, on the whole, a larger pool of Singapore users being captured. We believe this was a major contributing factor to the general increase in numbers recorded across the categories for Singapore. That said, it does not rule out the impact of external factors and other improvements made to our content moderation solutions. We set out below changes made to our policies and enforcement approach which may provide some further context to the changes seen in enforcement numbers applicable during the reporting period from April 1, 2024 to March 31, 2025, focusing on Child Endangerment and Dangerous Organisations and Individuals (Organised Hate): Child Endangerment: There were several changes to our Child Sexual Exploitation, Abuse, and Nudity (CSEAN) policy and enforcement approach: 1. Our policy was amended to prohibit real and non-real coercion and threats to expose intimate images, and other actions such as enticing children to engage in sexual activity. This change was to account for the increase in sextortion scams and AI-generated CSEAN. 2. We expanded our policy as it applies to content that sexualises children and content that features children and sexualised elements. We remove content that features stimulation of nipples, by-products of sexual activity and fetish content. Dangerous Organisations and Individuals (Organised Hate): Dangerous Organisations and Individuals is a dynamic policy area that is continuously evolving to keep up with the threat landscape. While no substantive changes were made to our policies or enforcement approach, our teams continue to work with external stakeholders such as civil society partners and governments to ensure content that glorifies, supports, and represents these entities is removed. This list of designated entities continues to evolve, and recently we have also added a number of criminal organisations and companies linked to criminal scam syndicates. Over the past year, significant investments¹¹⁴ have been made in this space to ensure we are able to take action against scam compounds by designating them under our DOI policies. We also note that there were	
--	--	---	--

¹¹⁴ <https://about.fb.com/news/2024/11/cracking-down-organized-crime-scam-centers/>

		changes in the geo-political landscape over the period that may have influenced the prevalence of violating content (e.g., the Israel-Gaza conflict, and instability in the Middle East). Regulated Goods (Firearms) : No substantive change was made to Meta’s firearms policies. Further detailed information about the policy changes we have made can be found in the Change Log located under each individual policy area in our Community Standards¹¹⁵.	
Paragraph 26(d): [Please insert data metric(s) that IMDA has agreed to] <i>The Service may propose suitable information and metrics to be included in its annual reports subject to agreement by IMDA. These may include but are not limited to:</i> <i>d. The number of accounts suspended or banned in Singapore, and the reasons for suspending or banning accounts in relation to the</i>		Total number of Singapore-created accounts actioned for policy violations as an aggregate (excluding fake accounts). All end-users in Singapore: From April 1, 2024 to March 31, 2025, we disabled over 924.3 thousand user accounts on Facebook created in Singapore, for violating our Community Standards (excluding fake accounts).	For most violations, if users continue to post content that goes against the Community Standards, despite repeated warnings and restrictions, Meta will disable their account. Accounts are generally removed after multiple strikes in which each strike may be a different violation type so it would be difficult and inaccurate to attribute account removals to a specific harm type.
	Sexual Content	Children: As a global social technology company with technologies where digital communications are borderless, our approach to measuring the efficacy of our efforts is to start with global enforcement metrics aimed at answering the question of how well Meta is at taking action on content and accounts that violate our Community Standards globally. We are not able to provide a breakdown of the proportion of accounts listed above that may have been accessible to users who were below 18 years old (if any).	
		All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
		Children: Not available. Please refer to the additional comments for explanation.	

¹¹⁵ <https://www.facebook.com/communitystandards>

<i>categories of harmful and inappropriate content in paragraphs 4 and 17</i>	Violent Content	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
		Children: Not available. Please refer to the additional comments for explanation.	
	Suicide and self-harm content	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
		Children: Not available. Please refer to the additional comments for explanation.	
	Cyberbullying Content	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
		Children: Not available. Please refer to the additional comments for explanation.	
	Content endangering public health	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	
	Content facilitating vice and organised crime	All end-users in Singapore: Not available. Please refer to the additional comments for explanation.	