04: KEY ISSUES BRIEF: Unlawful Material Codes and Standards

Talking Points

- eSafety established an Enforcement Taskforce in June 2025 to focus on investigating and enforcing non-compliance with systemic regulatory schemes under the Online Safety Act, including the Unlawful Material Codes and Standards.
- eSafety issued and published formal warnings to OmeTV (a 'chatroulette' style app) and a provider of some of the world's most visited AI-generated nude image ('nudify') websites resulting in suspension of these services in Australia and creating a safer online environment in Australia.
- Following engagement by eSafety with Apple and Google regarding their obligations under the App Distribution Service Code (App Code), both removed OmeTV from their app stores.
- eSafety regulatory engagement has also resulted in several online services committing to uplift their safety practices, with Roblox committing to put in place measures to prevent adults contacting children and Hugging Face amending their terms to recognise that users uploading powerful models must take steps to prevent child sexual exploitation and abuse, and pro-terror material.
- In January 2026 eSafety commenced an investigation into xAI's compliance with the Designated Internet Services Standard in response to reports that its AI chatbot Grok may have been used to create child sexual exploitation material. This is in addition to an investigation commenced in 2025 in relation to the X platform.

Key Issues

- The Unlawful Material Codes and Standards seek to prevent and minimise harms from the worst material online. The requirements are limited to Class 1 material, which is defined in the Act as material that is, or is likely to be, classified as refused classification under the National Classification Scheme.
- Class 1A material is defined in the codes and standards as the most harmful Class 1 material such as child sexual exploitation material, pro-terror material and material dealing in 'extreme' crime and violence.
- Class 1B material is defined as material that is still harmful, but less so than 1A, including crime and violence and drug-related material.
- The industry codes came into effect between September 2023 and March 2024, and the industry standards came into effect in December 2024. An 'enforcement grace period' ended in June 2025.

Contact: s 22
Phone: s 22

Cleared by: s 22, A/g Executive Manager
Date: 19/01/2026

Enforcement Taskforce

- eSafety is prioritising four types of services for assessment and enforcement where required.

    1. Services with high risk and high prevalence of child sexual exploitation material and/or pro-terror content
    2. Services enabling access to children, resulting in self-generated imagery, grooming and sexual extortion
    3. New generative AI harms from 'nudify' services and AI models that can create child sexual exploitation material
    4. 'Gatekeeper' services, such as app stores, which enable access to a wide range of services/apps, including those that may be unresponsive to eSafety outreach.
- While these are eSafety's core priorities, eSafety will take broader action where the risk, evidence and need for deterrence supports it.
- eSafety has issued **2 formal warnings** for breaches of Industry Standards and has a number of active investigations.

Ome.TV app

- Ome.TV is a randomised video chat app that was available on the Apple App Store and Google Play store.
- On 20 August 2025, eSafety issued a formal warning to OmeTV's parent company, Bad Kitty's Dad LDA, which is based in Portugal.
- The provider failed to comply with the requirements under the Relevant Electronic Services Standard to have appropriate safety features and settings, including measures to prevent adults from contacting children.
- The Commissioner wrote to Apple and Google to highlight that, because we found that OmeTV has breached Australian law, they are required to take appropriate action under the App Code.
- As a result of this engagement, Apple and Google removed the OmeTV app. The provider of OmeTV has since suspended all access, including to their website, from Australia.
- We continue to engage with Apple and Google in their capacity as providers of app distribution services. Where they fail to comply with the App Code, eSafety can direct them to comply, with civil penalties for a failure to comply with such a direction.

Nudify services

- eSafety has identified several highly accessible services that openly market themselves as 'nudify' platforms.
- On 2 September 2025, eSafety issued a formal warning to a provider of two of these services. Together, these services receive around 100,000 monthly views from Australian users.
- The services have been found to lack appropriate safeguards to prevent the creation of child sexual exploitation material, in breach of the Designated Internet Services Standard. Indeed, these services have the ability to create 'young' as well as 'adult' imagery and promote their ability to nudify 'any girl'.
- To avoid unintentionally alerting potential offenders, the names of the services have been withheld.
- The provider subsequently suspended certain services in Australia.
- We continue to assess the provider's compliance and will take further action if appropriate.

X and Grok

- eSafety commenced two separate investigations into the providers of X and Grok – X Corp. and X.AI LLC respectively.
- In January 2026, eSafety commenced an investigation into the Grok AI service after becoming aware of reports that it may have been used to create child sexual exploitation material. This

includes as a result of reporting from experts such as those at the Internet Watch Foundation, a UK based child safety organisation.
- eSafety has an existing investigation into X due to ongoing concerns about the prevalence of child sexual exploitation material on the platform. Reports about Grok generating child sexual exploitation material on X heightened these concerns.
- eSafety has sought information from both X Corp. and X.AI LLC as part of these investigations.
- The providers of both services (X and Grok) are required to have appropriate systems, processes and technologies in place to detect, remove, disrupt and deter their services from being used to disseminate unlawful material, regardless of whether it is AI-generated.
- While eSafety can seek the removal of specific image-based abuse of adults using other regulatory powers under the Online Safety Act, the Unlawful Material codes and standards only cover generation and distribution of child sexual exploitation material, so cannot be used to address systemic issues related to image-based abuse involving adults. Further, Age Restricted material codes only prevent children's access to explicit adult images which may amount to image-based abuse, rather than the generation of such material.
- Beyond this, it is not appropriate to comment on the conduct of an ongoing investigation.

<u>Compliance and Enforcement Outcomes</u>

- Through eSafety's compliance and enforcement activities to date, a number of services have made changes or committed to comply with the industry codes and standards:
    - **Roblox** committed to putting in place parental controls for users and default privacy settings for users who are under 16, supported by facial age estimation for all users. These changes will mean adults will not be able to contact children via chat features. We are monitoring implementation and will use our enforcement powers where required.
    - **Hugging Face** has changed its terms of service so now all account holders are required to take appropriate steps to minimise the risks of models being used to generate CSEM or pro-terror material. Hugging Face is required to enforce these terms where there are breaches.
    - Some major services introduced or improved their user reporting tools [If asked which companies – **Google** introduced user reporting in Gmail and Messages; **WhatsApp** introduced harm-specific user reporting].

Background / Difficult Questions

How does your work align with the Government's commitment to ban nudify apps?

- We welcome the Government's announcement. We will continue working with the government to ensure any legislation effectively addresses the challenges, which occur both on websites as well as apps.
- We will continue to use the powers currently available to us, as these harms are occurring and growing.

How many investigations do you have underway?

- As of 20 January 2026, we have 7 investigations under section 42 of the Act into industry codes and standards compliance. We have a wider programme of compliance assessments and engagement, that fall short of formal investigations.
- [If pushed on which companies] It would not be appropriate to comment on ongoing investigations.

What can you actually do about Grok and the alleged nudified images of children that were posted on X?

- If any service is found to be in breach of its regulatory obligations, eSafety has a range of enforcement options available under the Online Safety Act including formal warnings, infringement notices, enforceable undertakings, injunctions and seeking court-ordered civil penalties.
- eSafety does not have the power to ban a service.
- The Online Safety Act provides that eSafety can apply to the Federal Court for an order to prohibit an service from being available to Australians if the service fails to comply on two occasions with a civil penalty provision (such as a breach of a code or standard), *and* poses a significant community safety risk.

What obligations are there on services in relation to violent material such as the murder of Charlie Kirk?

- Where material is refused classification, or is likely to be considered RC, it would fall under the Unlawful Material codes and standards.
- The specific obligations on services depend on whether the violent material is 'extreme', and therefore considered Class 1A material under the codes and standards.
- Where material is 'extreme' and Class 1A, services have obligations both to remove the material, but also to take action to prevent further breaches, ensuring systemic action is taken.
- Material depicting the murder of Charlie Kirk and Irena Zarutska was re-classified R18+ by the Classification Review Board, on appeal by X Corp.
- The presence of that particular material on the service, therefore, would require services to take action in compliance with the Age-Restricted Material Codes upon their commencement.
- Relevant obligations for social media services which permit high-impact violence material include requirements to implement appropriate safety tools to limit end-user access or exposure to this material (including filtering, blocking or blurring this material). These safety tools must be set by default to an appropriate setting for Australian child end-users in relation to high-impact violence material.

What do you do with the complaints you receive?

- Complaints made about potential non-compliance with an industry code or standard can provide us with critical information for an investigation or alert us to systemic issues with a service.
- However, codes and standards are narrowly focussed on Class 1A and 1B material and systemic safety issues. We cannot resolve individual issues, or out of scope material that we receive complaints about such as scams. To the extent that eSafety receives complaints about specific class 1 material available on a service, these may be investigated under the Online Content Scheme in accordance with the Online Safety Act.

3: KEY ISSUES BRIEF: Complaints Schemes

Talking Points

Bondi Beach Shooting

- eSafety's on-call capability was activated immediately following the Bondi Beach attack to monitor footage of the attack circulating online.
- eSafety also alerted platforms to ensure they were proactively responding to content on their platforms.
- Initial material depicting the shooting was classified as MA15+ by the Classification Board.
- Further material was subsequently classified as RC.
- **106 complaints** have been received relating to footage of the Bondi Beach shooting.
- eSafety sought information from six providers (Meta, TikTok, Snap, YouTube, X Corp and Reddit) about their response to the attack and content that may be available on their services. All indicated they were removing content in accordance with their terms of service and had safety tools in place for under 18s for graphic content.

Antisemitism

- **3 Adult Cyber Abuse complaints** involving antisemitism have been received this financial year.
- Material must target a specific Australian adult to be classified as adult cyber abuse. Many antisemitic complaints are directed at groups or communities rather than an individual and therefore fall outside this threshold.
- Some complaints also do not meet the threshold because the serious harm element cannot be established, where the material reflects strong disagreement or controversial views rather than a deliberate attempt to cause serious harm to an individual

Impact of SMAR on Child Cyberbullying Complaints

- While less children are on social media, children are still online, including on messaging apps. The trend of increasing CB complaints continues, notwithstanding the SMMA.
- We have seen an increase in complaints about threats of violence, doxing and offensive/upsetting photos/videos but a decrease in complaints about nasty comments/name calling, unwanted contact, fake accounts/impersonation and meme pages
- Comparing complaint numbers received for the period 1 December 2024 - 22 January 2025 and 1 December 2025 - 22 January 2026 shows **a 15% increase**.

Key Issues by Complaint Scheme

Adult Cyber Abuse

- Adult Cyber Abuse complaints have increased 65% from the previous year to 3,073 ytd. Just 2% of these complaints are assessed as meeting the required legislative threshold.
- The most common type of complaints that don't meet the threshold are about material that is merely defamatory, which accounted for **1,147 complaints** in Jul-Dec 2025/26, up **131%** for the same period in the previous year.

Online Hate

- Between 1 July 2025 and 22 January 2026, eSafety received 19 complaints involving online hate, none of which met the threshold for adult cyber abuse under section 7 of the Online Safety Act.
- Although online hate is not defined by the OSA, through our complaint schemes eSafety has observed complaints regarding online hate referring to content that attacks or discriminates against individuals or groups based on characteristics like race, religion, gender, or disability.
- When receiving these types of complaints an assessment is made on whether the material meets the definition of 'serious harm', being serious physical harm or serious harm to a person's mental health, whether temporary or permanent.
- For eSafety to act under the Adult Cyber Abuse or Child Cyberbullying schemes, the online hate complained of must target a specific Australian adult or child—not a broad group or organisation. However, if content promotes or depicts violence against a group, eSafety may act under the Online Content Scheme.

- Specific **antisemitic complaints** received from July 2025 to present:
  - **Complaint** regarding numerous posts on X.com posted by an individual with a significant online footprint and a strong political position regarding the Israel/Gaza situation. This end-user targeted individuals with Jewish identities or those who publicly express pro-Israel views Assessment of the posts were that they would be regarded as offensive or malicious, however the serious harm element could not be established. Section 7 Definition not met.
  - **Complaint** where they were subjected to 'explicit antisemitic' abuse on social media. Material assessed as disagreeing views regarding Venezuela and the complainant was called a "zio-freak" however the serious harm element could not be established. Section 7 Definition not met.
  - **Complaint** where the complainant has been branded a liar and included screenshots of the complainants' comments which was posted to a pro-Palestinian Instagram account. The reported material is an Instagram 'Highlights' labelling certain accounts 'Zionists" No evidence of intention to cause serious harm. Section 7 Definition not met.

Collective Shout

- In April 2025 the online advocacy group Collective Shout launched an open letter campaign demanding credit card companies and PayPal block payments for games on Steam and Itch.io after identifying games they claimed had themes of rape, incest, sexual violence and or child abuse.
- As a result of the campaign, employees of Collective Shout have been the target of online harassment and abuse. eSafety has received 47 complaints under the adult cyber abuse and image-based abuse schemes from employees of Collective Shout:
  - **36 ACA** complaints - 9 met the threshold for adult cyber abuse. For 3 of these, material was no longer available online, so there was no action for eSafety to take. For the remaining 6, section 93(1) service provider notifications were sent to platforms resulting in the removal of material by the platforms.
  - **11 IBA** complaints - 3 led to removal of intimate images that were posted online, 3 did not meet the threshold of the IBA scheme, 4 related to the same reported material that was no longer accessible online, and one was a duplicate report. A further complaint was received on the 23 January 2026 however, the material had been removed prior to eSafety involvement. Due to the recurring nature of these complaints enquiries are being made to attempt to identify the end user, including with the JC3P who are conducting a concurrent criminal investigation.

Image-Based Abuse

- eSafety received 3,843 image-based abuse reports during the reporting period, a 34% increase compared to the same period in the last financial year.
- Sexual extortion was the highest category of harm with a 26% increase in reports compared to the same period in the last financial year.
- eSafety received 55 reports involving digitally altered intimate images during the reporting period, a 120% increase compared to the same period in the last financial year (25). This type of content is commonly created with the use of nudifying apps or more recently AI tools.
- 567 complaint notification were given, 75% of which resulted in removal, and 8 service provider notifications were given, 100% of which resulted in removal. *Note: there is no statutory requirement for content to be removed pursuant to a complaint alerts.*
- Few removal notices are issued under the scheme as content is often removed following a complaint notification, additionally over half of reports involve threats to share intimate images so there is no content to remove. Complaint notifications are an efficient and effective mechanism for facilitating rapid removal.
- Pornography sites, especially hack/leak/expose sites, account for most unsuccessful removal attempts due to lack of contact points and hidden hosting. Automated scraping and bot-run sites also pose major challenges, often lacking human moderation. Other difficult platforms include uncooperative message boards, archive sites, file sharing services, and overseas-hosted platforms. eSafety continues to work with international partners to pursue outcomes where possible.

Illegal and Restricted Content

- Complaints to the Illegal and Restricted Content (IRC) team have an **increase of 71%** in URLs reported compared to the previous year.
    - eSafety has received a total of **10,024 complaints** for the first half of the current FY (Jul 2025 – Dec 2026) about material relating to CSAM. This is an **increase of 33%** compared to the same time previous FY.
    - eSafety has received a total of **706 complaints** for first half of the current FY (Jul 2025 – Dec 2026) about material related to TVEC. This is an **increase of 394%** compared to the previous FY.
    - eSafety has received a total of **3,918 complaints** for the first half of the current FY (Jul 2025 – Dec 2026) about material relating to promotion, incitement or instruction in crime, violence, sexually explicit adult material, extreme or offensive content. This is an **increase of 102% compared** to the previous FY.
- Formal action by way of removal notices has also seen a significant increase in the 2025-26 FY with **36 removal notices** issued - an **increase of 350%** compared to the previous FY.
    - 1 notice relating to CSAM material
    - 35 notices relating to TVEC and Crime and Violence
    - 7 Revocation notices issued under section 113.

Bondi Beach Shooting

- Immediately following initial news reporting of the Bondi Beach shooting, eSafety monitored for any footage of the attack circulating online to assess whether the content would meet the threshold for removal, which is Class 1 material or 'Refused Classification' as defined under the National Classification Scheme.
- A total of 106 complaints have been received, mostly concerning uploaded bystander footage.
- Multiple videos of the attack and aftermath were quickly identified across platforms.
- eSafety also engaged with platforms to ensure they were being proactive in their own response to content on their platforms, including by removing violent content under their own terms of service and placing interstitials on gratuitous violence to prevent inadvertent or accidental viewing of the content.
- On 15 December, eSafety wrote to Meta, TikTok, Snap, YouTube, X Corp and Reddit requesting information on their response to the attack and content that may be available on their services. All providers responded to eSafety's information request and indicated they were

removing content in accordance with their terms of service and had safety tools in place for under 18s for graphic content.
- Initial footage was assessed and referred to the Classification Board, which classified the material as MA15+.
- A later complaint contained new footage was classified by the Board as Refused Classification (RC).
- Investigations confirmed that the RC material had already been removed by Meta, so no removal notice was required.
- Additional materially similar RC content was identified on a fringe website named Watch People Die (WPD) and X; removal notices were issued. X geo-blocked the content however the content is still available on WPD. eSafety is considering what further action may be appropriate.
- eSafety has recently been alerted by an overseas regulator to 'gamified' versions of the incident circulating online overseas, the nature and origin of these are not yet clear, however no content has yet been identified as available in Australia.

Charlie Kirk shooting and Zarutska Stabbing

Following the fatal shooting of Charlie Kirk at Utah Valley University on 10 September 2025, eSafety received 47 complaints about extremely graphic footage of his death circulating online. Following the fatal stabbing of Iryna Zarutska, eSafety received 13 complaints about online material depicting her death.

- The material was referred to the Classification Board and initially classified as Refused Classification (RC).
- Based on the RC classification, eSafety issued removal notices to Meta and X under the Act.
- X applied to the Classification Review Board for the classification to be reconsidered, and the Review Board reclassified the footage as R18+.
- The Review Board determined that the footage did not contain a "very high degree of impact," noting the depiction lacked detail.
- The Review Board found the violence was not excessively frequent or prolonged, as it was shown in a single, unedited shot.
- The Board observed that no techniques were used to heighten impact, such as close-ups or slow motion, and that neither the shooter nor the weapon were shown.
- The Review Board also noted the event was widely reported and part of public discourse.
- The Board advised that more detailed or edited versions of the same event could still be found to be RC.
- The Review Board required consumer advice warning: "High impact violence, blood and injury detail, distressing scenes."
- Following the reclassification to R18+, eSafety issued revocation notices to Meta and X.

Child Cyberbullying

- eSafety has recorded a total increase of **15%** in complaints, however only **18%** of these met the legislative threshold for child cyberbullying
- Although too early to determine if related to the SMMA, analysis of complaints for 1 December 2024 - 22 January 2025 compared to 1 December 2025 - 22 January 2026 has indicated:
    - Decreases in complaints relating to:
        - Nasty comments and name calling - 11% decrease
        - Unwanted contact - 15% decrease
        - Fake account and impersonations - 100% decrease
        - Meme page - 100% decrease
    - Increases in complaints relating to:
        - Offensive/upsetting pictures and or videos - 14% increase
        - Threats of violence - 0 complaints to 18 complaints
        - Doxing - 0 complaints to 16 complaints
        - All other categories - 6% increase

- Comparing complaints received from 1 December 2024 – 22 January 2025 with those from 1 December 2025 – 22 January 2026, reports from children under 16 have remained stable or increased across most age groups, with exceptions noted at ages 1, 8, and 15.

    - Complaints decreased in **3 age groups** (ages 1, 8, 15).
    - **3 age groups** recorded no change (ages 3, 7, 11).
    - The largest increases were for **ages 6 and 9 (+300%)**.
    - The largest decrease was at **age 1 (–100%)**.

Complaints and enquiries about Grok AI

- Between 2 and 6 January 2026, 3 image-based abuse complaints were received by eSafety in relation to the creation/posting of images by Grok AI.
- The images the subject of two of the complaints were not accessible at the time an eSafety member assessed the complaints, however, screenshots were provided with the complaints that showed the prompts used.
- The image the subject of the remaining complaint was accessible and was an intimate image; however, the account involved in the creation of the image had been suspended.
- eSafety alerted X to the image that was accessible.
- The image was subsequently removed and was no longer accessible.
- eSafety has received multiple media enquiries about Grok.
- Between 26 November 2025 and 15 January 2026, the Illegal and Restricted Content (IRC) Team received 18 complaints in relation to Grok.
- Where URLs were provided by the complainant, material was assessed; however, they did not meet a Class 1 threshold, specifically the material was not deemed to be CSAM or other types of illegal content such as terrorism. IRC has not observed any images created by Grok that amount to Class 1 material.

Other Issues

Reliance on the Classification Board to classify material

- **Q: How does eSafety determine whether material is class 1 under the Online Safety Act?**
  A: Class 1 material is material that has or is likely to be classified as Refused Classification under the National Classification Code. eSafety relies on the Classification Board to formally classify material as Refused Classification. eSafety can also determine that material is class 1 material on the basis that it is *likely* the Classification Board would classify it as such.

  Since X Corp's legal challenge during the Wakeley incident, eSafety has relied more on Classification Board decisions rather than making its own assessment that the material is likely to be classified as Refused Classification.

  More recently, X Corp challenged a Classification Board decision to classify footage of the Charlie Kirk assassination as Refused Classification.

- **Q: How long does the classification process take?**
  A: Applications are either standard, with up to a 20-business-day timeframe, or priority, with up to 5 business days, excluding weekends, public holidays and shutdown periods. There is no out-of-hours classification or advice mechanism.

- **Q: What impact do review decisions have on eSafety's enforcement powers?**
  A: Where the Classification Review Board changes a classification, eSafety may be required to change its regulatory response, including revoking any removal notices that were issued based on a previous classification.

5: KEY ISSUES BRIEF: Online Safety Codes – Age-Restricted Material

Talking Points

- Three Online Safety Codes relating to Search Engines, Hosting and Internet Carriage Services **came into effect on 27 December**. Further Online Safety Codes will come into effect **on 9 March**.
- These **Age-Restricted Material Codes** aim to prevent children from being exposed to class 1C (Refused Classification material depicting fetish and fantasy material but not falling within class 1A) and class 2 (R18+ or X18+) material such as pornography and material which promotes or instructs in self-harm, and to give users the ability to better control their online experiences.
- The Codes focus on preventing children's **accidental or unsolicited exposure to material such as pornography**. eSafety research has found that **40% of young people first encountered pornography unintentionally**, e.g. while searching for something else.
- These Codes codify existing good practice but also require the tech industry to uplift their safety practices e.g., protecting children from harmful content generated through AI and chatbots.
- The Codes require certain sections of the online industry to implement appropriate age assurance proportionate to their risk and in compliance with applicable privacy laws.
- The Codes complement the Social Media Age Restrictions by adding an extra layer of protection on services not covered by that framework, as well as making age-restricted social media platforms safer for those aged 16-17 and for any under 16s who bypass controls to gain or keep accounts. They also give Australians of all ages information, tools and options to avoid high-impact material if they don't want to see it.

Key Issues

- The **Search Engine Services Code** which came into effect on 27 December requires services like Google and Bing to take steps to minimise children's accidental exposure to pornography, and ensuring searches relating to self-harm and suicide to helplines and authoritative health resources in results, prioritising these over sites which may encourage or exacerbate these harms.
- The **Designated Internet Services**, **Relevant Electronic Services**, **Social Media Services** (**Core Features** and **Messaging Features**) and **App Distribution Services** Codes were registered on **9 September 2025** and will come into effect starting on **9 March 2026**. This will include some age assurance measures.
- Age assurance measures for 'gatekeeper' services (search engines and app stores) have an additional six-month lead time:
  - Age assurance in the Search Code will come into effect on 27 June 2026.
  - Age assurance in the App Store Code will come into effect on 9 September 2026.
- eSafety published Regulatory Guidance to assist industry with implementing the Unlawful Material and Age-Restricted Material Codes and Standards on 5 December 2025. The

guidance complements our guidance on intersecting schemes, including the Social Media Age Restrictions.

Background

- The Codes were drafted by 5 industry associations representing 8 sections of online industry:
  - Australian Mobile Telecommunications Association (AMTA), Communications Alliance (now Australian Telecommunications Alliance, or ATA), Consumer Electronic Suppliers Association (CESA), Digital Industry Group Inc (DIGI) and Interactive Games and Entertainment Association (IGEA).
- The Codes aim to prevent children from accessing or being exposed to age-restricted material, such as **pornography**, **self-harm** (and **disordered eating**) material, **high-impact violence** material and **simulated gambling** material. They also aim to give all end-users the ability to better control their online experiences.
- It took roughly 14 months from the issue of the section 141 notices and the publication of eSafety's position paper on 1 July 2024 until the registration of all the ARM Codes.

Age Assurance

- In developing the codes, the industry associations drafted age assurance requirements proportionate to the risks different services present in exposing children to age-restricted material.
- Examples of age assurance measures that can be applied under the codes include: photo ID, facial age estimation, credit card checks, digital identity wallets or systems, parental attestation of a child's age, and use of AI to estimate age based on relevant data inputs. There are no requirements to use a particular form of age assurance, including no requirement in the Code for companies to use Government ID as a method. Age assurance methods must meet the definition of 'appropriate age assurance' appearing in the Codes Head Terms.
- The government-sponsored Age Assurance Tech Trial provides a helpful overview of measures currently available in the market and how they can be applied in the Australian context in a way that is private, effective, and efficient.

Key points on why age assurance measures are needed across the tech stack

Code	Who/When	Why
Search engine services	Logged in account holders before returning results containing online pornography or high-impact violence	• eSafety's Age Verification Roadmap report (p 94) showed search engines can be gateways to pornography. • eSafety's 'Accidental, unsolicited and in your face' research (Sept 2023) (p 14): 40% of young people first encountered porn unintentionally, eg while searching for something else.
App distribution services	Before downloading apps rated 18+	• eSafety's Age Verification Roadmap report (p 291) showed app stores can provide meaningful safeguards around the types of apps children can access.
Social Media Services	Services that allow online pornography or self-harm material (e.g. Reddit, X)	• eSafety's Behind the Screen Report (Feb 2025) (p 11): showed 80% of 8- to 12-year-old children are accessing social media, despite ToS setting a min age of 13. • eSafety's 'Accidental, unsolicited and in your face' research (p 28): 60% of young people who had seen porn had done so via social media or a messaging service. • eSafety's Digital use and risk research (Jun 2025) (p 30): 75% of those who had been exposed to content associated with harm had most recently experienced this on social media.
Relevant Electronic Services	Predominant purpose of sharing online pornography or self-harm material	• eSafety's Digital use and risk research (Jun 2025) (p 30): while almost all children have used communication platforms, only 6% of children interviewed were most recently exposed to content associated with harm on

	(e.g. Chaturbate)	communication platforms. Accordingly, only the highest-risk services here are required to implement age assurance.
	Online games rated 18+ (including simulated gambling) (e.g. Grand Theft Auto 5 online)	• These games are considered age-restricted because they have been rated in line with the Classification Scheme. • Online games are also highly interactive, which makes them higher-impact under the classification guidelines.
Designated Internet Services	Services deemed highest risk for exposure to porn and self-harm material (e.g. Pornhub)	• eSafety's 'Accidental, unsolicited and in your face' research (Sept 2023) (p 27): 70% of young people who had encountered pornography had seen it on porn sites.
Generative AI/AI Chatbots	Services deemed highest risk for exposure to sexually explicit or self-harm material (e.g. nudify services)	• There has been significant reporting about how AI chatbots have engaged in sexually explicit conversations with children, or conversations which encourage suicidal ideation, self-harm and disordered eating. See e.g., the reporting in late August 2025 where ChatGPT allegedly instructed and encouraged a child aged 16 in California on how to commit suicide.

Other key measures in the Age-Restricted Material Codes:

Code	When	What
Search engine services (in effect now)	Non-logged in users	• Automatically blur pornography and high impact violence material for non-logged in users. • Redirect to help services in response to searches about self-harm or suicide
Gen AI/AI Chatbots	Lower-risk services	• Implement systems that prevent the service from being used to generate age-restricted content. • Regularly test and improve these systems
Social Media Services	Services that do **not** allow online pornography or self-harm material	• Automatically detect and remove this material in line with Terms of Service, and continuously improve these detection and removal systems (including to increase accuracy)

Question about Parliamentary scrutiny/public consultation

- By the Parliament's own design of the Online Safety Act, industry-drafted codes are not subject to Parliamentary scrutiny.
- Standards are subject to a period of Parliamentary scrutiny and disallowance.
- eSafety can only determine a Standard if:
    - industry associations fail to submit codes in response to a request, or
    - the Commissioner considers that a proposed code does not provide appropriate community safeguards to deal with one or more matters specified in a request, or
    - Any indicative targets for progress in code development specified in a request are not met, or
    - the Commissioner refuses to register a code submitted by industry (which is a reviewable decision before the Administrative Review Tribunal).
- Industry associations were required to conduct a minimum 30-day public consultation on the Codes and consider submissions. They did so from 22 October to 22 November 2024.
- eSafety also produced documents to and appeared before the Environment and Communications References Committee inquiry into the Internet Search Engine Services Online Safety Code. The Committee made no recommendations about the Online Safety Codes in their final report.

Other FAQs: Accessing online porn and adult content | FAQS | | eSafety Commissioner

6: KEY ISSUES BRIEF: BOSE (Expectations)

Talking Points

- Meaningful transparency is essential to drive change. Without access to information from industry participants, eSafety cannot fulfil its functions and industry cannot be held accountable. These processes shine a light on industry practice and drive change.
- The Expectations have been in effect since early 2022. Since then, eSafety has given **23 non-periodic notices, 8 periodic notices**, **7 information requests**, and published **7 transparency reports**.
- Our most recent report (published 5 February 2026) indicates mainstream platforms are taking some steps to improve safety including improving ability to detect and respond to child sexual exploitation and abuse. However, platforms should take further steps to detect new CSEA, address livestreaming of CSEA and stem the proliferation of sexual extortion of children and adults.
- We anticipate publishing a report of findings related to four generative artificial intelligence companion app providers in the coming months. This will be complemented by a report on the results of a pulse survey on the use of AI Chatbots by 10-17 year old children. Combined these reports will provide a view of how Australian children are using chatbots and what some companies are doing to protect children on these services in an Australian context.

Key Issues

Periodic notices

- Periodic reporting notices enable eSafety to understand whether providers are complying with the Expectations over time.
- The first round of periodic notices was given in July 2024 to **Apple, Discord, Google, Meta, Microsoft, Snap, Skype** and **WhatsApp.**
- The notices require these providers to report every 6 months for a two-year period on their compliance with the Expectations, focusing on child sexual exploitation and abuse (CSEA), including grooming, and sexual extortion of children and adults.
- eSafety published the first periodic transparency report in August 2025.
- On 5 February 2026, eSafety published the second periodic transparency report.
- Key takeaways from report 2 include:
    - Mainstream platforms are taking some steps to improve the safety of their services including improving or expanding the tools they use to detect CSEA. Examples include:
        - **Microsoft** - reported **expanded use of hash matching** for known CSEA images in email attachments in Outlook, and images and videos in OneDrive.
        - **Discord** - reported it **started hash matching** to detect known CSEA videos.
        - **Meta** - used **more sources** for its lists of terms and language indicators to detect sexual extortion (NCMEC and Thorn).
        - **Skype** - implemented proprietary **program for calls** in certain regions, including Australia, for high-risk users. If CSEA is detected, the video is disabled.

        - **Google joined Take it Down** – a global hash-matching service operated by NCMEC which helps remove and prevent the distribution of online nude, partially nude or sexually explicit photos and videos of children under 18.
        - **Apple** - continued development of its **Communication Safety** feature (and Sensitive Content Warning) with the aim of expanding to more of its services.
    - Despite these improvements, significant safety gaps remain. eSafety calls on industry to take further steps to:
        - address online CSEA / CSEA **livestreaming**;
        - detect **new CSEA**; and
        - stem the proliferation of **sexual extortion** of both children and adults.

Non-periodic notices

- Non-periodic notices focus on steps providers are taking to meet the Expectations during a particular period. Our approach is to focus on specific Expectations and issues of high harm.
- eSafety's most recent non-periodic notices were given in **October 2025 to four generative artificial intelligence (GenAI) companion app providers – Character AI, Chai, Nomi and Chub AI**. These notices will provide information about how these companies are mitigating harms relating to children on their services including child sexual abuse material, self-harm, suicide, and pornography, which includes sexualised role play.
- eSafety will publish a summary of the findings in the coming months. We will also publish the results of a pulse survey of children aged 10-17 and their use of chatbots. These 2 reports will be the first to provide a view of how Australian children are using chatbots and what some companies are doing to protect children on these services in an Australian context.
- As well as supporting public transparency, this work will help inform eSafety's compliance monitoring of the Unlawful Material Codes and Standards which are already in effect and the Age Restricted Material Codes which will all be in effect on 9 March 2026.
- eSafety will continue focusing on issues of high harm, including harms occurring on GenAI and on services excluded from the SMMA.

Information requests

- In addition to notices, eSafety may also obtain information from providers under section 20 of the BOSE Determination, which sets out the expectation that providers will provide certain information to the Commissioner within 30 days on request. Information requests are not backed by civil penalties, but can result in a 'statement of non-compliance'.
- This information-gathering mechanism was used to collect information from 7 providers in late 2024 in relation to age assurance and informed eSafety's **Behind the Screen report** published in February last year, which provided important insights to inform our work on the SMMA.

Difficult Questions

If providers don't meet Expectations, why haven't you taken enforcement action?

- There is no penalty for not complying with the Basic Online Safety Expectations, they are about transparency rather than enforcement.
- The periodic notices are intended to keep the pressure on these companies by requiring them to answer the same set of questions over two years (until August 2026).
- eSafety intends to publish regular transparency reports following responses to these notices. These reports serve to incentivise meaningful safety improvements across the technology sector and hold companies accountable for protecting their most vulnerable users. They also exemplify to industry the improvements in online safety being made and tech solutions available to keep users, particularly children, safe on their services.

- Where we identify ongoing safety gaps, eSafety will consider all options, including the codes and standards which require providers to detect, disrupt and deter CSEA.

If GenAI providers are not meeting the Expectations what action will you take?

- While the Expectations are not enforceable, the Age Restricted Material Designated Internet Services (DIS) Code is enforceable and will take effect on 9 March 2026. Requirements in the Code include:
    - **Design for safety** to prevent illegal and harmful content.
    - **Age assurance on highest-risk features** so only adults can access functions that could generate high-impact material, such as self-harm and online pornography.
    - **Extra protections on medium-risk services without age assurance** to prevent children generating harmful or age-inappropriate content.
    - **Quick and consistent moderation and escalation**, including working with police.
    - **Clear ways for users to report issues and seek help**, plus regular risk assessments.
    - **Regular testing and review**, with documented results and improvements over time.
- eSafety will use the full range of our powers to ensure compliance and deter non-compliance. This can include seeking penalties from the Federal Court of up to A$49.5m.

Why haven't you sent notices to Grok and ChatGPT?

- eSafety did give a notice to X which sought information on Grok's safety measures to prevent pro-terror material. X challenged that notice in the Administrative Review Tribunal (ART), and the Tribunal remitted the decision to eSafety.
- It is open to eSafety to issue further notices to additional companies with AI capabilities in due course to continue building a picture of online safety measures across the industry.
- In addition to BOSE periodic and non-periodic notices, eSafety also has powers under mandatory Codes and Standards to investigate allegations of non-compliance.

Why give a notice to Chub AI when it's not available in Australia?

- Prior to issuing these Notices, the Chub service was available to Australians.
- When Chub AI Inc. was engaged about specific expectations under the BOSE to protect children, it decided to withdraw its service from Australia on 1 October 2025.

How have you dealt with non-compliance with reporting requirements?

- eSafety has taken action in relation to 4 notices given to 3 companies:
    - **X Corp – 1 x Infringement Notice and 2 x Service Provider Notifications**
        - **Infringement Notice** of $610,500 & **Service Provider Notification** to X Corp. re: February 2023 CSEA Notice. X Corp. applied for judicial review, which was dismissed in October 2024 and dismissed again on all grounds with costs on appeal to the Federal Court. eSafety filed civil penalty proceedings in December 2023 following non-payment of the infringement notice, which are ongoing.
        - **Service Provider Notification** to X Corp. in January 2024 for giving inaccurate and incomplete response to June 2023 Online Hate Notice.
    - **Google – 1 x Formal Warning**
        - Formal Warning to Google for failure to adequately answer some questions in the February 2023 CSEA Notice.
    - **Telegram – 1 x Infringement Notice**

- Infringement Notice of $957,780 for failure to respond to the notice by the deadline. Telegram's response was provided to eSafety 160 days after the due date. Telegram did not pay the infringement notice. Telegram applied to the Federal Court for judicial review of the notice, claiming the legal entity to which eSafety gave the Notice (i.e. Telegram FZ LLC) is not the legal entity that provides the Telegram service.
- Telegram has discontinued the proceedings. eSafety is considering next steps, and whether to take additional action for failure to comply with the notice.

Are your enforcement powers ineffective?

- The 2024 review of the Online Safety Act recommended strengthening our powers, which we support.
- We have seen improvements in online safety through the actions we have taken to date and the progress shown in the latest transparency report is evidence of this.
- While the Expectations themselves are not enforceable, the requirement to respond to a transparency notice is enforceable, with a maximum penalty of $825,000 per contravention.

7: KEY ISSUES BRIEF: Social Media Minimum Age (SMMA)

Talking Points

- Social media companies removed access to about 4.7 million accounts identified as belonging to children under 16 within the first half of December to comply with Australia's social media minimum age.
- eSafety continues to engage with the major social media platforms to monitor and assess compliance. We are looking at systemic compliance, not individual instances of accounts.
- The process of age assurance requires time to complete fairly and accurately, but early indications are that platforms are making meaningful attempts to prevent under-16s from holding accounts.
- eSafety is taking a proportionate and risk-based approach, focusing on services with the greatest number of children, where there are higher risks of harm.

Key Issues

- The SMMA obligation requires age-restricted social media platforms to take reasonable steps to prevent Australian children under 16 from having accounts on their platforms. The obligation is on platforms, not children or their parents or carers.
- As our regulatory guidance makes clear, the social media minimum age (SMMA) obligation is about having appropriate and effective systems and processes in place, not individual instances of accounts.
- eSafety is taking a proportionate and risk-based approach, focusing on services with the greatest number of users, where there are higher risks of harm, and the steps companies are taking to prevent the youngest users from having accounts.
- Overall, early indications are that platforms are making meaningful attempts to prevent under-16s from having accounts.
- While the SMMA obligation is not a complaints-based scheme, eSafety has a SMMA form on our website through which the public can provide information to eSafety about platform implementation and compliance with the obligation.
- eSafety is monitoring migratory patterns of social media use by children under the age of 16 and will adjust its regulatory focus as needed to ensure age-restricted social media platforms operating in Australia are complying with their obligations.

Background

Regulatory guidance

- Consistent with other international approaches, eSafety has taken a **principles-based** approach. Reasonable steps should respect and protect fundamental human rights and be:
  - **Reliable, accurate, robust and effective**

    - **Privacy-preserving and data-minimising**
    - **Accessible, inclusive and fair**
    - **Transparent**
    - **Proportionate**
    - **Evidence-based and responsive to emerging technology and risk.**
- We expect age-restricted platforms take steps to:
    - **find** accounts held by under-16s, **and deactivate/remove** those accounts with kindness
    - **prevent** under-16s from opening **new accounts**
    - **prevent workarounds** that may allow under-16s to bypass the restrictions
    - have **processes to review** decisions and correct errors, so no one is removed unfairly
    - **provide accessible and clear ways for people to report** underage accounts.
- eSafety has encouraged providers to take **a layered approach**
    - Depending on the platform and userbase, this could involve a range of methods for estimating or inferring age at sign up, and then only requiring more information when there is a signal they are under 16, e.g. if they are reported by another user.
    - Steps reflect platforms' circumstances; no specific form of age assurance is required.

Assessing compliance

- We are considering a range of insights to monitor platforms' compliance.
- eSafety has information-gathering powers to require age-restricted social media platforms to give **any information** relevant to their compliance with the SMMA obligation.
- An age-restricted social media platform must comply with an information-gathering notice to the extent that they are capable of doing so. Non-compliance may result in enforcement action taken and subject to civil penalty of up to $825,000 for each contravention.
- We will be as transparent as possible, however eSafety will not be publishing specific numbers or detailed information obtained using our information-gathering powers, to maintain the integrity of investigations and ensure any potential enforcement action is not compromised.
- While the SMMA is not a complaints-based scheme, eSafety has provided an SMMA online form through which the public can provide information to eSafety about platform implementation and compliance with the obligation. We also have a 'contact us' form that the public can use for enquiries. From **10 December 25 to 20 January 26**, we have received **123 SMMA form submissions** and **380 enquiries**, ranging from broad support of the SMMA, to concerns about accounts not being removed, as well as concerns about the policy in general.

Difficult Questions

If asked whether specific platforms are complying/potential compliance action

- We are considering a range of insights to assess platforms' compliance, including information received in response to information-gathering notices, submissions from members of the public, and engagement with a range of stakeholders, including discussions with age assurance providers and the platforms themselves.
- Our regulatory guidance outlines some examples of the types of information we can require platforms to provide us. This includes information about the tools and technologies the platforms are using, and how many accounts have been actioned since 10 December.

If asked about reports of circumvention

- This is something we are actively considering
- As our regulatory guidance makes clear, the SMMA obligation is about **having appropriate and effective systems and processes** in place, **not individual instances of accounts**.
- The process of age assurance can require time to complete fairly and accurately.
- Many platforms already use age-inference models; a user might pass initial age checks but should later be detected if there is a signal they are under-16.

- We continue to make clear our expectation that platforms continuously improve their efforts to detect and deactivate under-16 accounts.

If asked about migration to other services

- eSafety is monitoring migratory patterns of social media use by children under the age of 16 and will adjust its regulatory focus as needed to ensure age-restricted social media platforms operating in Australia are complying with their obligations.
- We are aware there have been increases in downloads of some emerging apps, for example, Lemon8, Yope and Coverstar. We are monitoring this over time to stay alert to any significant migration of under 16s to new platforms. To date there is no indication this is occurring.
- We did expect this would happen and we continue to engage with platforms to ensure they are meeting their regulatory obligations. We have been engaging with companies located across the world, including France, Singapore and China to encourage platforms to self-assess whether they meet the definition of an age-restricted social media platform and to monitor for migration trends.

If asked why GenAI platforms/other platforms aren't included

- The definition of an age-restricted social media platform is outlined in the Online Safety Act.
- eSafety has provided guidance to platforms to support them to assess whether they might meet the definition of an age-restricted social media platform and are required to comply with the SMMA.
- In the lead up to the SMMA taking effect, eSafety published information on preliminary assessments of a range of online services against the definition of 'age-restricted social media platform' to provide clarity to the public about which services we believe are subject to the SMMA obligation.
- Given the sheer number of online services and the way they are constantly changing, it is not possible for eSafety to maintain an exhaustive list of all services which meet the conditions for age-restricted social media platform at any given time. The onus is on providers who offer services in Australia to assess for themselves whether they meet the conditions, and if so, to comply.
- eSafety has engaged with a wide range of services beyond those listed on our website, and a number of services have assessed that they are covered by the SMMA, including BlueSky, Lemon8, Wizz and Yubo.
- Importantly, eSafety does not have an ability to formally determine whether a service is subject to the SMMA obligation. Rather, eSafety will conduct an assessment of a service when we are considering regulatory action, such as giving an information-gathering notice. If a service disagrees with eSafety's assessment and chooses not to comply with a notice or the SMMA obligation, it will ultimately be a matter for the courts to determine.
- The SMMA is just one of many measures designed to encourage greater transparency and accountability, prevent online risks, and limit harm – especially for children.

If asked about users 16 and over being removed

- We expect that platforms implement measures to comply with the SMMA object in an empathetic and transparent way, communicating clearly with their users about any actions taken on their account, and providing them with ways to access and save their information.
- Our guidance to industry includes an expectation that platforms provide accessible review mechanisms for users who believe they've been wrongly flagged as under 16.
- We understand where there may be instances where platforms initially get it wrong – however we expect that platforms are treating their users with care and fairness.

If asked about the High Court challenge

- Reddit is exercising its democratic right to challenge these laws.
- We will continue to implement this legislation and assess compliance.
- Notwithstanding the legal challenge, Reddit is engaging with eSafety cooperatively on compliance.

9: KEY ISSUES BRIEF: Corporate Matters

This brief is based on the 2025/26 Portfolio Budget Statements. The updated financial position will be reflected in the 2025/26 Portfolio Additional Estimates Statements. The Government expects to table the 2025/26 PAES in both the House of Representatives and Senate between 3-12 Feb.

Talking Points

- eSafety has a **departmental** appropriation budget of **$59.4m** for the 2025/26 financial year

  (this includes a whole of government savings target of $0.9m, applied in July 2025, but not reflected in the 2025-26 PBS)
- eSafety has an **administered** budget of **$2.5m** for the 2025/26 financial year
- eSafety total funding, as per program 1.3 in the Portfolio Budget Statements is **$71.7m**

  (this includes departmental and administered funding, ACMA direct appropriation, other revenue and depreciation expenditure)
- eSafety spent a total of **$57.8m** in the 2024/25 financial year

Key Issues

2025/26 Budget as per the 2025-26 PBS

Current Funding as per the 2025-26 PBS	$'000
Administered Funding - Grants	2.500
Total Administered Expenses	**2.500**
Base funding	47.501
NPP eSafety General awareness Initiative	0.100
NPP Be Connected	4.034
NPP National Strategy to Prevent Child Sexual Abuse	0.644
NPP TFA Technical Support	5.600
NPP Protecting Australians Online	1.633
NPP Internal legal and Compliance	0.782
Total Appropriation receipts	**60.294**
External revenue	0.190
ACMA direct appropriation funding	8.909
Less acma capital funding	-1.821
Expenses not requiring appropriation (depreciation)	1.677
Total expenses	**71.749**

Contact: s 22
Phone: s 22

Cleared by: Kathryn King, General Manager
Date: 29/01/2026

The **2025/26** budget includes the following key adjustments:

1. Increase from the previous year in relation to the Social Media Minimum Age (SMMA) program of **$9.353m** in operational funding and **$3m** in capital funding, a total increase of **$12.353m.**

   The year-by-year funding profile for SMMA is noted below:

SMMA Funding	2024/25 $m	2025/26 $m	2026/27 $m	2027/28 $m	Total $m
Operational funding	3.810	13.163	13.260	12.418	**42.651**
Capital funding		3.000			**3.000**
Total	**3.810**	**16.163**	**13.260**	**12.418**	**45.651**

2. Funding of **$0.6m** was provided in the 2025/26 Budget for the continuation of the National Strategy to Prevent and Respond to Child Sexual Abuse by one year to 30 June 2026.

   - This measure focuses on **safeguarding children and young people from online harm**, including sexual abuse and exploitation through **resources, guidance and training**. Through this project we support parents, carers and community organisations – increasing their knowledge and confidence in talking about online safety in ways that decrease shame and stigma.

3. A whole of government savings target was applied in July 2025. This represents a reduction of **$0.908m.**

2024/25 Financial expenditure breakdown

Expense type		2023/24 actuals $m	2024/25 actuals $m
Departmental	Operational expense	39.968	55.751
	Capital expense	0.730	0.043
	Total	**40.698**	**55.794**
Administered	Grants provided	2.335	1.979
	Be Connected expense	3.762	0
	Total	**6.097**	**1.979**
Total		**46.795**	**57.773**

The growth in overall expenditure represents an increase in funding for programs, including SMMA and the internal legal and compliance NPP.

The Be Connected program moved from administered expenditure to departmental expenditure in 2024/25.

Current Staffing

Staffing	As at 31 December 2025	As at 31 December 2024	Percentage change
APS (FTE)	241	186	30% increase
Contractors (FTE)	27	53	49% decrease
Total	**268**	**239**	**12% increase**

The growth in overall staffing numbers represents an in increase in staff associated with new programs, including SMMA (40 ASL) and the internal legal and compliance NPP (4 ASL).

The reduction in contractors represents the conversion of contractor roles to APS roles as per the Government's Strategic Commissioning Framework, and the conclusion of some labour hire contracts.

Background

Extract from the **2025/26 Portfolio Budget Statements**

Table 2.1.1: Budgeted expenses for Outcome 1

	2024-25 Estimated actual $'000	2025-26 Budget $'000	2026-27 Forward estimate $'000	2027-28 Forward estimate $'000	2028-29 Forward estimate $'000
Program 1.3: Office of the eSafety Commissioner					
Administered expenses					
Ordinary annual services (Appropriation Bill (No. 1) and Supply Bill (No. 1))	2,500	2,500	2,500	-	-
Administered total	**2,500**	**2,500**	**2,500**	-	-
Departmental expenses					
Departmental appropriation	61,110	67,382	61,608	59,313	54,914
s74 External Revenue[a]	1,442	190	150	-	-
Special accounts					
Appropriation receipts[d]	49,126	60,294	54,536	52,241	47,842
less expenses made from appropriations credited to special accounts[e]	(49,126)	(60,294)	(54,536)	(52,241)	(47,842)
Expenses not requiring appropriation in the Budget year[b]	1,724	1,677	1,536	1,264	1,126
Departmental total	**64,276**	**69,249**	**63,294**	**60,577**	**56,040**
Total expenses for program 1.3	**66,776**	**71,749**	**65,794**	**60,577**	**56,040**

8: KEY ISSUES BRIEF: Online hate and antisemitism

Talking Points

- Online hate—including antisemitism is not explicitly covered in the Online Safety Act (OSA), but eSafety can act when content meets thresholds within other schemes (e.g. cyber abuse, industry codes, Basic Online Safety Expectations).
- In response to the Bondi Beach shooting, eSafety undertook immediate surveillance of circulating footage, engaged platforms to ensure proactive moderation, and **managed 106 complaints** involving MA15+, Refused Classification or likely to be Refused Classification (RC) content.
- While major platforms removed graphic material under their own policies, some Refused Classification content remains on fringe sites such as Watch People Die. eSafety has issued removal notices and is assessing further regulatory action.
- eSafety has **received 19 online-hate complaints**—including three regarding antisemitism—none of which met the Adult Cyber Abuse threshold, as the material did not target a specific Australian individual or satisfy the 'serious harm' test.
- Following the Bondi Beach attack, eSafety has strengthened engagement with Government, the Department, and the Special Envoy to Combat Antisemitism to enhance regulatory responses to online hate and progress duty-of-care reforms under the Online Safety Act.

Key Issues

eSafety's approach and regulatory measures address online hate, including antisemitism

- Since the Bondi Beach shooting, eSafety has worked with Government, the Department, and the Special Envoy to Combat Antisemitism to strengthen measures addressing online hate.
- Guidance on online abuse is available on eSafety's website and can support people experiencing online hate.
- eSafety has strong referral arrangements with law enforcement, including a joint initiative focused on online threats and extremism, such as antisemitic content.

Overview of complaints related to antisemitism/online hate

- Since the 7 October 2023 atrocity, there was a marked but modest increase in reports related to hate speech:
    - the adult cyber abuse scheme received 102 complaints involving hate speech. Of these complaints:
        - 9 related to antisemitic content

        - 8 related to Islamophobic content
        - 23 related to other religion discrimination
        - 18 voluntary notifications were sent to services about potential breaches of their term of service
        - 6 Service provider notifications were sent to services for material that met the threshold for adult cyber abuse
        - 4 removal notices were issued to services for material that met the threshold for adult cyber abuse
        - 9 information gathering notices were issued to services to provide end user contact details and information for material that met the threshold for adult cyber abuse.

- Only 1 antisemitic complaint met the adult cyber abuse threshold - a complaint about antisemitic content targeting an ABC journalist was referred by the AFP's Special Operation Avalite. The platform removed the post for breaching its terms of service, and eSafety disclosed account details to the AFP under the OSA.
- On 30 April 2025 eSafety referred three adult cyber abuse complaints to the AFP's Special Operation Avalite for investigation. The complaints related to an Australian end user account posting antisemitic comments on social media in response to posts made by two prominent Jewish entertainers. The material did not meet the threshold for adult cyber abuse as the comments were not targeting an individual Australian adult so could not be actioned under eSafety's powers.
    - the Child cyberbullying scheme received 18 relevant complaints with elements of online hate, including:
        - 2 related to antisemitic content
        - 2 related to Islamophobic content
        - 14 related to other religious discrimination
        - 4 voluntary notifications were sent to services about potential breaches of their term of service
        - 1 Service provider notification was sent to a service for material that met the threshold for child cyber bullying
        - 5 information gathering notices were issued to services to provide end user contact details and information for material that met the threshold for child cyber bullying
        - 2 school notifications were sent.
    - The IRC scheme received the 49 relevant complaints to online hate and violent extremism, including:
        - 36 related to antisemitic content (from 5/2/24 to 24/12/25)
        - 4 related to Islamophobic content (from 5/2/24 to 24/12/25)
        - 9 related to the Israel/Gaza conflict (from 10/10/24 to 24/12/25)
        - 32 removal notices were issued for material relating to manifestos and other refused classification material (from 2023 to current)
    - The IRC scheme received a further 232 complaints with the following content tags:
        - ISIS home page
        - Christchurch shooting video
        - Towards a better society manifesto- Jordan Patten

    - 2083: A European Declaration of Independence manifesto
    - Militant Accelerationism- Terrogram, The hard Reset – Terrogram, and The Great Replacement
    - The West Has Fallen- Solomon Henderson
    - The Great Replacement Manifesto
- Of those, 205 were further tagged as being Refused Classification (1)(a), (1)(c) or 9A.
- Between 1 July 2025 and 22 January 2026, eSafety received 19 online-hate-related complaints; none of which met the Adult Cyber Abuse threshold.
- Complaints typically relate to content targeting individuals or groups based on race, religion, gender, disability, etc.
- To act under Adult Cyber Abuse or Child Cyberbullying schemes, hate must target a *specific* Australian person—not a group.
- Content promoting violence against a group may be actionable under the Online Content Scheme.

Specific antisemitic complaints (July 2025–present)

- **Complaint:** Posts on X targeting Jewish individuals or pro-Israel voices—offensive/malicious but did not meet the "serious harm" requirement.
- **Complaint:** Complainant called a "zio-freak"—offensive but not serious harm.
- **Complaint:** Instagram highlights labelling accounts as "Zionists"—no intent to cause serious harm.

Bondi Beach Shooting

- The eSafety on-call capability was activated immediately following the Bondi Beach attack to commence monitoring for any footage of the attack circulating online. The team identified multiple videos capturing the shooting and aftermath were uploaded and circulated across social media platforms, including Instagram, X and Threads, as well as some fringe sites.
- eSafety monitored for footage of the attack and assessed content against Class 1 and Refused Classification thresholds for removal.
- **106 complaints** were received comprising mostly of bystander videos.
- eSafety also engaged with platforms to ensure they were being proactive in their response to content on their services, including removing violent material under their own terms of service and applying warnings on gratuitous violence to prevent inadvertent or accidental viewing.
- Initial footage was assessed and referred to the Classification Board, which classified the material as MA15+.
- A later complaint containing new footage was classified by the Board as Refused Classification (RC). Content with an RC rating is considered illegal and cannot be hosted, shared or distributed in Australia.
- RC content was identified on a fringe website named Watch People Die (WPD) and X; removal notices were issued. X geo-blocked the content however the content is still available on WPD. eSafety is considering what further action may be appropriate.

- eSafety has recently been alerted to ‘gamified’ versions of the incident circulating online overseas, the nature and origin of these are not yet clear, however no content has yet been identified as available from Australia.

Background / Difficult questions

Questions about how eSafety’s regulatory measures address online hate, including antisemitism

- eSafety can issue removal notices where content meets legislative thresholds for cyber abuse or cyberbullying—targeting must be toward a specific Australian individual.
- Material showing or encouraging terrorist acts/extreme violence can be actioned under the Illegal and Restricted Content scheme.
- eSafety enforces industry codes and standards across eight sectors to reduce systemic availability of harmful material.
- Transparency powers allow eSafety to request information and has published reports in 2024 (Online hate on X) and 2025 (Terrorist/Extremist material).

Questions about eSafety’s view on online safety reform to address online hate

- eSafety supports industry-wide uplift in responding to online hate.
- Government has committed to a duty of care under the OSA, which the Minister has stated will address online hate.
- eSafety continues to work with the Department/Government on OSA reforms, including the duty of care.
- Engagement continues with the Special Envoy to Combat Antisemitism following the Bondi shooting.

Questions about eSafety’s engagement with law enforcement

- eSafety has established partnerships with federal, state, and territory law enforcement agencies facilitating the rapid referral of harmful material. This includes complaint referrals to the AFP’s Special Operation Avalite (see above).
- Ongoing shared intelligence informs regulatory actions on terror/extreme violent content.
- eSafety refers matters to law enforcement where criminal investigation is appropriate.

Questions about eSafety’s response to the Bondi Beach terror attack

- eSafety monitored for attack footage, assessed complaints, and engaged platforms to remove violent material that met the threshold for removal (Class 1 material or ‘Refused Classification’ as defined under the National Classification Scheme.
- On 15 December, eSafety wrote to six platforms requesting information on their response to the attack and content that may be available on their services. These platforms included Meta, TikTok, Snap, YouTube, X Corp and Reddit.

- All platforms responded to eSafety's information request and indicated they were removing content in accordance with their terms of service and had safety tools in place for under 18s for graphic content.

- RC content was identified on a fringe website named Watch People Die (WPD) and X; removal notices were issued. X geo-blocked the content however the content is still available on WPD.

Questions about eSafety's position on the Special Envoy's Plan to Combat Antisemitism

- eSafety defers to Government on the Plan.
- The Plan calls for additional measures to address online hate.
- eSafety has engaged with the Special Envoy to provide online safety advice.

Questions about the Special Envoy to Combat Islamophobia's report (National Response to Islamophobia)

- Implementation decisions are matters for Government.
- The report also calls for strengthened responses to online hate.
- Policy and legislative matters sit with the Department.

Antisemitic complaints and suggested response:

Q: Why do some antisemitic complaints not meet the cyber abuse threshold?
A: The OSA requires that material targets a *specific* Australian adult (s36). Many complaints relate to groups, not individuals.

Q: Other reasons?
A: Some do not meet the "serious harm" requirement (s7), particularly where content reflects strong disagreement or political argument rather than intent to cause serious harm.

10: BACK POCKET BRIEF: eSafety Key Statistics

Investigations stats - Bryan Downie					
1 July 2025 – 31 December	**Adult Cyber Abuse**	**Cyberbullying**	**Image-based Abuse**	**Illegal and Restricted Content**	**Totals**
Complaints received (% change from PY[1])	2,666 (+62%)	1,801 (+2%)	4,184 (+30%)	14,648 (+52%)	23,299 (+44%)
% that don't meet legislative threshold	95%	72%	28%		
# complaint notifications[2])	5	113	567	389	1,074
# of SPNs	57	143	8	0	208
# formal removal notices	7 issued, 6 removed	3 issued, 1 removed	0 issued, 0 removed	36 issued, 16 removed	46 issued, 23 removed
# of link deletion notices sent for class 1	N/A	N/A	N/A	0	0
# of remedial directions	N/A	N/A	1	N/A	1
Top 3 sub-categories by volume	*Defamation (43%)*	*Nasty comments/ serious name calling (39%)*	*Sexual extortion (49%)*	*Child sexual abuse / child abuse / paedophile activity (68%)*	
	Nasty comments/ name calling (26%)	*Offensive/ upsetting pictures and/or videos (24%)*	*Child sexual exploitation (12%)*	*Extreme, offensive or adult content (15%)*	
	Doxing (13%)	*Unwanted contact (11%)*	*Posted online (8%)*	*Sexually explicit (6%)*	
% of complainants who identify as gender diverse	1.2%	1.0%	1.8%	N/A	

No formal warnings or infringement notices were issued under the complaint schemes in 1 July – 31 December period

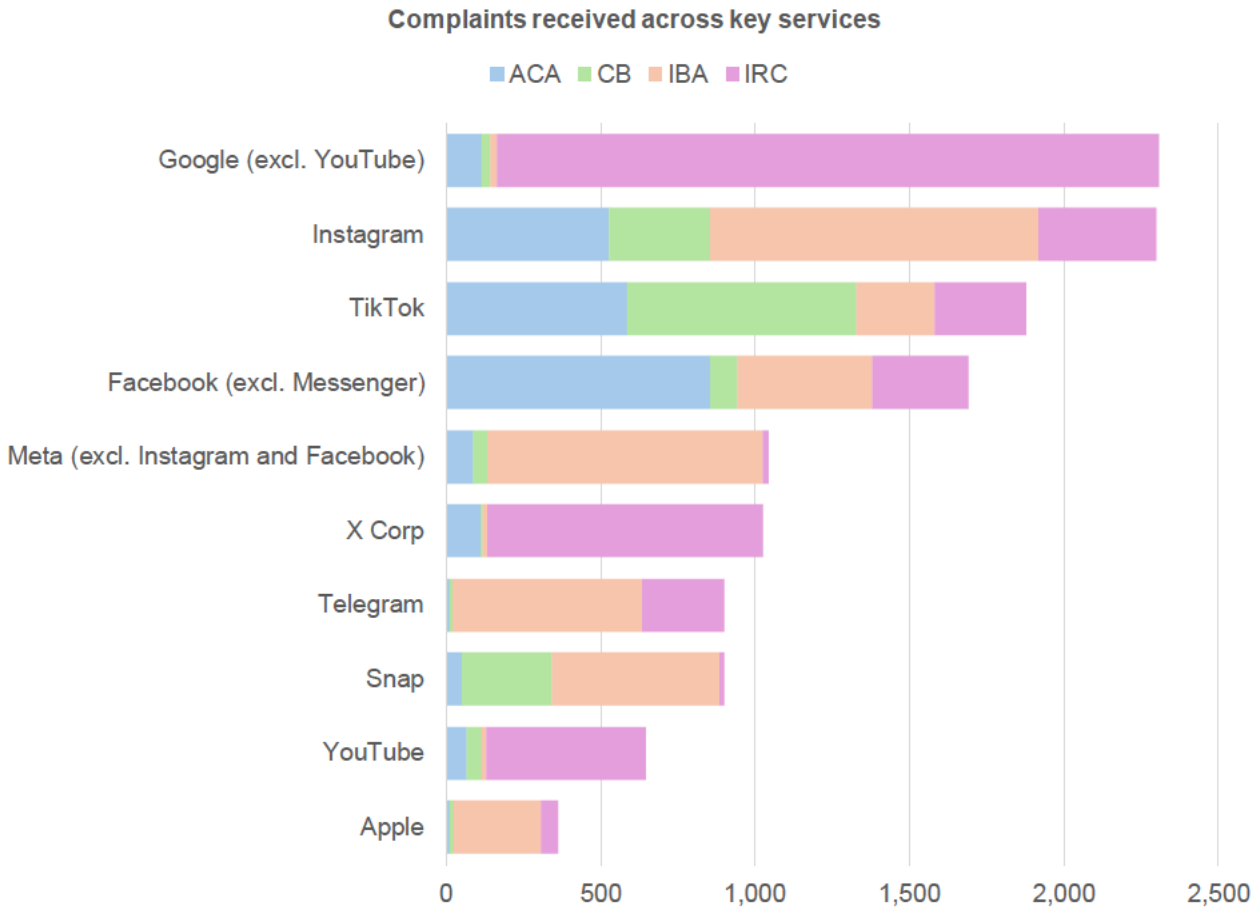
[1] Compared to 1 July to 31 December 2024.
[2] A complaint notification may include multiple urls

Topical issues under the complaint schemes	
TVEC	Makes up 6% of complaints received under our IRC scheme. This financial year saw a 496% increase in number of complaints made. 89 investigations completed each containing content that had been classified as RC. A refinement of assessment processes means complaints are only progressed to investigations where the content is sufficiently serious has resulted in a significant decrease in the number of investigations.

Online hate	Race-related Adult Cyber Abuse Complaints
	
	• eSafety received 12 complaints involving religious or race-related abuse:
	  • 3 antisemitic
	  • 6 Islamophobic
	  • 1 targeting Aboriginal/Torres Strait Islanders
	  • 2 other race-based discrimination
	
	No antisemitic complaints met the adult cyber abuse threshold.
	
	Gender related Adult Cyber Abuse Complaints
	
	In 1 July – 31 December 2025, the adult cyber abuse scheme received 5 complaints relating to gender.
	
	Child Cyberbullying
	
	• eSafety received 3 complaints that involved belief-based abuse
	• None of these complaints met the cyberbullying threshold
	• 5 complaints were received relating to the complainant's gender
	• 2 of these complaints did not meet the cyberbullying threshold
	• 37 complaints were received relating to the complainant's race/ethnicity
	• 19 complaints met the cyberbullying threshold

CSAM	Comprised 12% of IBA complaints (9% of total reports in July – December 2025). 73% increase in number of CSAM reports made compared to July – December 2024. Comprised 68% of IRC complaints (78% in July – December 2024). 33% increase in number of complaints made compared to July – December 2024. 2,062 investigations were completed that contained Class 1 (RC) content. Noting the investigation numbers are commonly higher than the complaints numbers due to the amount of own motion investigations that are instigated following the detection of CSAM. For example, a single complaint may contain multiple URL's of CSAM which correlate to multiple own motion investigations. As with TVEC complaints, a refinement of assessment processes means complaints are only progressed to investigations where the content is sufficiently serious has resulted in a significant decrease in the number of investigations.

Complaints across key services (IBA and IRC figures are based on keyword searches and may include false positives (or exclude false negatives))				
Service	**ACA**	**CB**	**IBA**	**IRC**
Google (excl. YouTube)	118	26	23	2,142[3]
Snap	53	290	542	16
Tik Tok	587	741	254	297
Apple	12	13	285	54
Meta	88	47	891	19
Instagram	529	326	1,062	383
Facebook (excl Facebook Messenger)	854	91	434	312
You Tube	69	49	13	517
X Corp	115	4	16	892
Telegram	12	11	614	265



[3] Caution should be had in interpreting this figure as it includes results for google searches where the content is ultimately on another service

Comms stats – Awareness and reach Performance Q1,Q2 2025/26 (comparison is vs same period in 2024/25)		
1 July 2025 – 31 December 2025	**eSafety.gov.au**	**BeConnected.gov.au**
Webpage visits	**+ 8.7 million users** (+167%) Of this traffic: • 75% was organic traffic • 12.4% was direct traffic • 4.1% was paid traffic *Source: Google Analytics*	**+207K users** (-8%) Of this traffic: • 42.13% was organic traffic • 41.44% was direct traffic • 7.67% was email traffic • 5.9% was referred traffic *Source: Google Analytics*
Web resource downloads	+146K downloads *Source: Google Analytics*	+13K downloads *Source: Google Analytics*
Email activity	+893K emails sent (+46%) from 91 email campaigns (+21%) *Source: MarComms reporting - per Monthly MO Reports*	• Be Connected learner and subscriber community has grown to 46k (+15.68%) *Source: Campaign Monitor Reporting*
	More than 87,000 subscribers at end December 2025 (+42%) *Source: MarComms reporting - per Monthly MO Reports*	• +34K Be Connected learner newsletter subscribers (+ 3%) • +12K Be Connected newsletter only subscribers (+ 94%) *Source: MarComms reporting - per Monthly MO Reports*
Media stats	eSafety issued 34 media releases and statements *Source: eSafety website*	
	Almost 8,000 media mentions *Source: Isentia Reporting*	
	Estimated media reach of more than 475m (audience), or media Advertising Space Rate (ASR) of more than $211m. *Source: Isentia Reporting*	
Social media stats	Over 95,000 social followers at December 2025 (12% growth across Jul to Dec 2025, though the figure is lower as eSafety's X account is no longer in active use) *Source: MarComms reporting - per Monthly MO Reports*	

	331 social media posts published	
	2.6m post impressions in Jul to Dec 2025 (963k in Jul to Dec 2024, 174% increase)	
Events delivered	The Commissioner delivered around **15 external engagements, reaching an estimated 3,000 attendees (in-person and online).** *Source: JIRA Commissioner Engagement Reporting*	
Awareness activities	From 15 May to 17 July 2025 eSafety delivered an **awareness campaign** aimed at **young people aged between 18 to 24** years and their support networks, to raise awareness of **tech-based coercive control**, to educate them on specific behaviours that form part of a broader pattern of control and to empower them to take action. The channels included paid search, Meta, TikTok, programmatic display, and a paid partnership with The Daily Aus. Key results: - Reached approximately 824k - 712k video ad views - 16.5k unique users to eSafety's campaign landing page. From 6 July to 21 August 2025 eSafety executed a campaign aimed at **parents and carers** to drive registrations to our **free webinar program**. The channels included paid search, Meta, and programmatic display. Key results: - 1.6 million unique reach - 55k users to the webinar pages - Over 4.5k registrations. From 22 October to 28 November 2025 eSafety delivered an **awareness campaign** aimed at **young people aged 14-18**, to raise awareness of the subtle and often hidden **tech-based abuse tactics used by abusive caregivers** against children and young people, and to guide them to the eSafety website to find support resources. Channels included gaming platforms, digital display, Meta, Snapchat, TikTok and Spotify. Key results: - Reached approximately 1 million - Over 53k users to the campaign landing page. From the period 1 July to 30 November 2025, eSafety's paid search campaigns have driven over 140k users to our website across 6 keyword campaigns: - 'Social media ban': approx 78k - eSafety brand terms: approx 55k - Domestic violence: approx 2k - Reporting online harms: approx 1.6k - Parents and online safety: approx 1.4k - Online safety issues: approx 700. Between 1 July and 31 December 2025, eSafety supported the Dept's 'For the good of' education campaign to support the launch of the Social Media Minimum Age legislation. Although the Dept managed all aspects of the paid campaign, eSafety drove traffic to the website hub through email and organic social.	

	Key results: - Over 25k users and over 50k page views from eSafety emails - Over 25k users and over 37.5k page views from eSafety organic social posts.

Education Prevention and Communities- training attendees						
	Gender and tech	**Primary school children**	**Educators**	**Parents and carers**	**Seniors**	**Support services and broader community**
Attendees at webinars/ presentations	2,633	124,365	26,016	5,111	4,596	8,784
Audiences include:	*Domestic, family and sexual violence workers and other gendered presentations including Social Media Self-Defence*	*Virtual Classrooms for Primary school aged children (years 3-4)*	*National Student Wellbeing Officer professional learning program, teachers and support staff, Pre-service teachers, tertiary sector.*	*Parents and carers*	*Senior Australians as part of the BeConnected Program*	*Health, wellbeing and support services government agencies, corporate audiences, sporting and other community groups (eg disability, First Nations, LGBTIQ+)*

Trusted eSafety Providers				
	Students	**Parents & Carers**	**Educators**	**Total audience**
Reach	1,374,726	34,400	53,902	1,625,668

Corporate top line stats – Jason Armstrong				
Staffing numbers at 31 December 2025	APS FTE	241	Under the Strategic Commissioning Framework, eSafety has converted approximately **27** positions from contractors to ASL [24-25 and 25-26]	
	Contractors FTE	27		
	Total	**268**		
Budget As per the 2025-26 PBS	Departmental	$59.4m		
	Administered	$2.5m		
	Total	**$61.9m**		
SMAA funding received	2024-25	2025-26	2026-27	2027-28
	$3.8m	**$16.2m**	**$13.3m**	**$12.4m**
External legal expenditure at 31 December 2025	2024-25 FY	2025-26 YTD	2025-26 expense excludes costs paid to X Corp re Wakeley case of $0.624m.	
	$1.018m	**$1.419m**		

Travel expenditure at 31 December 2025	Domestic travel 2024-25 FY	Domestic travel 2025-26 YTD	Overseas travel 2024-25 FY	Overseas travel 2025-26 YTD
	$0.670m	**$0.311m**	**$0.349m**	**$0.152m**

AAT results – selected examples			
Age Verification technologies – 27 vendors participated, TRL range 8-9.			
Vendor / Technology name	**Technology type**	**TRL**	**Comments**
Austroads	Enables verified credentials and selective age attestations via government issued ID (Driver's licences)	8-9	Austroads manages the National Exchange of Vehicle and Driver Information System (NEVDIS) and is working on Mobile Driver Licence (mDL) standards. Relying party queries a government API or wallet service for a "match/no match" check - no DOB is returned
AgeChecked	Uses data matching, credit reference checks, electoral rolls, facial age estimation, document verification with liveness detection and adult-linked credit card checks in a cascading process	9	
Australian Payments Plus - ConnectID	Relies on banks' KYC-verified date of birth to return yes/no age assertions;	9	Peer to peer exchange model ensures that no Personal Identity information is ever visible to ConnectID; only consented minimal data (e.g. 'over 18') is shared.
LexisNexis - IDVerse	Uses AI for real-time age verification via biometric face matching, liveness detection and OCR	9	
Luciditi	System includes facial age estimation, document verification via selfie-ID match, NFC passport reading and open banking or telco records, with fallback to a reusable digital ID app.	9	Verified users across 13+, 16+ and 18+ thresholds with high accuracy.
Age Estimation technologies – 13 vendors participated, TRL range 7-9			
Yoti	Uses facial age estimation either on-device or at the edge.	9	Certified for privacy and security compliance, GDPR aligned; does not store images. Mean Absolute Error (MAE) values under 2 years for ages 13–20. 80% of users reported being either satisfied or very satisfied with the experience.

IDMission	Facial age estimation	8	Lab and school testing demonstrated MAE above 4.5 indicated low precision and high false positive rates. Trial suggested further validation needed before commercial deployment. **NB**: Demonstrates that the accuracy varies by vendor – not just by technology type.
Persona	Uses facial estimation with optional ID fallback	8	MAE ranged from 0.86 to 3.31 across different cohorts
Needemand - BorderAge	Uses hand gesture dynamics, captured via a device's camera; does not use biometric or identity data	8	Gesture based preferred by many mystery shoppers due to non-invasive feeling.
Age Inference technologies – 9 vendors participated, TRL range provided 5-8			
Verifymy	Age inference using metadata, email domains and app interaction context	7	While operational, its age inference capability is narrower and more targeted (e.g. "likely under 13") and evidence suggests it is part of layered access control.
AgeChecked	Integrated with payment processors to test users' ability to initiate a zero-dollar authenticated credit card transaction, with verification via one-time passcode.	*	Since credit cards are legally restricted to those 16+ in Australia (as additional cardholders), a successful result offered a medium confidence, binary signal of being over 16 at least.
Yoti	Behavioural and contextual age inference – eg. account metadata (age of account, frequency of use), content engagement patterns, device and browser context.	*	Triggers fallback to facial age estimation where confidence in inference result is low.

** Yoti and AgeChecked's inference products not given separate TRL to vendors bundled solutions.*

All systems assessed/tested in the trial were TRL 7 and above. Some emerging approaches (TRL <7 where discussed, but unable to be tested).

TRL 9	**System Proven and Ready for Full Commercial Deployment**: Actual system proven through successful operation in an operating environment, ready for full commercial deployment.
TRL 8	**System Incorporated in Commercial Design**: Actual system/process completed and qualified through test and demonstration (pre-commercial demonstration)
TRL 7	**Integrated Pilot System Demonstrated**: System/process prototype demonstration in an operational environment (integrated pilot system level)