Meta Responses To Questions on Notice

1. Questions on Notice from the Hearing (Transcript)

The following are drawn from the 4 September 2024 hearing transcript. Page numbers refer to transcript page numbers.

I am interested in what's included in the three per cent count and how Meta defines news. Is it just original news material at source as put on Facebook by *New York Times*, *Sydney Morning Herald*, *ABC* or whoever it might be? Or does Meta also count the commentary and engagement that's promoted by news? You might have people sharing articles, observations or comments around world news and global affairs. Does placement of news trigger a daily conversation? If so, how does Meta measure this? How does this have value for your users or you as a company? *(Ms Claydon MP, p5)*

The 3% figure reflects views to content in Facebook Feed that contains a link to a news article on a publishers' owned and operated app or website, irrespective of who it was posted by. Publishers predominantly share links to their own websites on Facebook so they can drive traffic back to those owned and operated properties which they can monetise through advertising and/or subscriptions. This is the average experience of people using Facebook. Content containing news links may appear in Facebook Feed for a variety of reasons and some people may see more or less content containing news links depending on their preferences.

As we have previously explained to the Committee, people use Facebook, as with all our products, to build connections with the people, content and communities that matter most to them. It is no surprise, therefore, that the vast majority of the content that is displayed on Facebook relates to individual and social experiences of users, which can vary greatly depending on their individual preferences. The 3% figure demonstrates that access to news content is not the core reason people use Facebook. Access to news has minimal impact on the extent to which people continue to use our services As such, we do not monitor and report on the frequency with which users access news on the platform - let alone whether this is direct or indirect, from "traditional media" or otherwise (however that is defined), what type of news users may access or whether content is "news-related" or otherwise..

If news is only three per cent of Facebook feed globally, can we see a breakdown of the remaining 97 per cent of content? I'm extremely worried about that 97 per cent of content where there is no reviewed factual news media content. Does that mean that 97 per cent of your remaining content is very vulnerable to misinformation or disinformation? *(Ms Claydon MP, p5)*

As set out in our previous response, 3% reflects views to content on Facebook Feed that contains a link to a news article on a publishers' owned and operated app or website and is an average figure. The vast majority of Facebook users engage with content meaningful for them - including friends, family, communities, Groups and video content.

The amount of news shared on our services is not connected to our work to combat misinformation. In terms of misinformation, Meta takes a number of steps to ensure that people are connected to reliable information on our platform. We work with third-party fact-checkers - certified through accreditation bodies like the non-partisan International Fact-Checking Network - who review and rate viral misinformation on our apps. We have built the largest global fact-checking network of any platform by partnering with more than 100 independent fact-checking organisations around the world who review content in more than 60 languages. We have contributed more than $150 million to programs supporting our fact-checking efforts since 2016 to combat the spread of misinformation and we will continue to invest in this area. In Australia, this includes partnerships with Australian Associated Press, Agence France Presse and RMIT FactLab.

To provide greater insight on our misinformation measures, Meta is a founding signatory of the DIGI Australian Code of Practice on Disinformation and Misinformation. Under this Code, Meta has committed to safeguards to protect people in Australia against harmful mis- and disinformation, and to adopting a range of scalable measures that reduce its spread and visibility. We have opted into all seven of the Code's objectives across Facebook and Instagram.

To date, Meta has published four transparency reports under the Code, with the latest launched in May 2024. Our 2024 report outlines the steps we took during the 2023 calendar year to meet the 38 commitments we opted into over that reporting period.

Provide 'more rigorous and substantive information on this to place trust in that 3% number' given the University of Canberra survey. *(Ms Daniel MP, p6)*

We do not have more data to share with respect to this figure than we have already shared with the Committee. Put simply, on average news represents a fraction of content viewed on Facebook. People do not principally come to Facebook for news and when they do not see news on the platform, they simply engage with other content. This is evident from what we have seen in Canada. Just as the number of people around the world using our technologies continues to grow, the number of daily and monthly activities on Facebook in Canada has increased since ending news availability. In addition, time spent on Facebook in Canada has continued to grow since ending news availability. If this were not true, we would observe a very different result.

It's important to recognise that there are a lot of different statistics which seek to look at different things. The University of Canberra study does not survey or measure how much news there is on Facebook - it seeks to show based on surveys the proportion of people engaging with forms of news content via different sources - this is based on stated recollection and is not limited to news published by mainstream media companies or news articles.

We think the most relevant reference point is the proportion of views to content that users see on our platform containing links to news articles out of all views globally. Our data reflects this, and the University of Canberra study provides no evidence to contradict that.

How many news publishers would be impacted by Meta pulling news from its service in Australia and how many Australian users would be impacted? *(Ms Claydon MP, p4)*

We are not able to speculate on hypothetical situations.

Who sits on the Product Leadership Team? *(Ms Claydon MP, p18)*

The Product Leadership Team (PLT) at Meta is responsible for setting the overall product strategy and direction for the company. This includes leaders from the following teams: product, product management, finance team, global affairs, finance team, legal, strategy and executive teams.

Provide further information about Meta's adherence to requests from the Australian Government to remove content from its platforms outside the processes provided under the OSA. *(Ms Daniel, p6)*

We receive many requests for different parts of the Australian, state and territory governments with respect to content on our services. When we receive a request, we review it firstly for compliance with our Community Standards and will action it if it violates these policies. However, when something on Facebook or Instagram is reported to us as going against local law, but doesn't go against our Community Standards, we may restrict the content's availability in the country where it is alleged to be unlawful.

In Australia, from July to December 2023, we restricted access to 1,450 pieces of content for violating Australian law and 38 pieces of content were restricted in Australia in response to global restrictions being imposed.

As we disclosed in our Content Restrictions Report, for Australia we restricted access in Australia to over 1,000 items reported by agencies such as the Therapeutic Goods Administration and the Tertiary Education Quality and Standards Agency for allegedly violating local laws on regulated goods and services (for example, the Tertiary Education Quality and Standards Agency Act 2011), and to over 160 items reported by users and legal counsel for defamation. We also restricted access to 3 items reported by the Australian Electoral Commission for alleged violation of local electoral laws. Of these, 1 item was restricted only temporarily. The remaining items were restricted due to alleged violations of other local laws.

The point is that, if we're looking at age assurance under 16, we would want to know how many children under 16 are on the platform, what times of day

they're on the platform and what your estimates are around the number of users under 18, under 16, under 13 and such. How many children under 16 are on the platform and what your estimates are around users under age? *(Ms Daniel, p16)*

We are in the process of considering the eSafety Commissioner's request for information under the Basic Online Safety Expectations (BOSE), which includes a similar question.

In order to assist the Committee, we can confirm that, based on self-reported ages of our Australian monthly active users, less than 10% of Instagram accounts belong to teens under 18, and less than 5% of Facebook accounts belong to teens under 18.

How many scam ads that have impacted users in Australia have been reported? Provide the number of scam ads identified and taken down in Australia. When a scammer puts an ad up, if it gets through your vetting and verification system, are they able to use the various tools on the platform to target specific people—whether it be people over 65, Indigenous Australians or those living in a remote area? Are they able to use tools to target specific groups such as those, and others? Are you able to provide us, on notice, with some data about the revenue and the number of scam ads that you've identified? Obviously there may be some that you have not identified and taken down, but, where you've identified it and taken it down, what I'd like to know is what revenue for Meta was connected to that. … What I'm interested in is: when you've received revenue that you know has been connected to a scam—whether it's a scam that you've taken action against or that someone else has—what do you do with the revenue that was earned, once those scam ads are removed? *(Ms Templeman MP, pp 13-14)*

We use a range of different tactics to prevent scammers. Meta takes a multi-faceted approach to protecting users on our platforms from scams. This includes policies and systems that prohibit or disrupt this type of behaviour across our services, on and off-platform enforcement, tools and features to help people report fraud and better protect themselves, and education campaigns and partnerships with local government and non-governmental stakeholders.

Content that purposefully intends to deceive or exploit others for money violates our policies, and we remove this content when it's found. Beyond removing content, we take a range of responses when we become aware of a scam. By way of example:

- In addition to suspending and deleting accounts, Pages, and ads, and seeking to prevent bad actors from creating new accounts, we have also taken legal action against bad actors responsible for violating our Terms to create real world consequences for their actions on our platforms.
- To have the biggest and most lasting impact, we target investigations and disruption on persistent and organised threat actors using a range of signals including our own detection and incoming reports from trusted partners. Between January 2023 to January 2024, for example, we have taken action against hundreds of thousands of accounts, targeting several countries including Australia.

When a scam occurs, typically our services represent only one part of the attack chain, meaning we do not have visibility of the scam from end to end. While we do not have records available in relation to losses incurred by Australians to scams, to give an overview of the nature of scam reports we have received and actioned locally, we share below details of reports made through our partnerships with Australian regulators and law enforcement.

Since September 2017, we have provided a direct scam reporting channel to the Australian Competition and Consumer Commission's (**ACCC's**) Scamwatch so they can promptly share complaints from Australian consumers with respect to scams (this is in addition to our in-app reporting tools that consumers can use). We have also worked with Australian law enforcement and the Office of the eSafety Commissioner (**eSafety**) in relation to investigations into scam and fraudulent activities.

The damage and cost to our business far outweighs any ad spend, as well as the cost incurred of having to add teams to track and address these bad actors. We believe this kind of misleading content has a negative impact on people's experiences and the platform overall.

Our business relies on providing safe and enjoyable experiences for our users. As fraud and scams degrade the experience for both users, business users, creator communities and advertisers, there is a strong business incentive to address them.

2. Written Questions on Notice from Mr Andrew Wallace MP (received 5 September 2024)

On National Security and Social Cohesion:
Do you accept that your algorithms and recommender systems, which favour engagement and user experience, encourage controversial and extreme content?

We do not agree with this characterisation. As we shared in our submission to the Committee, the word "algorithm" is often used but infrequently defined. In general, an algorithm is just a set of rules that helps computers and other machine-learning models make decisions. Yet, in the context of social media, "algorithms" are often cited as a concern regarding the claimed influence of social media in promoting social polarisation and the spread of mis- and dis-information. These concerns regarding social media algorithms overlook the role of algorithms on our services in ranking and recommending content, the transparency and controls available to users to better understand and manage them.

Content ranking and distribution
At Meta, we use a range of different algorithms to help us rank content. The ones that people are often most familiar with are those that we use to rank content in their Feeds on Facebook and Instagram. Those algorithms that help with ranking play different roles. Some help us find and remove content from our platform that violates our Community Standards, or filter content that is potentially problematic or sensitive. Others help us understand what content is most meaningful to people so we can order it accordingly in their feeds.

It is important to bear in mind that the content people see in their Feeds is not solely due to algorithms: what people see is heavily influenced by their own choices and actions. Content ranking is a dynamic partnership between people and algorithms. Even though the people that use our services play a significant role in the ranking process, we recognise that they are only going to feel comfortable with these algorithmic systems if they have more visibility into how they work and then have the ability to exercise more informed control over them. That is why we have been releasing products, tools and greater transparency about the way algorithms work on our services. Our Content Distribution Guidelines and Recommendation Guidelines on Facebook and Instagram set a higher benchmark than our Community Standards; they apply to content that would not otherwise violate our rules on Facebook and Instagram.

The Content Distribution Guidelines also share more detail on the types of content that we demote in Feed, and likewise for Instagram Feed and Stories. While the Community Standards make it clear what content is removed from our services because we do not allow it, the Content Distribution Guidelines make it clear what content receives reduced distribution because it is problematic or low quality. Many of these guidelines have been shared in various announcements, but in efforts to make them more accessible, we have brought them together in one easy-to-navigate space in our Transparency Center and Help Center.

The changes we make, particularly ones focused on limiting the spread of problematic content, are based on extensive feedback from our global community and external experts.

There are three principal reasons why we might reduce the distribution of content:

- **Responding to People's Direct Feedback.** We listen to people's feedback about what they like and do not like seeing and make changes to their Feeds in response.
- **Incentivising Creators to Invest in High-Quality and Accurate Content.** We want people to have interesting new material to engage with in the long term, so we're working to set incentives that encourage the creation of these types of content.
- **Fostering a Safer Community.** Some content may be problematic or sensitive for our community, regardless of the intent. We'll make this content more difficult for people to encounter.

Providing guidelines for recommendations

Across our apps, we make personalised recommendations to help users discover new communities and content we think they are likely to be interested in. Some examples of our recommendations experiences include Pages You May Like, "Suggested For You" posts in Feed, People You May Know or Groups You Should Join.

Since recommended content does not come from accounts that people have already chosen to follow, it is important that we have high standards for what we recommend. This is why the Recommendation Guidelines on Facebook and Instagram set a higher benchmark than our Community Standards. This helps ensure we don't recommend potentially sensitive content to those who don't explicitly indicate that they wish to see it.

Transparency
As well as providing transparency at the user level, we recognise that there continue to be discussions about the best ways to provide model and systems documentation that enables meaningful transparency around how these systems are trained and operate. Our transparency initiatives at system level include the release of more than 22 AI System Cards that explain how the AI systems in our products work. They give information, for example, about how our AI systems rank content, some of the predictions each system makes to determine what content might be most relevant to users, as well as the controls users can use to help customise their experienc*e.*

How does Meta judge what a 'positive' or 'negative' user experience is?

We want our services to be useful and relevant to the people who use them and that the time that they spend on our services to be intentional and positive. We use a range of different measures to identify if people have a positive or negative experience on our services. We also have rules, as set out in our Community Standards, around what content can be shared and appropriate behavior on our products. For example we provide:

- On platform surveys asking people if they found content they saw useful and/or want to see more of it;
- A range of signals outlined in our Systems Cards about how people interact with content on our services; and
- Our proactive detection rate and prevalence rate for harmful content disclosed in our Community Standards Enforcement Report

These are examples of just some of the measures.

What other data is collected to feed back into algorithms?

With respect to the Facebook Feed ranking algorithm, we use thousands of different signals to make predictions about whether a person will find something more or less valuable. The categories of signals listed below represent the vast majority of the signals currently used in Feed ranking for connected content to make these personalised predictions. Some of the signals include:

- How long a person has been using Facebook
- Language Facebook is being used in
- Location-related information such as IP address and other device signals if you allow us to receive it
- The device and software being used, and other device characteristics; for example, the type of device, details about its operating system, details about its hardware and software, battery level, signal strength etc.
- The number of different posts that a person has shared, for example, videos, photos, reels etc.
- Data specific to the post being ranked
- Data specific to the individual user and the post being ranked
- Data about how a person has interacted with the post
- Data about how a person has interacted with posts similar to the one being ranked
- Data about the user and the actor who created the post
- Data about the user and the actor who shared the post (when different from actor who created the post)

We provide more details about all of these signals in our Transparency Center.

How is Meta responding to the growing scourge of antisemitism online, when it comes to its own platforms?

1. **In particular, which keywords trigger automatic reviews of content?**
2. **Would support for Hamas, Hezbollah or other terror groups trigger a review? What about removal?**
3. **Would support for the atrocities on October 7 trigger a review? What about removal?**
4. **Would holocaust denial, or denial of the atrocities of October 7 trigger a review or removal?**

Since the terrorist attacks by Hamas on Israel last October, and Israel's response in Gaza, expert teams from across our company have been working hard to monitor our platforms and protect people's ability to use our apps to shed light on important developments on the ground.

We do not allow hate speech on Facebook and Instagram. We define hate speech as violent or dehumanizing speech, statements of inferiority, calls for exclusion or segregation based on protected characteristics including race, ethnicity, national origin, and religious affiliation. From April to June 2024, we removed 7.2 million pieces of content on Facebook for violating our hate speech policies, 96.2% of which we removed proactively before anyone reported it to us. While Meta does not publicly disclose the exact list of keywords used for automatic content review, some examples of keywords that may trigger a review include words or phrases that promote or glorify hatred towards individuals or groups based on their race, ethnicity, religion, or other protected characteristics.

Hamas and Hezbollah are designated by the US government as both Foreign Terrorist Organisations and Specially Designated Global Terrorists. They are also designated under our dangerous organisations policy. This means Hamas and Hezbollah are banned from our platforms, and we remove praise and substantive support of them when we become aware of it, while continuing to allow social and political discourse — such as news reporting, human rights related issues, or academic, neutral and condemning discussion.

In 2020, we updated our hate speech policy to prohibit any content that denies or distorts the Holocaust. We also remove claims that individuals are lying about being victims of any terrorist attack, including the October 7th attacks.

We continue to receive feedback from partners globally as well as in Australia on emerging risks and move quickly to address them.

Can you define what a 'social topic' is?

- **Could this include issues around health, government accountability, legislative changes?**
- **If Meta are serious about dispelling misinformation, why would you limit legitimate government, parliamentary or social movements?**
- **Do you accept that this restricts the ability for important movements for transparency, democracy, and media freedom – whether they include political figures like me or activists for gender equality in Afghanistan; for freedom in Taiwan; or for political change in Venezuela?**

We assume your question relates to our approach to political content. In response to feedback from the people who use our services, who have told us they want to see less political content when they are using our services, we have spent years refining our approach to reduce the amount of political content,

The topics included within the political content control are broadly categorized into three main areas:

- *Governments:* This includes content related to the functioning, actions, policies, and personnel of government bodies.
- *Elections:* This encompasses content that discusses electoral processes, campaigns, results, and related political activities.
- *Social Topics:* These are issues that impact society and may include discussions on civil rights, environmental policies, education policies, international relations, natural disasters, violence and crime, and other topics that affect groups of people

For those people who do want to find and interact with political content, we want to make sure they are able to connect with and find it. When ranking political content in Facebook Feed, our AI systems consider personalized signals, like survey responses, that help us understand what is informative, meaningful, or worth people's time. We also consider how likely people are to provide us with negative feedback on posts about political issues when they appear in Facebook Feed. We have shifted away from ranking political content in Facebook Feed based on engagement signals – such as how likely you are to comment on or share content – since we've found that they are not reliable indicators that the content is valuable to someone.

However, people can personalise what they see on Facebook through customization tools; we offer in their Feed Preferences tool and directly in places in their Feed. Anyone can provide direct feedback on a post by selecting 'Show more' or 'Show less' and use 'Reduce' to adjust the degree to which we demote some content. Other tools to manage the content people see includes the Feeds tab, which will rank posts chronologically, or adding people to the Favorites list so a person can always see content from their favorite accounts.

We also offer a "Political Content Control" tool that people can use to ensure that they see more or less political content.

Our commitment to combatting misinformation is focused on removing disinformation and harmful misinformation, and also on our investment in a third party fact-checking network as well down ranking content rated as false by our fact checkers. It is not predicated on the amount of political or social content on our services.

On Children's Access:
What is the total *actual* number of child users – that is, the number of Australians under the age of 18 – on Facebook and Instagram?

We are in the process of considering our response to the eSafety Commissioner's request for information under the Basic Online Safety Expectations (BOSE), which includes a similar question.

However, at this time, we can confirm that, based on self-reported ages of our Australian monthly active users, less than 10% of Instagram accounts belong to teens under 18, and less than 5% of Facebook accounts belong to teens under 18.

How many children under the age of 13 were "kicked off" Instagram and Facebook in Australia in the 2023 calendar year? How many in the year to date?

Under Meta's terms of service, users must be at least 13 years old to access services like Facebook and Instagram. We take steps to enforce this requirement when we learn that a user is under the age of 13, however age verification remains a challenge for many apps. This is why we support both greater parental controls for the use of our services by teens, and greater age control measures at the app store or operating system level so that age verification can operate seamlessly across the app ecosystem. We do not have the data to share in response to this question at this time.

How do you know they're under 13?

We require everyone to be at least 13 years old before they can create an account on Facebook and Instagram. We remove accounts that don't meet our minimum age requirement when we become aware of them.

Understanding users' real age is key to all of these efforts. This information allows us to create new safety features for young people, and helps ensure we provide the right experiences to the right age group.

However, understanding user age is a complex, industry-wide challenge that requires thoughtful industry-wide solutions to appropriately balance privacy, effectiveness, and fairness. For example:

- People misrepresent their age
- It is important to offer privacy-preserving tools and more options than just ID upload to verify age, as not everyone has access to formal documentation or feels comfortable sharing this information online

At Meta, we take a continuous, multi-layered approach to refine our understanding of age throughout a user's online journey, recognising that no single method will work 100% of the time for every user. This includes:

- Requesting users provide their date of birth when they register new accounts, a tool called an age screen. Those who enter their age (under 13) are not allowed to sign up. The age screen is age-neutral (ie. it does not assume that someone is old enough to use our service), and we restrict people who repeatedly try to enter different birthdays into the age screen.
- Allowing anyone to report suspected underage accounts on Instagram and Facebook. We have dedicated channels to review these reports.
- Investing in AI technology to detect likely teens and ensure they receive age-appropriate experiences, for example, restricting adults from sending messages to teen accounts who do not follow them

All those methods have varying degrees of accuracy and also have technical limitations particularly for users under the age of 18 while also requiring personally identifiable information (e.g. ID verification or biometric data). For example:

- Face-based-age-prediction is challenging to implement in relation to granular prediction for U18s due to technical limitations - e.g. 16 vs 15 years old
- AI classifiers do not work for new users; they require the user to have been using the service for a period of time

Those limitations - as well as broader considerations around safety and privacy - mean that we are part of a growing industry alliance (coordinated by the International Center for Missing and Exploited Children (ICMEC)) advocating for industry-wide solutions and standards to age assurance - namely, age assurance at the app store and device/OS level, to supplement existing age assurance measures individual apps have.

On Child Sexual Abuse and Exploitation:
In a Question on Notice, the Member for Flinders asked about the number of Child Sexual Abuse Material reports which have been made by end-users on your platforms in Australia (p37, QoN). You said that from January to March 2024, you actioned 14.4 million pieces of content. Am I correct that this is 14.4 million pieces of content globally, or is this from Australian end-users alone?

Correct - this is a global figure.

What do you think is driving such a high volume of child exploitation on your platforms and on social media broadly?

It is important to distinguish between the sharing of content and the conduct of criminal activity. As you may know, under US law, companies such as Meta are legally obligated to report apparent violations of child sexual exploitation, including child sexual abuse material (CSAM) they become aware of to NCMEC's CyberTipline. In addition to reporting content we become aware of, we have developed sophisticated technology to proactively seek out this content, and as a result we find and report more CSAM to NCMEC than any other service today. We make this technology available to the industry to help protect children from exploitation across the internet.

While NCMEC already publishes the total number of CyberTips it receives from ESPs on an annual basis, we will begin publishing additional data that demonstrates the types of reports we're making to NCMEC. For example, we will start to provide insight into reports made to NCMEC that may include inappropriate interactions with children.

What we found when we analysed our reports to NCMEC was that the vast majority of them were reshares. Specifically, in Q2 2023, we reported the following number of CyberTips to NCMEC from Facebook and Instagram:

- Facebook and Instagram sent over 3.7 million NCMEC Cybertip Reports for child sexual exploitation.
- Of those reports, 48 thousand involved inappropriate interactions with children. Cybertips relating to inappropriate interactions with children

may include an adult soliciting CSAM directly from a minor or attempting to meet and cause harm to a child in person. These CyberTips also include cases where a child is in apparent imminent danger.

- 3.6 million related to shared or re-shared photos and videos that contain child sexual abuse material (CSAM).

This insight led us to develop and promote the campaign “Report It. Don’t Share It.” across our services in many countries, including Australia.

In a previous Public Hearing of this committee, dated Friday, 28 June 2024 (p19, Hansard Transcript), Ms Davis said, when asked about 18+ content, including pornography, that *"We don't have pornography on our site, so let me just correct that statement."* Do you still deny that your platforms are being used to share and indeed, market pornography and adult sexual content, YES or NO?

Our Community Standards prohibit the display of nudity or sexual activity, with careful allowances for real world art and certain medical, educational, and awareness-raising content, which are detailed in our policy. Under our policies, we remove real photographs and videos of nudity and sexual activity, AI- or computer-generated images of nudity and sexual activity, and digital imagery, regardless of whether it looks “photorealistic” (as in, it looks like a real person). We default to removing sexual imagery to prevent the sharing of non-consensual or underage content. Our Commerce policies similarly do not allow the promotion of any form of human trafficking, prostitution, escort, or sexual services.

To help enforce our policies, Meta is investing in technology that can find violating content proactively - and in some cases, prevent it from being shared in the first place. We also use artificial intelligence (AI) and machine learning to proactively detect harmful content before anyone reports it, and sometimes before people even see it.

Many pornographers host links to adult content on X, OnlyFans, JustForFans and others through Instagram’s ‘link in bio’ or link in bio tools. Organised criminals do the same to connect people to their Telegram accounts. What are you doing to stop people from linking their site to pornography or harmful content through those kinds of tools?

We ingest vetted lists of external sites known for hosting CSAM and block access to those sites from our platform. We also block known terms and hashtags that use or share violating content.

We recently enhanced our search system to restrict additional search terms and hashtags associated with this type of content. As new terms are added to our system, the terms are restricted across Facebook and Instagram simultaneously.

Because this is an industry-wide issue, we will continue to work with experts and partners throughout the industry to combat predators on the internet.

Do you accept that the issue is not just that content must be removed, but that it should not be uploadable in the first place?

1. **What are you doing to PREVENT pornographic, sexualised or harmful adult material in comments, as well as PREVENTING this content from being uploaded in the first place?**
2. **What is stopping you from automatically blocking this content at the time of posting?**

We have steadily increased our investment in proactive detection technology over the years such as that, for example, in Q2, 2024, we actioned 32.2 million pieces of adult nudity and sexual activity content on Facebook and 11.9 million pieces of such content on Instagram, in relation to which 95.6% or 98.3% respectively of which we did so proactively, before people reported it.

We continue to build technologies like RIO, WPIE and XLM-R that can help us identify harmful content faster, across languages and content type (i.e. text, image, etc.). These technologies alongside our continued focus on AI technologies help us to scale our efforts quickly in keeping our platforms safe.

As part of our ongoing commitment to transparency and accountability, we provide data about our enforcement work in our Community Standards Enforcement Report, which we publish quarterly. This report includes metrics such as how much content we are actioning, and what percentage was detected proactively. Currently, we report these metrics against 14 policy areas on Facebook and 12 on Instagram.

Our investment in technology to proactively find violating content - and in some cases, prevent it from being shared in the first place - includes investment in industry-leading initiatives. One example is our investment to combat Non-Consensual Intimate Image (NCII).

It has long been our policy on Facebook and Instagram to remove NCII, and in 2018 we began a pilot in 9 countries - including in Australia with the Office of the eSafety Commissioner - to help victims proactively stop the proliferation of their intimate images.

Following the success of this pilot, in 2021 we launched the expansion of the program globally, known as StopNCII.org. StopNCII.org operates in partnership with more than 50 non-governmental organisations around the world, including the Office of the eSafety Commissioner.

This is the first global initiative of its kind to safely and securely help people who are concerned that their intimate images (photos or videos of a person which feature nudity or are sexual in nature) may be shared without their consent.

When someone is concerned their intimate images have been posted or might be posted to online platforms like Facebook or Instagram, they can create a case through StopNCII.org. When they select their image, the tool uses hash-generating technology to assign a unique hash value (a numerical code) to the image, creating a secure digital fingerprint. The original image never leaves the person's device. Only hashes, not the images themselves, are shared with StopNCII.org. If someone tries to upload a matching image on one of the participating tech companies' platforms, they will review the content on their platform to check if it violates their policies and take action accordingly.

We have developed this platform with privacy and security at every step thanks to extensive input from victims, survivors, experts, advocates and other tech partners. By allowing potential victims to access the hashing technology directly we are giving them more privacy and control of their images.

You outlined on page 14 of your submission to this inquiry that Meta already uses biometrics and social network analysis to match content with programs like PDQ and TMK-PDQF. Why can't similar tools be used to prevent the upload of concerning material?

As stated in our submission, we use a combination of technology and behavioural signals to detect child sexual abuse material. We continue to invest in our proactive detection tools, which are constantly evolving. From April to June 2024, we found and actioned 9.7 million pieces of child sexual exploitation content on Facebook and 2.8 million on Instagram, of which we respectively actioned 97.8% and 96.5% proactively, before it was reported to us.

On page 26 of your submission, you indicated that you disable accounts for sextortion, and that you warn users when an account has recently been accused of sextortion. What further action do you take when you identify that someone has engaged in sextortion?

1. **Do you report the issue to the police?**
2. **What about the sextortion of those under 16, or under 18?**
3. **Do you report this to the police?**

We report apparent instances of child exploitation appearing on our site from anywhere in the world to the National Center for Missing and Exploited Children (NCMEC), as required by US law. NCMEC coordinates with law enforcement authorities from around the world. When we have a high priority NCMEC report, we also flag the report number directly to the Australian Centre for Countering Child Exploitation (ACCCE).

We have strict rules against content or behavior that exploits people, including sharing or threatening to share someone's intimate images. We encourage anyone who sees content they think breaks our rules to report it - and we have a dedicated reporting option to use if someone is sharing private images. When we become aware of this content, we work to take action.

We have specialised teams working on combating sextortion. We have identified patterns associated with this behavior, and built automated systems that detect and remove these accounts at scale. We also have dedicated teams that investigate and remove harmful content and report them to NCMEC, in accordance with our terms of service and applicable law. We work with partners, like NCMEC and the International Justice Mission, to help train law enforcement around the world to identify, investigate and respond to these types of cases.

Our work with law enforcement to help people on our platforms stay safe includes, in certain circumstances, providing information to law enforcement officials that will help them respond to emergencies, including those that involve the immediate risk of harm, suicide prevention and the recovery of missing children. We may also supply law enforcement with information to help prevent or respond to fraud and other illegal activity, as well as violations of our policies.

We are committed to working with expert partners to combat sextortion around the world and have been dedicated to this work for many years. For example, Meta is also a founding member of the Lantern program, managed by the Tech Coalition, which enables technology companies to share signals about accounts and behaviors that violate their child safety policies. We provided the Tech Coalition with the technical infrastructure that sits behind the program and continue to maintain it. Participating companies can use this information to conduct investigations on their own platforms and take action. This is incredibly important because we know that predators don't limit themselves to any one platform—so we need to work together to tackle this.

Like many crimes, financial sextortion crosses borders, and over recent years there's been a growing trend of scammers — largely driven by cybercriminals known as Yahoo Boys — targeting people across the internet, both with these and other types of scams. We've banned Yahoo Boys under Meta's Dangerous Organizations and Individuals policy - one of our strictest policies - which means we remove Yahoo Boys' accounts engaged in this criminal activity whenever we become aware of them.

We also recently announced the strategic network disruption of two sets of accounts in Nigeria that were affiliated with Yahoo Boys and were attempting to engage in financial sextortion scams. We removed around 63,000 Instagram accounts in Nigeria attempting to target people with financial sextortion scams, including a coordinated network of around 2,500 accounts. We also removed a set of Facebook accounts, Pages and Groups run by Yahoo Boys that were attempting to organise, recruit and train new scammers.

In Australia, child sexual abuse and exploitation are issues for which ALL adults are mandatory reporters, with very few exceptions. Do you accept that if an adult in your employ is made aware of child sexual exploitation, and does not report it, that you are breaching your social license, duty of care, and legal responsibility to your end-users and the Australian community, YES or NO? What kind of training and support is provided to those in your employ to ensure they comply with their legal duties and to protect their own liability

In compliance with US law, we report apparent instances of child exploitation appearing on our site from anywhere in the world to the National Centre for Missing and Exploited Children (NCMEC). When we have a high priority NCMEC report, we also flag the report number directly to the Australian Centre for Countering Child Exploitation (ACCCE). As noted above, in addition to reporting content we become aware of, we've developed technology to assist with proactively finding and enforcing this content.

We provide guidance and tools to help content reviewers understand the latest behaviors and terms used by predators, in many different languages. For example, content reviewers will now see information about coded terms used in posts they're reviewing to understand the subtext of those terms, and how they're used by predators. This will help content reviewers better recognize this behavior and take action.

We continuously improve our systems to help prioritize reports for content reviewers. For example, we're using technology designed to find child exploitative imagery to prioritize reports that may contain it.

On Mental Health and Wellbeing:
In her 2021 testimony to the US Congress, it was reported by whistleblower Frances Haugen that Facebook had commissioned internal studies about the potential harms of social media on children under 13 years of age. That same whistleblower asserted that instead of acting to counter those harms, the company considered the pre-teen age bracket "a valuable but untapped audience". ***Has Meta conducted any such studies of impacts on users under 13 years of age since 2021?***

We dispute Ms. Haugen's characterizations of Meta's internal research. Meta does not allow people under age 13 on Instagram and Facebook.

Do you acknowledge that Australians are trying to reduce their screen time? What are you doing to protect users from spending too much time on screen?

We want the time people spend on Facebook and Instagram to be intentional and positive, and we have developed tools to help users understand how much time they spend on our platforms so they can better manage their experience. These include:

- **Improving Feed quality.** We have made several changes to Feed to provide more opportunities for meaningful interactions, and reduce passive consumption of low-quality content. We demote things like clickbait headlines and false news. We optimise ranking so posts from the friends you care about most are more likely to appear at the top of your feed. Similarly, our ranking promotes posts that are personally informative. We also redesigned the comments feature to foster better conversations.
- **Activity Dashboard.** The Activity Dashboard was introduced in 2018 to help people manage their time on Facebook and Instagram. The Dashboard allows people to see the average time spent on the app , and allows them to set reminders once they've reached the amount of time they want to spend on the app.
- **Hide Likes on Facebook and Instagram.** We tested hiding like counts to see if it might depressurise people's experience on Instagram. What we heard from people and experts was that not seeing like counts was beneficial for some and annoying to others, particularly because people use like counts to get a sense of what's trending or popular. We now give users the option to hide like counts on all posts they see in their feed. They also have the option to hide like counts on their own posts, so others can't see how many likes their posts get.
- **Take a Break.** In December 2021, we announced a tool called Take a Break which will empower people to make informed decisions about how they are spending their time. If someone has been scrolling for a certain amount of time, we ask them to take a break from Instagram and suggest that they set reminders to take more breaks in the future. We also show them expert-backed tips to help them reflect and reset.

More recently, on 17 September 2024, we announced the introduction of Instagram Teen Accounts, to provide additional in-built protections for Instagram users under 16. These additional protections include time limit reminders (teens will get notifications telling them to leave the app after 60 minutes each day) and sleep mode (sleep mode will be turned on between 10 PM and 7 AM, which will mute notifications overnight and send auto-replies to DMs). Teens under 16 will need a parent's permission to change any of the built-in protections to be less strict within Teen Accounts.

General Comment

Do you still stand by your comments in our previous hearing that you *"…don't think that social media has done harm to our children… [that it] social media provides tremendous benefits"?*

In our testimony on 28 June 2024, we said *'I don't think that social media has done harm to our children. I think that social media provides tremendous benefits. I think that issues of teen mental health are complex and multifactorial. I think that it is our responsibility as a company to ensure that teens can take advantage of those benefits of social media in a safe and positive environment.*
Regardless of what I think of the research or not, we are committed to trying to provide a safe and positive experience. For example, if a teen is struggling with an eating disorder and they're on our platform, we want to try to put in place safeguards to ensure that they have a positive experience and that we aren't contributing or exacerbating that situation the teen may be dealing with…' (pp11-12).

We stand by these comments.

3. Written Questions on Notice from Ms Sharon Claydon MP (received 12 September 2024)

Decisions made when and by whom
In answer to questions taken on notice, about Meta's data and surveys, Meta states that it surveys thousands of users daily on-platform and off-platform, along with many other signals, to inform relevant business decisions.

- **Would Meta please provide the Committee with periodic surveys for the last 24 months about Australian users and news consumption**
- **How many signals does Meta draw user insights from, and please provide a list of the signals.**
- **Please provide the Committee with the annual schedule of Meta's period surveys including information about what topics or issues they cover.**

We continually assess and take business decisions to adapt and evolve our products to deliver the most valuable experience to consumers. To help inform these continual and ongoing changes, we survey thousands of our users daily on-platform and off-platform in a range of different ways, for different purposes, periodically and in an ad hoc way. These survey results may be used as part of research and studies, along with many other signals, to determine relevant insights which then inform business decisions. Given this is vast, complex, continual and evolving, it's not possible or feasible to identify and provide this data.

There are several data points that provide evidence of the shift that has occurred towards user preference for short-form video and creator content. Some of these include:

- As of Q1 2024, video content continues to grow across our platform and it now represents more than 60% of time spent by users on both Facebook and Instagram. Reels remain the primary driver of that growth. Video is not the majority of what publishers post. The majority of publisher posted content contains links.
- When news content was surfaced for users in a dedicated tab on Facebook (Facebook News), the data clearly showed that users did not engage with it and engagement declined dramatically over the time it was available. The number of daily active users of Facebook News in Australia dropped over 80% in 2023. The same was the case for the US.

- We have observed no meaningful impact to user engagement following the restrictions on the viewing and sharing of news content in Canada. Just as the number of people around the world using our technologies continues to grow, the number of daily active users and monthly active users on Facebook in Canada has increased since ending news availability. In addition, time spent on Facebook in Canada has continued to grow since ending news availability.
- We have reported that there has been a decline in the amount of referral traffic to Australian news publishers from Facebook Feed over time – declining from approximately 5.1 billion organic referrals or clicks in 2020 to more than 3.5 billion in the 12 months to March 2022, which declined again to more than 2.3 billion in 2023 – reflecting a continuing shift in user preferences.
- Additionally, we have made changes to the Facebook Feed algorithm to reflect these changing user preferences, and we have not observed any decline to user engagement. For example, in 2018, we announced changes to the Facebook algorithm so that posts on a user's Feed that lead to conversations and interactions between individual users were prioritised. These changes reflected user feedback that they wanted to see less public content like news and wanted to connect with more meaningful posts from friends and family. In February 2021, we announced that we would be reducing the political content on Facebook Feed. This has been in response to feedback from users and consistent with well-being research. In July 2022, we made further changes to content ranking by placing less emphasis on shares and comments for political content.

Why did Meta agree to negotiate with news publishers to pay for news in 2021, but now refuses to pay in 2024? What has changed?

In 2021, we had a new product to bring to market – Facebook News – a dedicated space to connect those people who are interested in seeing news on our services, with publishers. We had a suite of deals to support that product.

Unfortunately, Facebook News was not successful. We saw an 80% drop in usage and made the difficult decision to deprecate the product. Whilst we paid out all of the deals we entered into in support of the product, it did not make sense to renew them.

Facebook once used to encourage and support news publishers to use its services – correct?

- **Why did Facebook do this? Was it because it recognised the value of news in attracting users and driving engagement?**
- **Why does Facebook now assert news has no value? How do you reconcile this notion with Facebook's earlier behaviour?**

Our partnerships team has, in the past, worked with news publishers to respond to their questions and support the development – with the product teams – of various products that were designed specifically for publishers. We did this to test new products, including for user uptake and product fit. One example is the creation of the Facebook News product. Our experience has been that these products and these investments have not translated to a successful product fit despite these efforts and engagement. Instead, we have had to respond to the rapid consumer shift to short form video, and prioritise our investments there.

Influence and social license

At the last hearing Ms Garlick said that "It made no sense" for Meta to continue supporting the Facebook News product in Australia. While Meta is an advertising business that will make commercial decisions, others consider that it makes a lot of sense for a large social media platform to support the news media and civic engagement in democratic society:

- **Does Meta acknowledge the high proportion of users of its social media services in Australia?**
- **Does Meta acknowledge that content shared over its services may influence what Australians see, hear, understand or believe about world events?**
- **Does Meta acknowledge that its platforms operate, to some degree, as a new 'town square' or 'civic space'?**
- **Does Meta acknowledge the importance of the fourth estate to Australian society and democracy?**
- **Does Meta accept that large social media platforms have a social responsibility to support the provision of news in the democratic society in which they operate?**
- **Does Meta accept that removal of news undermines the flow of commentary through civil society?**

News makes up a small proportion of the average person's experience of our services. We invest in the Facebook News product as a way to connect publishers with the small proportion of people who do want to engage with news content on our services. However, it has not been successful, with an 80% drop in usage. Consequently, it did not make sense to continue to invest in the product. Publishers can and do continue to use our services to share their content in Feed. However, Australians have many channels through which to consume news and information – not just on Facebook.

We have consistently expressed concern about legislative models that force digital platforms to pay publishers since we first saw the *Treasury Laws Amendment (News Media and Digital Platforms Mandatory Bargaining Code) Bill 2020* because it misunderstands the economics of our business, especially in relation to news. Specifically, as our July 2020 submission outlined: "[t]he draft law fundamentally misunderstands the economic reality of the value exchange between Facebook and publishers. It is based on the misconceptions that Facebook profits from taking news content with no consent or control by publishers, that we do not pay for news or drive sufficient value for news publishers, and that government intervention is required to correct this."

Forcing companies into contracts that hold no commercial benefit is not going to address the long standing issues the news industry faces.

Measurement and documentation
What documentation, if any, does Meta have about news content and the consumption of news content on its platforms?

- **How does Meta categorise, define and measure what content is on its platforms?**
- **How many Australians spend their time on Meta's platforms each month? Please provide a breakdown by demographic for the last year.**
- **How much time do Australians spend on Meta's platforms each month? Please provide breakdown for the last year.**
- **What categories of content are Australians consuming the most, over the past year?**

As set out in our previous responses, on average less than 3% of what users both globally and in Australia view on Facebook Feed is posts that contain a link to a news article on a publishers' owned and operated app or website. By contrast, and as evidence of the shift that has occurred towards user preferences for short-form video and creator content, as of Q1 2024, video content continues to grow across our platform and it now represents more than 60% of time spent by users on both Facebook and Instagram. Reels remain the primary driver of that growth.

How does Meta measure how changes in its terms of service or algorithms impact the surfacing and consumption of content, including news?

- **How many changes have there been in the last year that impact news publishers on platforms?**
- **How many changes or updates have there been in the last year that impact Australian politicians?**
- **What were these changes and who decided them? Engineers or corporate managers? Please list.**

We continually evaluate the effectiveness of News Feed ranking signals and update or remove them when it makes sense. We share updates (see, for example, a 2021 series of updates) about these changes. We also provide transparency about the signals, predictions and data that inform Facebook Feed ranking in our Transparency Center.

How many signals are in the Facebook Feed algorithm?

- **What are they? Please list.**
- **Does Meta give higher preference to paid or promoted content than unpaid/organic content?**
- **How does Meta decide what to promote and what to demote? What document outlines this approach? Please provide the document to the Committee.**

We use thousands of different signals to make predictions about whether you'll find something more or less valuable. We provide details about some of these in our Transparency Center.

The delivery of ads is via a different system to the ranking algorithm that determines Feed. The Meta ad delivery system uses an ad auction and machine learning to determine where, when and to whom we show your ads. These processes work together to maximize value for both people and businesses.

Our Content Distribution Guidelines provide details about the types of content we demote.

Has Facebook downgraded messages from MPs and Senators, as political content?

- **Has Facebook labelled posts by Australian MPs as spam?**
- **Has Facebook demoted or removed posts by Australian MPs?**
- **Who decides this? Who controls it?**

We want users to have a valuable experience when they use Facebook, Instagram, and Threads, which is why we use AI systems to personalise the content they see based on the choices they make. People have told us they want to see less political content, so we have spent the last few years refining our approach on Facebook to reduce the amount of political content – including from politicians' accounts – users see in Feed, Reels, Watch, Groups You Should Join, and Pages You May Like. We recently extended this approach in Reels, Explore and In-Feed Recommendations on Instagram and Threads, too.

As part of this, we aim to avoid making recommendations that could be about politics or political issues, in line with our approach of not recommending certain types of content to those who do not wish to see it.

At the same time, we are preserving users' ability to find and interact with political content that is meaningful to them if that is what they are interested in on Facebook Feed. When ranking political content in Facebook Feed, our AI systems consider personalised signals, like survey responses, that help us understand what is informative, meaningful, or worth users' time. We also consider how likely people are to provide us with negative feedback on posts about political issues when they appear in Facebook Feed. We have shifted away from ranking political content in Facebook Feed based on engagement signals – such as how likely users are to comment on or share content – since we have found that they are not reliable indicators that the content is valuable to someone.

In addition, users can personalise what they see on Facebook through customisation tools we offer in their Feed Preferences and directly in places like their Feed. They provide direct feedback on a post by selecting Show more or Show less and use Reduce to adjust the degree to which we demote some content. If users don't want AI systems to personalise their Feed at all, they can use the Feeds tab, which will rank posts chronologically. They can also add people to their Favorites list so they always see content from their favorite accounts.

How does Meta assess the risk of its platforms to Australian users?

- **What research or analysis or monitoring of your platforms is undertaken to assess risk? Please provide frameworks to Committee.**
- **How does Meta assess the risk of filter bubbles on its platform? Please provide documents to Committee.**
- **How does Meta assess the threat of regulation in Australia and other jurisdictions around the world? Please provide the relevant documents detailing risk of regulation in Australia.**

All new products are subject to an internal risk assessment to identify and evaluate the potential risks associated with such products. In addition, we are required to perform risk assessments under the Phase 1 Online Safety Codes and Standards under the Australian Online Safety Act CTH 2021.

4. Written Questions on Notice from Mr Andrew Wallace MP (received 13 September 2024)

What types of information do companies selling and marketing harmful and addictive products on Meta, like alcohol and gambling, upload to the Meta platforms?

- **How is this data used by Meta?**
- **Is it used in any way by Meta algorithms/recommender systems in determining what content and/or advertisements people see?**

Meta has specific policies on alcohol advertising and online gambling and gaming.

Ads that promote or reference alcohol must comply with all applicable local laws, required or established industry codes, guidelines, licenses and approvals. Advertisers must follow all applicable laws, including targeting their ads in accordance with legal requirements. At a minimum, alcohol and gambling ads may not be targeted to people under 18 years of age.

For gambling, we only allow authorised gambling partners who either have a license for a specific country to run gambling ads on the platform targeted to that country. Ads are geo-gated and restricted to 18+ as well.

Australian research has shown that alcohol companies upload data about children to Meta platforms and that Meta platforms tag children with alcohol and gambling related advertising interests. Can Meta provide information about which alcohol companies have uploaded data about children under 18, including accounts which you suspect belong to children, to Meta platforms over the last 12 months?

- **Which information is included in this data?**
- **How many children have been targeted with this kind of behaviour?**

Alcohol ads are not allowed to be served to people under 18 on our services. Meta has a specific alcohol advertising policy, which explicitly states that advertisers may only run ads that promote or reference alcohol if those ads:

- Follow the targeting requirements of the location of the audience; and
- Do not target people under the age of 18.

For gambling, we only allow authorised gambling partners who either have a license for a specific country to run gambling ads on the platform targeted to that country. Ads are geo-gated and restricted to 18+ as well.

We welcome further evidence being provided on any ads that may be in violation of this policy in Australia so that we may have the opportunity to conduct further investigation.

Further, while we take measures to enforce these policies, age verification remains a challenge for many apps. This is why we support both greater parental controls for the use of our services by teens, and greater age control measures at the app store or operating system level so that age verification can operate seamlessly across the app ecosystem.

How much money has Meta made from alcohol and gambling advertising in the 2023-24 Financial Year, or 2023 Calendar Year – whichever Meta uses in its financial reporting?

Meta does not, in the ordinary course of business, separately track revenue by advertisement type for financial reporting purposes.

What protections does Meta have in place to ensure that people most at risk of harm from addictive products, like alcohol and gambling, such as people recovering from alcohol dependency or gambling addiction, are not targeted with marketing for these products when using Meta platforms? Does Meta access or use data, by any means, that might indicate a person is at risk of harm from addictive products like alcohol or gambling, including information about alcohol purchase frequency, frequency of gambling engagement and a person searching for help-seeking material or attending help-seeking practices related to alcohol or gambling?

- **Is any of this information accessed or used in Meta marketing algorithms/recommender systems and in which ways is this information accessed or used?**

Meta has specific policies on alcohol advertising and online gambling and gaming. Ads that promote or reference alcohol must comply with all applicable local laws, required or established industry codes, guidelines, licenses and approvals. Advertisers must follow all applicable laws, including targeting their ads in accordance with legal requirements. At a minimum, alcohol and gambling ads may not be targeted to people under 18 years of age.

For gambling, we only allow authorised gambling partners who either have a license for a specific country to run gambling ads on the platform targeted to that country. Ads are geo-gated and restricted to 18+ as well.

Advertisers can use our tools to exclude certain categories of people from their ad targeting, for example, using the BetStop list for this purpose.

5. Written Questions on Notice from Ms Zoe McKenzie MP (received 13 September 2024)

Age verification
In your opening statement on 4 September 2024, you mentioned that you *'require users to provide their date of birth when they register new accounts, a tool called an age screen. Those who enter their age (under 13) are not allowed to sign up.'*

You have highlighted that age is asked for at signup, and only if an age appropriate birth date is entered, will an account be opened (Antigone Davis, Meta witness hearing, 4 Sept 2024).

- **Is this the only age assurance method used at signup?**
- **Are there no other age assurance methodologies used at signup?**
- **If Meta can use Yodi to assure someone's age via a selfie when a teen 'tries to age up' as per your opening statement, why can't this technology be used for an initial setup of a new Meta account?**

You also reference Meta's use of AI age estimation tools

- **When are these tools used?**
- **What data informs an AI generated determination of someone's age?**
- **Is it biometric data?**
- **Do you gather any biometric data on any users, for any purposes, at any time?**
- **Please provide a list of data types you do not gather.**

At Meta, we take a continuous, multi-layered approach, recognising that no single method will work 100% of the time for every user. So rather than relying on a single-step process, we believe that it's more effective to build and invest in a suite of tools.

This includes:

- **Age collection at sign up:** When new users sign up, we request date of birth at account registration through an age-neutral screen with technical restrictions to make it harder for users to provide false information. For example, it is a neutral age collection screen (not pre-set to age 13 or today's date, the user must input their actual birth date) - this is an intentional design so that users must actively enter their date of birth. If multiple ages are entered the user will be blocked from registering.

- **Community reporting**: Anyone can report suspected underage accounts on Instagram and Facebook and in Oculus, and we have dedicated channels to review these reports
- **Training content reviewers**: Our content reviewers are also trained to flag reported accounts that appear to be used by people who are underage. If these people are unable to prove they meet our minimum age requirements, we delete their accounts.
- **Educating parents**: We remind parents of the minimum age in the Instagram Parents' Guide and our Parent Education Hub in VR and on IG.
- **Building AI to detect user age**: We invest in AI technology to detect likely teens and ensure they receive age-appropriate experiences for a variety of use cases e.g. restricting adults from sending messages to teens who do not follow them.
- **Age verification menu of options**: In some instances, users will be required to verify their age (e.g. if we have reason to believe they are misrepresenting their age, or if they attempt to change their age from under 18 to over 18 on Facebook and Instagram). When users need to verify their age, we currently provide them with two options to do so:
    - ID verification - we offer ID verification as a way for users to verify their age and accept various equivalent documents for Facebook and Instagram.
    - Face-based-age-prediction, offered through a third party Yoti - where users upload a video selfie of themselves to verify their age.

Today, we take a risk based and proportionate approach to age verification, and only require users to use Yoti or IDs to verify their age when we have signal that this is necessary. While we're pleased with how Yoti is operating in the use cases we have rolled out, we have only tested it to assure whether someone is above or below 18, not to assure an exact age.

Additionally, we have published a blog post with more information on how we use AI to better understand people's ages on our platforms.

- To develop our adult classifier, we first train an AI model on signals such as profile information, like when a person's account was created and interactions with other profiles and content. For example, people in the same age group tend to interact similarly with certain types of content.

From those signals, the model learns to make calculations about whether someone is an adult or a teen.

- To evaluate the performance of the model, we develop an “evaluation dataset.” That dataset is created by having teams manually review certain data points that we believe to be strong signals of age, such as birthday posts. Identifying details are removed before these posts are shared with the team to make a determination about the age of the person who posted it. Once the team has made that determination, they label the data with a note indicating whether the post was made by an adult or a teen. These labeled data points then make up our evaluation dataset.
- We then evaluate our classifier on a country-by-country basis. Before applying the classifier to a new country, we look at its performance across several criteria, including overall accuracy and accuracy across different groups of people. For example, since we use interactions with content as a signal, we look at how our model performs for people who have not been on our platform for very long and therefore have not yet interacted with much content. But the work is not done once the classifier is up and running. To check that our determinations are up-to-date, we regularly rerun the classifier to include the latest information.
- Each time we retrain the model, we check its age detections against the labeled evaluation dataset to measure the model’s accuracy. We have a sophisticated framework to ensure that our evaluation dataset is representative of the people using our services and that our model accuracy metrics are generalizable to the population of people using our services.

This technology forms the basis of important protections we have introduced to keep young people safe, for example, restricting adults over 18 from starting private chats with teens they are not connected to on Instagram and Messenger, and limiting the type and number of direct messages people can send to someone who does not follow them to one text-only message.

However, we believe there is a better way to implement legislation that will create simple, efficient ways for parents to oversee their teens’ online experiences, and that is an app store/OS approach to age verification.

Age assurance at the app store/OS-level is a simple approach that would:

- reduce the onus on parents to find and navigate a different age verification system on all the multiple apps and websites their children accessuse;
- minimise the number of places and times people have to share potentially sensitive data to verify age;
- allow parents to be more involved in the apps their children use from the time of download; and
- mitigate the risk of children moving from apps and websites with age assurance measures to less safe apps and websites that do not have such measures.

You suggest that the low effectiveness of existing age estimation technology means it is not a reliable technology. What effectiveness level would allow you to use age modelling and age estimation technology? What is impeding you from developing or procuring more effective technology? Is it correct that Meta believes that the way to assure age assurance on social media platforms is to get OS and app stores to police and enforce this? Would Meta still not be responsible for any Meta user activity outside the app store, for example users who access Meta products exclusively via the internet, not the app?

No age assurance technology currently available is 100% effective, and all have technical limitations. As the eSafety Commissioner noted in its Roadmap for Age Verification, many age assurance technologies are still in the early stages of development. As such, it is important to take a multi-layered approach, with age modelling and age estimation technology a valuable part of that, which signals to us whether someone may be misrepresenting their age. This allows us to either act on this signal, or ask the user to prove their age.

The technical limitations - as well as broader considerations around safety and privacy - mean it is important to consider at what layers of a teen's online experience age assurance measures should be placed.

We know that teens move interchangeably between many websites and apps. The average teenager uses dozens of applications on their phone - in some cases as many as 40 apps or more. Many of these apps have different standards or safety features, which are constantly changing or have new features added which can be challenging for parents and guardians to keep up. Only by creating industry-wide protections will teens actually be safer.

We believe there is a better way to implement legislation that will create simple, efficient ways for parents to oversee their teens' online experiences, and that is an app store/OS approach to age verification.

Age assurance at the app store/OS-level is a simple approach that would:

- reduce the onus on parents to find and navigate a different age verification system on all the multiple apps and websites their children access;
- minimise the number of places and times people have to share potentially sensitive data to verify age;
- allow parents to be more involved in the apps their children use from the time of download; and
- mitigate the risk of children moving from apps and websites with age assurance measures to less safe apps and websites that do not have such measures.

What *'signals on the platform'* (Antigone Davis, Meta witness hearing, 4 Sept 2024) do you use to determine someone's age?

- **When do you use these signals?**
- **Are these signals based off biometric data? Or usage data?**
- **Please provide a complete list of the signals used on your platform in Australia, and how/what data is collected and stored to develop these signals.**

We use AI technology to help determine whether someone is an adult (18 and over) or a teen (13–17). When people first sign up to use our services, we ask them to enter their birth date. But people aren't always accurate (or honest), and we've seen in practice that misrepresenting age is a common problem across the industry.

We train this technology with signals like profile information, when a person's account was created and interactions with other profiles and content. From those signals, we can begin to make calculations about the likelihood of whether someone is an adult or a teen, even if a teen has listed an adult birthday on their account.

To evaluate the performance of the model, we develop an "evaluation dataset." That dataset is created by having teams manually review certain data points that we believe to be strong signals of age, such as birthday posts. Identifying details are removed before these posts are shared with the team to make a determination about the age of the person who posted it. Once the team has made that determination, they label the data with a note indicating whether the post was made by an adult or a teen. These labeled data points then make up our evaluation dataset.

We then evaluate our classifier on a country-by-country basis. Before applying the classifier to a new country, we look at its performance across several criteria, including overall accuracy and accuracy across different groups of people. For example, since we use interactions with content as a signal, we look at how our model performs for people who have not been on our platform for very long and therefore have not yet interacted with much content. But the work is not done once the classifier is up and running. To check that our determinations are up-to-date, we regularly rerun the classifier to include the latest information.

Each time we retrain the model, we check its age detections against the labeled evaluation dataset to measure the model's accuracy. We have a sophisticated framework to ensure that our evaluation dataset is representative of the people using our services and that our model accuracy metrics are generalisable to the population of people using our services.

Publishing the full list of signals that we use to determine age may risk the integrity of the system and help children and adults to circumvent the safety measures we have implemented.

Recommender systems and algorithms
Meta asserted that 'There is some content on the platform that we may want to allow because it is particularly newsworthy or relevant' (Antigone Davis, Meta witness hearing, 4 Sept 2024).

- **How does Meta determine if content is newsworthy or relevant?**
- **How does Meta use algorithms to ensure this information is relevant in a user's feed?**

When making a newsworthy determination, we assess whether that content surfaces an imminent threat to public health or safety, or gives voice to perspectives currently being debated as part of a political process. We also consider other factors, such as:

- Country-specific circumstances (for example, whether there is an election underway, or the country is at war)
- The nature of the speech, including whether it relates to governance or politics
- The political structure of the country, including whether it has a free press

We provide more details about this in our Transparency Center.

Are algorithms designed to connect people to their friend's original authored content? Why does Meta not think that *'making algorithms opt in would be a positive change for our users'* (Meta responses to QONs, from 28 June hearing).

- **If consumers wanted the choice to opt in, would Meta provide this?**
- **In what ways does Meta benefit from not making algorithms opt in?**

The ranking algorithms are designed to connect people with the most relevant content that they will find interesting from friends, family, businesses and Groups. As the volume of content that is shared online has increased, without a ranking algorithm, we found that people were missing up to 70% of posts from their connections. So we developed and introduced a Feed that ranked posts based on what you care about most.

Recommendation algorithms are designed to show people content that they are likely to find interesting and valuable, from sources with which they are not otherwise connected.

If people want to switch to see the most recent posts, they can manage this in the Feeds tab.

Meta has asserted that *'Sensitive content control will help to filter [negative] content'* (Antigone Davis, Meta witness hearing, 4 Sept 2024), but the eSafety commissioner has said in her Mind the Gap research that *'almost two-thirds of young people aged 14–17 were exposed in the past year to negative content, such as content relating to drug taking, suicide or self-harm, or gory or violent material.'*

- **Do you accept that sensitive content controls currently used by Meta are not working?**
- **Do you think that Meta is a safe place for children?**

- **Do you think that Meta negatively impacts the mental health of its users?**
- **Do you accept that Meta can share inappropriate content with users?**

We are not familiar with the specifics of the services that are being referred to in the report cited. We provide an overview of our policies, tools, technology and resources to make our services a positive experience for people, including young people, in our Safety Center.

Meta asserted that *'we work with experts to ensure we are providing an age appropriate experience'* (Antigone Davis, Meta witness hearing, 4 Sept 2024).

- **What does Meta believe constitutes an age appropriate experience?**
- **b. Does this involve screentime recommendations per age group?**
    - **If so, why?**
- **Please advise the names of the experts you work with in Australia**
- **Do you think the viewing and posting of a livestreamed stabbing of a priest at the Wakely Christ the Good Sheperd Church is an age appropriate experience for those under 18 year olds who viewed it?**
- **Do you think the viewing and posting of a livestreamed stabbing of a priest at the Wakely Christ the Good Sheperd Church was appropriate content for your platform?**
- **Can you guarantee that every child on your platform has an age appropriate experience?**

When we refer to an "age appropriate experience," we mean creating environments and interactions on its platforms that are suitable for the developmental stage and maturity level of different age groups, particularly younger users. This involves implementing specific measures to ensure that the content, settings, and features are suitable for users' ages, particularly those between 13 to 18 years old. These measures include default protections and controls that allow young users to manage their online experiences effectively.

Meta's approach to providing age-appropriate experiences is informed by its "Best Interests of the Child" framework, which guides the design and operation of its services to ensure they are safe and beneficial for children and teenagers. This includes using technology to verify users' ages and adjust their experience accordingly to prevent access to inappropriate content. Additionally, Meta engages in ongoing consultations with young people and their parents to continuously improve the age-appropriateness of its platforms. We have a Global Safety Advisory Board and an Australian based Online Safety Advisory Board and work with many different partners such as the Australian Federal Police, Office of the eSafety Commissioner, Kids Helpline, Orygen, Reachout, PROJECT ROCKIT – among many others.

With respect to the livestream of the Walkley stabbing, this violated our policies and was removed promptly upon us becoming aware of it. We also took technical measures to proactively prevent the video being shared on our services.

Meta asserted that that *'we are working to ensure we aren't [sharing inappropriate content]'* (Antigone Davis, Meta witness hearing, 4 Sept 2024).

- **What specifically is this scope of work?**
- **Does this scope of work include technology based safety measures?**

We make significant investments in our ability to keep people safe. This includes investing in ongoing policy development, automated and human enforcement of our policies, awareness and educational initiatives, partnerships, as well as tools that allow people to customise their experience on our services, over and above the baseline safety and security efforts we deploy. We have around 40,000 people overall working on safety and security, and we have invested over US$20 billion (~AU$30 billion) on safety and security since 2016.

This investment includes building and maintaining our content governance and integrity systems, as well as user transparency tools and controls, and partnerships and programs through which we receive feedback and promote digital skills and literacy.

With respect to content governance, we use a strategy called "remove, reduce, inform" to manage content across Meta technologies. This means that we remove harmful content that goes against our policies, reduce the distribution of problematic content that doesn't violate our policies, and inform people with additional context so they can decide what to click, read or share. We also offer a range of tools so that people can customise their experience above and beyond the baseline investment we make in safety and security.

To help with this strategy, we have policies that describe what is and isn't allowed on our technologies. Our teams work together to develop our policies and enforce them. Increasingly, we have been deploying proactive detection technology to identify and action harmful content before anyone reports it to us. For many categories, our proactive rate (the percentage of content we took action on that we found before a user reported it to us), is more than 99 per cent across high-risk content types.

Beyond actioning harmful content, we also work to promote a safe and positive experience on our services by using technology and offering tools to help users customise their use of Meta's services. These features are informed by our consultations with industry, experts, and civil society organisations, including in Australia.

Meta has identified to the committee that content in a feed is ranked by positive or negative interaction from a user (Meta responses to QONs, from 28 June hearing) ***'if many people have interacted in a positive way with a post on Instagram or with similar content, the post will appear higher in a person's feed. Alternatively, if those interactions were negative ...the content is removed or ranked lower in the feed.'***

- **What qualifies as a positive experience?**
    - **Is it percentage of the piece of content viewed?**
    - **Is it shares?**
    - **Is it reactions?**
    - **Is it speed at which a piece of content is scrolled past?**
- **What qualifies as a negative experience?**
    - **Is it percentage of the piece of content viewed?**
    - **Is it shares?**
    - **Is it reactions?**
    - **Is it speed at which a piece of content is scrolled past?**

Our Systems Cards provide more insights in to what is considered to be a positive or negative interaction from a user with content by the signals used to predict whether a piece of content will be valuable to the end user. These signals might include who created the post and how you previously interacted with them, whether the post is a photo, a video or a link, or how many of your friends liked the post. A person can also hide a post, which helps to minimise similar content from appearing in your Feed.

Are system cards a complete insight into algorithms?

- **What data does a system card omit?**
- **What data does a system card include? (Please provide a comprehensive list)**
- **Are system cards available globally across all Meta products? What products and in which countries are system cards not in use?**
- **Were system cards implemented in response to the EU Digital Services Act?**
- **Have you done any market testing on system cards? What are people's responses? What cohorts have you used in these studies?**

System cards serve as valuable tools for promoting understanding, accountability, and ethical considerations in the development and use of AI systems. System cards provide an in-depth view into the complex world of AI systems, which are made up of many models and dynamic rules. Meta's system cards were written in a way that can be understood by experts and non-experts alike. We now provide 25 systems cards across a wide range of our products and surfaces.

There are several limitations that we outline in the MetaAI Blog to Systems Cards, specifically:

- AI systems constantly learn and evolve and so Systems Card are not definitive
- Technical information can be difficult to simplify, and the landscape changes in real time
- There is a need to consider unintended consequences
- Revealing the exact workings of certain AI systems could compromise the systems' security or open up a model to adversarial attacks, thus potentially harming the people who use our products

We developed the Systems Cards in consultation with experts and through pilot testing, as this blog post outlines and published a technical paper with the research underpinning them. They were launched as part of our compliance with the EU Digital Services Act but made global because they are part of our ongoing work on greater transparency.

Meta highlighted that users can be given a 'nudge' notification (Antigone Davis, Meta witness hearing, 4 Sept 2024) if they are viewing content on a specific topic for a certain amount of continuous time.

- **Is this correct?**
- **What qualifies a nudge to be given to a user?**
- **Does a nudge notification being used as part of the user experience imply addictive behaviours in the user experience?**
- **Why do nudge user notifications exist?**
- **Do nudge user notifications help manage screentime?**

The most updated information to share to allow the best response to this question is now with reference to Instagram Teen Accounts. Within these accounts, time limit reminders mean that teens will get notifications telling them to leave the app after 60 minutes each day. More details are here.

Scam ads

Does Meta gain revenue from the money paid to the platform by a user boosting or running an ad on Meta platforms? Does Meta gain revenue from the money paid to the platform to boost or run an ad on Meta platforms even if the ad is an identified scam ad? Does Meta, once identifying an ad is a scam ad, keep the revenue from the identified scam ad? You mentioned that Meta is not profiting from criminals making scam ads, due to the 'sizable' investment Meta has made to crack down on scam ads. Are you asserting that Meta does gain revenue from scam ads being hosted on Meta platforms, but that the amount Meta spends on crackdown activities means this revenue is mitigated by expenditure? How much in FY22-23 did Meta spend on anti-scam activities? What revenue did Meta make in 2022-23 off identified scam ads, on revenue generated by scam ads while they were boosted or run on Meta products?

Our business relies on providing safe and enjoyable experiences for our users. As fraud and scams degrade the experience for both users, business users, creator communities and advertisers, there is a strong business incentive to address them.

The damage and cost to our business far outweighs any ad spend, as well as the cost incurred of having to add teams to track and address these bad actors. We believe this kind of misleading content has a negative impact on people's experiences and the platform overall.

It is important to us at Meta that our services are positive for everyone who uses them. That is why Meta has around 40,000 people overall working on safety and security, and why we have invested over US$20 billion in safety and security since 2016, including US$5 billion in the last year alone.

Duty of care
What mental health experts and organisations in Australia have you partnered with?

We have a number of partnerships in Australia to ensure that our safety efforts are complemented and informed by local expertise.

We have a dedicated Australian Online Safety Advisory Group to consult and provide a local perspective on policy development. This group comprises safety and mental health experts from leading organisations such as Orygen, Butterfly Foundation, Kids Helpline, PROJECT ROCKIT, ReachOut, WESNET and CyberSafety Solutions, as well as many others.

In addition, we provide significant support to our safety partners to ensure that our users - especially young people - can connect and communicate safely.

Most recently, we have funded and supported the following online safety and mental health initiatives with local partners:

- *Butterfly Foundation:* In May 2024, we launched ‘Enter the Chat’, an educational campaign that brought together a group of Australian creators to discuss the impact that certain types of online content may have on body image, how to create content more consciously and what safety tools are available on Instagram to support body image and wellbeing.
- *PROJECT ROCKIT:* In November 2023, we partnered with youth-driven organisation PROJECT ROCKIT to create ‘Intimate Images Unwrapped’, a series of educational videos that aimed to build greater literacy and awareness around the dynamics of sharing of intimate images. Additionally, we have supported PROJECT ROCKIT for over a decade to deliver the Digital Ambassadors Program, s a youth-led, peer-based anti-bullying initiative. This program aims to utilise strategies to safely connect and tackle online hate, directly empowering more than 25,000 young Australians to tackle cyberbullying.
- *Kids Helpline and ACCCE:* In November 2023, we partnered with the Australian Federal Police-led Australian Centre to Counter Child Exploitation, Kids Helpline and US-based organisation NoFiltr (Thorn) to

inform young people about sextortion. The campaign included educational resources encouraging preventative behaviours online, the signs to look out for, where to report and where to seek support.

- *ReachOut:* In 2023, we partnered with youth mental health service, ReachOut, to launch a creator-led campaign aimed at fostering social and emotional wellbeing in the lead-up to, and following, the Voice to Parliament referendum. The campaign focused on supporting and empowering young First Nations people in navigating the complex social and emotional wellbeing challenges resulting from the referendum and its surrounding debate.

What mental health and health organisations in Australia have you donated to, and how much per transaction in

- **FY23-24**
- **FY22-23**
- **FY22-21**

We don't have these details to share. We partner with organisations by funding specific initiatives and campaigns, by supplying ad credits and by supporting charitable work through dedicating employee time.

Meta stated that (Meta responses to QONs, from 28 June hearing) '*The phrase "duty of care" is vague and undefined such that we do not think agreeing to it would be helpful to users or the industry. Instead of a "duty of care," we support clearly defined standards that would apply equally to all social media platforms.'*

- **Would a duty of care obligation not support Meta's recommender system algorithms or ad algorithms as they currently exist?**
- **If a duty of care was defined clearly through standards, would Meta find these helpful?**

We support relevant, proportionate and risk based systemic requirements on platforms such as Meta to ensure the safety and integrity of our products.

Screen time

Meta provides services that allow parents to set time limits (Antigone Davis, Meta witness hearing, 4 Sept 2024)

- **What is the purpose of a screen time limit?**

- **Does the existence of a screen time limit mean that children on your platforms are having issues managing a social and appropriate screentime while on your app?**
- **How would parents manage their children's screen time after a Meta account has been set up?**

Meta states that (Meta responses to QONs, from 28 June hearing) *'we understand that screen time can be a concern in relation to all media apps, including social media, video game, and streaming platforms.'* What does Meta understand is concerning about screen time?

We provide time limit reminders and a range of parent supervision controls in the new Instagram Teen Accounts that was announced last week. More details are here.

The Facebook Files, and the information whistleblower Frances Haugan leaked as a part of this release asserted that 70% of all inappropriate adult-minor contact came from the 'people you may know' function. When asked a question on this Meta's response was *'People rely on people you may know... to be connected to be people in their connections...that said, we take discoverability of teens very seriously, we've put safeguards in place'* (Antigone Davis, Meta witness hearing, 4 Sept 2024).

- **Does the statistic of 70% of all inappropriate adult-minor contact not raise safety alarms for Meta?**
- **Why does this function still exist?**
- **Does Meta believe that this identified problem has been completely resolved?**

We work hard to prevent potentially unwanted or unsafe interactions to keep teens safe. In addition to removing accounts that violate our Child Sexual Exploitation, Abuse and Nudity policies, our reviewers and automated systems consider a broad spectrum of signals to prevent potentially unwanted or unsafe interactions. As outlined in the Transparency Centre, we may restrict access to products and features for adults based on their interactions with other accounts, searching for or interacting with violating content, or membership in Groups we remove for violating our policies. These restrictions apply to People You May Know (PYMK), where, based on these signals, we will restrict these adults from discovering teens in PYMK, and also prevent teens from discovering these adults in their PYMK.

Preventing Potentially Unwanted or Unsafe Interactions

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. More details about the recently announced Instagram Teen Accounts is available here.

When our systems detect an adult may be engaging in unsafe interactions:

- We don't show young people's accounts in Explore, Reels, 'People You May Know' or 'Accounts Suggested For You'.
- We limit the ability to search for teen accounts.
- We restrict adults from identifying teen accounts through follower lists.

User behaviour
How many users under 18 does existing age estimation technology suggest are on your platform in Australia? How many users under 16 does existing age estimation technology suggest are on the platform in Australia?

We are currently considering our response to the eSafety Commissioner's request for information under the Basic Online Safety Expectations (BOSE), which includes similar questions to these. However, at this time, we can confirm that, based on self-reported ages of our Australian monthly active users, less than 10% of Instagram accounts belong to teens under 18, and less than 5% of Facebook accounts belong to teens under 18.

How many users under 16 does existing age estimation technology suggest are on the platform in Australia (on average across FY23-22, and/or in this current week)

- **After 10pm?**
- **After 1am?**

Meta does not, in the ordinary course of business, separately report on the time of day during which users are using our apps.

Does Meta believe that children under the age of 18 using Meta instead of getting their recommended hours of sleep per night is problematic use?

- **Is this the intended use of Meta products?**

We do not and cannot measure whether an individual user is experiencing problematic use. Assessing problematic use is an individualised determination that depends on multiple factors, including how an individual spends time on our app, what activities an individual is not engaging in while they're on our app, and how the individual's use is impacting their everyday life and relationships. Without insights into all of those factors, it is not possible to determine whether a user is experiencing problematic use. Basing evaluations of problematic use solely on the time at which an app is used is arbitrary and would miss the fuller picture.

Meta states that *'Assessing problematic use is an individualised determination that depends on multiple factors, including how an individual spends time on our apps, what activities an individual is not engaging in while they're on our apps, and how the individual's use is impacting their everyday life and relationships'* (Meta responses to QONs, from 28 June hearing).

- **What would Meta qualify as 'problematic use'?**

It is not possible to provide a universal definition of 'problematic use'. As stated in our previous response, what constitutes 'problematic use' will differ across individuals.

Extremism and radicalisation
On ABC's Insiders on 11th August 2024, ASIO Director General Mike Burgess said *'social media does make our job harder'*, and highlighted that social media and the internet were an *'incubator of violent extremism'*. He cited *'the algorithms companies use to direct content'* as one of the vectors of this.

- **Do you agree with these statements?**

Burgess also stated that 'a youth only has to search for something once, and then in their search feed, they get plenty of violent extremism or extreme material, which is unhelpful and hurtful to their young forming brains'

- **Do you agree with this statement?**

We are not familiar with which social media platforms were being referred to by this statement but this is not a fair assessment with respect to those services provided by Meta.

- **Why does extremist material proliferate on recommender systems?**

We are not familiar with which social media platforms were being referred to by this question but this is not an accurate assessment with respect to those services provided by Meta.

We do not allow organizations or individuals that proclaim a violent mission, or are engaging in violence, to have a presence on Facebook and Instagram. We do not allow content that praises, supports or represents individuals or groups engaging in terrorist activity or organized hate.

From April to June 2024, we removed 7.5 million pieces of content for violating our policies that prohibit terrorist content, 99.2% of which we removed proactively. In the same period, we removed 7.4 million pieces of content for violating our policies that prohibit violence and incitement, 9.8% of which we removed proactively.

- **How does this content develop such momentum via algorithms?**
    - **Is it because algorithms value interaction?**
    - **Do algorithms consider the type of interaction someone may have with their content? For example, if a piece of content is a clip of someone saying controversial religiously motivated extremism, and this clip doesn't breach community safety guidelines, will an algorithm judge someone watching this whole video and sharing it as a 'positive' experience, which has helped someone connect to a community of like minded people?**

This is not an accurate characterisation of how our services operate. From April to June 2024, we removed 7.5 million pieces of content for violating our policies that prohibit terrorist content, 99.2% of which we removed proactively. In the same period, we removed 7.4 million pieces of content for violating our policies that prohibit violence and incitement, 9.8% of which we removed proactively. We also take steps to demote content that is likely to violate our policies but has not yet been confirmed to do so.

The AFP's submission to this inquiry states *'The saturation of such extremist content can create an "echo chamber" lacking alternative content, and feeds into algorithmic-based preferencing that seeks to increase viewership and engagement by offering viewers further, related content. This is resulting in young people unwittingly self-radicalising themselves through increased exposure to content. This persuasive technology can further exploit an individual's desire to connect and engage with extremist groups and highlights the impacts of allowing harmful material to stay online.'*

- **Does Meta agree with the AFP's assertion?**

- **Does Meta think its community guidelines are working, if algorithms have become such prolific tools of violence that the APF is now referencing them as a significant problem in our security landscape?**
- **Does Meta agree that algorithms are increasingly proliferating extremist content?**
- **Does polarising content rank well on Meta's recommender algorithms?**

Does Meta agree with the AFP that there is '*The widespread availability of extremist material…on open social media platforms'* (AFP submission to this inquiry)?

The AFP's submission to this inquiry states *'Social media provides a platform for violent and extremist material to be viewed, shared, and promoted. JCTT investigations often identify subjects engaging in extremist ideological dialogue and viewing and sharing abhorrent and violent extremist material including beheading videos and other violent content linked to extremist ideologies.'*

- **Does Meta agree with the AFP's assertions?**
- **Does Meta believe the APF's assertions may relate to their platforms?**

We are not familiar with which social media platforms were being referred to by this statement but this is not a fair assessment with respect to those services provided by Meta.

From April to June 2024, we removed 7.5 million pieces of content for violating our policies that prohibit terrorist content, 99.2% of which we removed proactively. In the same period, we removed 7.4 million pieces of content for violating our policies that prohibit violence and incitement, 9.8% of which we removed proactively. We also take steps to demote content that is likely to violate our policies but has not yet been confirmed to do so.

Does Meta believe that, as asserted by the AFP in their submission to this inquiry, that *'the removal of harmful extremist material from online platforms is a challenging and time consuming process'?*

We are not familiar with which social media platforms were being referred to by this statement but this is not a fair assessment with respect to those services provided by Meta.

The AFP submission to this inquiry states that '*[Child sexual abuse] Offenders seek to engage with children on popular platforms including Facebook, WhatsApp, and Skype, and further obfuscate their offending through virtual private networks and encrypted technology.'*

- **Does Meta agree with this assertion?**
- **Does Meta believe that its platforms are a safe place for children**

When Australians are using Meta's family of apps, we recognise that we have a responsibility to keep people safe, to comply with all applicable laws, to be responsive to community concerns, and to promote accountability and transparency.

Keeping people safe online has been a challenge since the start of the internet. As threats and trends constantly evolve, it's important that we continue to adapt so that people have safe and positive experiences across Meta's services. We are focused on building technology that people find useful and feel safe when doing so.

At Meta, child protection is always a top priority. We use a combination of technology and behaviour signals to detect and prevent child sexual abuse material, including grooming or potentially inappropriate interactions between a minor and an adult.

In our submission, we outlined our approach to combatting child sexual abuse material on our services.

With respect to encrypted services, to protect users' personal information and enable users to share personal information securely and privately, we apply end-to-end encryption to WhatsApp and Messenger personal messages and calls.

We recognise that there is general agreement across industry, civil society and within Government about the value of encryption to promote privacy, safety, and security. While there are concerns that have been raised about the ability to promote safety on encrypted services, for Meta, the values of safety, privacy, and security are mutually reinforcing. An independent Human Rights Impact Assessment of Meta's expansion of end-to-end encryption - conducted by NGO Business for Social Responsibility in line with UN Guiding Principles on Business and Human Rights - found, among other areas, that encryption increased the realisation of privacy, freedom of expression, protection against cybercrime threats, physical safety, freedom of belief and religious practices and freedom from state-sponsored surveillance and espionage.

In line with these findings, we continue to invest in behavioural analysis and metadata as effective harm prevention rather than undermine encryption.

Additionally, we provide a range of features to empower users to keep themselves safe. For WhatsApp, that includes:

- **Unknown senders**: the first option users are given when someone who is not a contact messages them is whether they would like to block or report them.
- **Block and report:** we advise users to block and report suspicious messages, turn on two step verification for extra security and never click on links or share personal details with someone they do not know. When users choose to report a message, group or other user, that content is reported to WhatsApp for review. Reporting the content means it can be seen by our trust and safety team, who can then pass it onto law enforcement if it is illegal.
- **Privacy settings:** users can adjust their privacy settings to control who sees their information, including their "last seen" and "online", profile photo, about, or status to determine who can see their profile photo, about, or status and who can add them in groups.
- **Mute:** we give users options to mute notifications and archive chats to avoid unwanted interactions. Users can also silence calls from unknown callers.

We encourage users to think carefully before sharing something with their WhatsApp contacts. When a chat, photo, video, file or voice message is shared with someone else on WhatsApp, they will have a copy of these messages and can forward or share with others if they choose to.

As part of our roll out of end-to-end encryption on Messenger (announced late last year), we introduced new privacy, safety and control features. This includes delivery controls that let people choose who can message them and 'app lock', which uses a device's privacy settings like fingerprint or face authentication to unlock the Messenger app. These supplement existing safety features like report, block and message requests.

We work closely with outside experts, academics, advocates and governments to identify risks and build mitigations to ensure that privacy and safety go hand-in-hand.