IJM Submission On:

Draft Industry Codes of Practice for the Online Industry

Table of Contents

International Justice Mission (IJM) welcomes this opportunity to provide a formal Submission on the draft Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B) Material, Phase 1 under the *Online Safety Act 2021*, jointly prepared by IJM Australia and IJM's Center to End Online Sexual Exploitation of Children.[1] We commend industry associations for detailing measures in the Codes by which digital service providers can proactively detect and remove the most harmful online content and take greater responsibility to ensure a safer online environment.

In brief, IJM recommends that:

- Providers of online platforms and services be required to use technological tools to detect not only known CSAM, but also first-generation CSAM and livestreamed CSAM.
- Providers of encrypted electronic services be required to use technological tools and behavioural indicators to detect CSAM before it enters the encrypted space.
- The digital industry tangibly support through their policies, tools, and rules the privacy and security of victims and survivors to create a safer online environment for all.

[1] https://osec.ijm.org/; https://www.linkedin.com/company/ijmendosec/

I. Background and Call to Action on Livestreaming Child Sexual Abuse and Exploitation.

a. IJM Has Deep, On-the-ground Expertise in Combating Online Sexual Exploitation of Children Committed by Australian Offenders.

Since 2011, IJM has worked closely with all levels of the Philippine Government, international law enforcement, community service organisations, survivor leaders, and other relevant stakeholders to combat online sexual exploitation of children (OSEC), with focus on the trafficking of children to produce first-generation child sexual exploitation material (CSEM) especially via livestreaming video. This form of child sexual abuse online, along with "self-generated" abuse in livestreams, are all live crime scenes happening on tech platforms.

To date, IJM has supported 274 law enforcement operations, safeguarding 1012 victims or at-risk individuals, leading to the arrest of 314 suspects and conviction of 153 offenders. Leveraging IJM Philippines' promising practices in combatting the livestreaming of child sexual abuse, IJM's Center to End Online Sexual Exploitation of Children launched in November 2021, seeking to strengthen the global response to this crime. IJM partners with the Philippine Internet Crimes Against Children Center, a cooperation between Philippine and foreign law enforcement, including the Australian Federal Police.[2]

Livestreamed child sexual abuse requires urgent attention by tech platforms because it involves repeated hands-on sexual abuse of predominantly pre-pubescent children by trusted adults in real-time as directed and paid for by foreign sex offenders. Hiding behind their screens, many Australians direct and pay for the sexual abuse of young children in livestreams on popular video chat apps.[3] One study found that 18% of online sexual exploitation cases in the Philippines were initiated by Australia-based offenders.[4] CSAM is also produced and distributed live through grooming of children directly by Australian and other offenders online.

CSAM livestreamed in video calls allow Australian offenders to produce child sexual abuse material of children anywhere in real-time, with less digital evidence than image- or video-based CSAM distribution. Detection, reporting, and technological prevention of this type of online abuse is critical because the victims are being repeatedly abused "live." IJM's 2020 study of livestreamed child sexual abuse in the Philippines found that victims were abused on average for two years prior to intervention, in part because technology and financial sector companies failed to detect and report in real-time the crimes happening on and through their platforms.[5] Greater investment is needed to develop and broadly implement appropriate technology to safeguard the privacy and safety of online users and children and to further platform trust and integrity. Detection methods commonly used (i.e. PhotoDNA, scanning of still images on platforms) do not detect livestreamed sexual abuse of children.

---

[2] Philippine Internet Crimes Against Children Center is a model for an enhanced global response against online sexual exploitation of children. PICACC is a cooperation among local and international law enforcement, namely the Philippine National Police's Women and Children Protection Center (PNP-WCPC), the National Bureau of Investigation's Anti-Human-Trafficking Division (NBI-AHTRAD), the Australian Federal Police, the United Kingdom National Crime Agency (UK NCA), and the National Police of the Netherlands; in partnership with non-government organization, International Justice Mission (IJM). https://osec.ijm.org/documents/12/rev_PICACC_2nd_Anniv._Magazine.ia.pdf

[3] AIC (2021). For example, a study by the Australian Institute of Criminology found that 256 Australians spent more than $1.3 million over 13 years to commission and watch livestreamed sexual abuse of Filipino children. https://www.aic.gov.au/sites/default/files/2021-10/ti639_live_streaming_of_child_sexual_abuse.pdf

[4] IJM (2020) *Online Sexual Exploitation of Children in the Philippines: Analysis and Recommendations for Governments, Industry, and Civil Society.*

[5] IJM (2020), p. 11.

Likewise, user reporting is often ineffectual as the majority of these children are abused not in the context of being platform end-users but rather as victims of other users, usually adults. To the extent the children are users, they find it difficult to report as their in-person abusers are often family members, as documented in IJM's 2020 study and by other experts.

b. Child Sexual Abuse Material Production and Distribution Via Livestreaming Is Growing, Requiring Industry Action Across Platforms and Codes.

Livestreamed class 1A online conduct and content present on online platforms operating in Australia by Australian end-users require urgent corporate attention, now, across platforms and codes. These are live crime scenes committed daily on online platforms.

IJM's direct casework experience on livestreaming is primarily in the Philippines—the global hotspot for victims of this cross-border crime being committed via globally accessible technology platforms. Recent cases and victims have also been identified across dozens of countries, including Romania,[6] Ghana,[7] and Thailand,[8] to name a few. And children anywhere are susceptible, as evidenced by the *Disrupting Harm*[9] country reports.

It is critical to note that as early as 2011, Australian children were also victims of child sexual abuse production and distribution via livestreaming.[10] Ten years ago, the Australian Federal Police (AFP) called it a "very concerning trend." And yet over time, occasioned by industry failures to detect, disrupt, or prevent this harm on online, this abuse has only grown and worsened. Last year in 2021, the AFP reported that:

> **Australian children as young as eight are being coerced into performing live-streamed sexual acts by online predators**, who often record and share the videos on the dark net and sexually extort victims into producing even more graphic content.[11] (emphasis added)

Evidence indicates that CSAM production and distribution in livestreaming have continued to grow and pose increasing risks for children across platforms and jurisdictions.

- According to the WeProtect Global Alliance's *2021 Global Threat Assessment*, "[l]ivestreaming is on the rise, enabled by connectivity and the availability of inexpensive streaming devices. It often manifests as a cross-border crime that demands a co-ordinated international response."[12]
- Europol warns that "livestreaming of child sexual abuse increased and became even more popular during the COVID-19 pandemic."[13]

---

[6] https://www.independent.co.uk/news/uk/crime/paedophiles-philippines-romania-national-crime-agency-b2112832.html

[7] https://www.nationalcrimeagency.gov.uk/news/registered-sex-offender-paid-to-watch-live-streamed-child-abuse

[8] DISRUPTING HARM IN THAILAND: Evidence on online child sexual exploitation and abuse, available at https://www.end-violence.org/sites/default/files/2022-02/DH_Thailand_ONLINE_final.pdf, p. 58 ("The victimisation of children via video calls is a common form of OCSEA, according to [the Thailand Internet Crimes Against Children task force] TICAC, and live-streaming of CSEA has appeared in the caseload of DSI. In addition, one foreign law enforcement agency notes that Thailand accounts for 5% of its total reports to date on live-streamed CSEA.")

[9] Disrupting harm (unicef-irc.org)

[10] Paedophiles are watching abuse of children live online, say police, https://www.smh.com.au/technology/paedophiles-are-watching-abuse-of-children-live-online-say-police-20120311-1uskh.html

[11] AFP warn about fast growing online child abuse trend, Sept. 2021, https://www.afp.gov.au/news-media/media-releases/afp-warn-about-fast-growing-online-child-abuse-trend

[12] WeProtect Global Alliance *2021 Global Threat Assessment*, p. 60. Global-Threat-Assessment-2021.pdf (weprotect.org)

[13] See pp. 6, 8, 34, 39, 41 of Europol (2021) Internet Organized Crime Threat Assessment 2020). https://www.europol.europa.eu/cms/sites/default/files/documents/internet_organised_crime_threat_assessment_iocta_2020.pdf

- INTERPOL also reports that "[l]ive-streaming of child sexual exploitation for payment has seen an increase in recent years," as demand surged during the pandemic as an alternative to 'in-person' abuse.[14]

Not only is livestreamed CSAM growing, but it contains some of the worst child sexual abuse acts, further demanding urgent tech action against this crime. Internet Watch Foundation (IWF) research on child sex abuse livestreaming reveals 98% of victims are 13 or under.[15] Forty percent of the livestream captures or recordings were classified by IWF as containing 'serious' sexual abuse, *with 18 percent involving the rape and sexual torture of children*. This is consistent with IJM's on-the-ground casework experience in the Philippines. In the over 270 cases IJM has worked on, the livestreamed abuse suffered by children at the behest of Australian and other offenders who watch on video calls is rarely limited to erotic displays: it usually includes forcible sexual penetration constituting rape in most jurisdictions. Children are forced to engage in sex acts with other children, sexually abused by an adult, and sometimes harmed in other degrading ways, such as in bestiality. IJM social workers and lawyers have journeyed with hundreds of survivors as they pursued healing and justice from these traumatic harms perpetrated both in person and online.

Australian online users likewise harm children in abusive and degrading ways in livestreams. In 2021, the Australian Institute of Criminology (AIC) analysed chat logs from seven now convicted Australian offenders who watched and directed 145 instances of sexual abuse of 74 children via livestream. The study "revealed the brutality of the abuse ordered by Australian offenders, which might otherwise never be known."[16] "The things that I've seen them request happen to these children is just awful," said AIC Principal Research Analyst Sarah Napier. Five of the seven offenders requested a victim of a specific age, with one offender inquiring about a four-year-old victim. In several cases, Australian online end-users using Australian online platforms, requested in written chats that children be tied, bound, beaten, whipped or otherwise subject to something that involves pain, or suffer bestiality.[17]

Moreover, this livestreamed abuse and new CSAM produced rarely remain with the receiving offender. Instead, as ECPAT International notes, recorded livestreaming may be "substantially adding to the volume of child sexual abuse materials…available on the web as a whole."[18] In other words, by detecting and disrupting or preventing livestreamed CSAM, industry is going upstream to prevent the increase in CSAM distributed online.

Adopting a preventive posture to livestreamed CSAM (i.e., using indicators, signals, client-side detection, or other AI-driven technology) allows industry to keep platforms and children safe, preserve the privacy of users and survivors, while creating friction and no place to hide for "bad actors" seeking to violate the human rights of children in flagrant disregard for platform terms of service.

Below are some examples of existing technology or actions taken by platforms in real-time to address livestreaming:

[14] Europol (2021) Serious and Organized Crime Threat Assessment 2021, socta2021_1.pdf (europa.eu)

[15] *See* https://www.iwf.org.uk/news-media/news/iwf-research-on-child-sex-abuse-live-streaming-reveals-98-of-victims-are-13-or-under/; Internet Watch Foundation 2018. Trends in child sexual exploitation: Examining the distribution of captures of live-streamed child sexual abuse. Cambridge, UK: Internet Watch Foundation. https:// www.iwf.org.uk/resources/research

[16] https://www.sbs.com.au/news/article/malone-was-sexually-abused-online-aged-eight-many-perpetrators-are-in-australia/xc8epp10a; Live streaming of child sexual abuse: An analysis of offender chat logs, available at https://www.aic.gov.au/sites/default/files/2021-10/ti639_live_streaming_of_child_sexual_abuse.pdf

[17] *Id*.

[18] Andrea Varrella, 'Live Streaming of Child Sexual Abuse: Background, Legislative Frameworks and the Experience of the Philippines' (2017) 12 ECPAT Journal 47, 49.

- The social livestreaming platform, Yubo, proactively screens live video to keep children safe online, implementing automated prompts to users to change behavior and disabling violative livestreams.[19]
- Safety technology company, SafeToNet, has created a real-time video & image threat detection technology, SafeToWatch, capable of determining whether visual data represents undesirable and illegal content such as pornography, sexually suggestive imagery, cartoon pornography, and/or CSAM.[20] The machine-learning algorithm will hash images with harmful content and render the content harmless.[21] SafeToNet can provide more information to the eSafety Commissioner or industry associations upon request.
- DragonflAI is a prevention and disruption tool that moderates livestreams completely on-device before they are streamed to platform. It detects indecent content such as CSAM and prevents content from being uploaded. It has been tested to detect 97.9% of nudity.[22]

Industry may know of other examples. Technology thus already exists that can be further refined and improved upon with proper investment from the very companies on which the abuse occurs. The commitment by industry to do so is referenced below and IJM encourages industry to set related, specific timelines and goals. The status quo, however, is simply unacceptable to children, survivors, governments, child protection organisations, and society as a whole. Video chat apps and livestreaming services cannot continue to be routinely misused to produce and distribute new child sexual abuse material with near impunity and minimal, if any, industry actions to stop these gross violations of human rights and their own terms of service. Even though the CSAM is produced and distributed using platform-provided video-chat services (so-called "private channels")[23] and thus invisible to most public users of the platform, high level corporate attention, policies, tools, and rules to detect, disrupt, and prevent this harm are critical to prevent platform end-users from abusing children.

These are live crime scenes. The response cannot continue to be sticking heads in the sand, while knowing such severe harm occurs at scale, or hiding behind the banner of "privacy" when providers are the ones who set user terms of service in the first instance or explaining that it is technologically difficult. Industry associations collectively possess the financial and intellectual resources and expertise, as well as intimate knowledge of their own platforms, to help solve this problem.

II. Key Strengths of the Draft Industry Codes

IJM applauds the industry associations for putting forward these draft Industry Codes to create a safer online environment for children and all users. The Codes show the commitment of the digital sector in collaborating with the eSafety Commissioner to engage in the difficult work of protecting children from online sexual abuse. These Codes are a step

[19] Yubo (n.d.) https://www.yubo.live/blog/real-time-intervention-on-social-video

[20] E&T (2020) https://eandt.theiet.org/content/articles/2020/08/ai-based-software-could-block-livestreamed-graphic-content/

[21] https://www.youtube.com/watch?v=2xwKfsSwv1I

[22] https://www.dragonflai.co/

[23] *Malone was sexually abused online aged eight. Many perpetrators are in Australia*, available at https://www.sbs.com.au/news/article/malone-was-sexually-abused-online-aged-eight-many-perpetrators-are-in-australia/xc8epp10a ("Offenders typically don't use the dark web. Instead, the crimes often occur in private messages on widely available video-chat platforms, including Skype, WhatsApp and Facebook Messenger."); *Live streaming of child sexual abuse: An analysis of offender chat logs*, available at https://www.aic.gov.au/sites/default/files/2021-10/ti639_live_streaming_of_child_sexual_abuse.pdf (identifying Skype, Viber, Facebook as used in the chatlogs of known Australia-based livestreaming offenders).

forward in creating a safer internet and putting practical solutions to child protection at the forefront of industry standards. The Codes include policies prohibiting class 1A material, encouraging social media services and relevant electronic services providers to detect and remove known child sexual abuse material. The removal of known CSAM will protect survivors from ongoing re-victimisation, as their images and videos are re-shared. The *Social Media Services Online Safety Code* and *Relevant Electronic Services Online Safety Code* have included a requirement for providers to engage NGOs, experts, academics, and researchers in their fight against online sexual abuse, which will only strengthen the responses of these companies. Finally, the ongoing investment that Tier 1 and Tier 2 companies will be providing in developing better detection tools and more equipped personnel has the potential to help advance global detection technologies and support widespread accountability to international tech companies to implement their own advanced detection technologies.

III. General Observations and Recommendations on the Industry Codes

IJM's comments, feedback and recommendations on the Codes focus on the online sexual exploitation of children. In this section, we provide observations and recommendations pertaining to the Codes, in general, and grouped according to six themes (A-F). Section IV provides specific recommendations on three of the individual industry codes.

A. Industry should require proactive detection of first-generation CSAM and livestreamed CSAM

A key Outcome in the Codes is the requirement for industry to take "Proactive steps to detect and prevent access or exposure to, distribution of, and online storage of class 1A material." The *Explanatory Memorandum* sets out the decision made in drafting the Codes to require industry to proactively detect and remove instances of known CSAM using technological tools but not to extend that requirement to the detection and removal of first-generation CSAM:

> Industry has also considered what it can do to detect first-generation child sexual abuse material. Some very large companies have invested in technology that can detect first-generation child sexual material (i.e., material not previously identified and stored in an appropriately maintained NGO database), however this technology is still in an early stage of development. While the accuracy of technology to enable detection of first-generation material is improving, it is generally accepted that it is not as accurate as technology for the detection of known CSAM and requires greater human review of detected materials. The proactive detection of online materials has therefore been limited to the detection of known child sexual abuse material. (p. 8)

The draft Codes and *Explanatory Memorandum* fail to specify the accuracy (by percentage or error rate) of technology to detect first-generation CSAM, although this is the stated reason to not include it. They similarly fail to explain what level of accuracy is sufficient for tech companies to use image and video classifiers (or other technology) to detect first-generation CSAM and what is unacceptable about that specific error rate; in other words, what is the rationale. For instance, the draft Codes do not explain how a user's privacy is impacted by AI-powered tools lawfully trained on actual images of child sexual abuse and implemented to only identify such contraband. Industry has not, for example, argued that if law enforcement receives a report of suspected CSAM that turns out to not be CSAM, that any enforcement action would occur against a user for possessing or sharing a legal photo.

To the contrary, industry readily explains that "very large companies have invested in technology that can detect first-generation child sexual material (i.e., material not previously identified and stored in an appropriately maintained NGO database)." And that the "accuracy of technology to enable detection of first-generation material is improving,. ." Evidence indicates that such technology is not only improving but *already effective* in detecting first-generation CSAM.

For instance, in her keynote address at the September 21-22, 2022 InHope Summit, Michelle DeLaune, CEO of the National Center for Missing & Exploited Children (NCMEC), reported that 17% of the 5 million images and videos received by NCMEC in July 2022 were new to NCMEC; in other words, first-generation abuse not previously identified and hashed. Alarmingly, that means *800,000 images and videos of first-generation CSAM were received by NCMEC in one month alone.* And apparently only by those companies currently detecting first-generation CSAM. Not only does that provide a small window into the scale of the problem of CSAM production, and therefore urgency to respond, but it also proves that tech companies can and, in fact, do detect and report first-generation CSAM. So why would Australian industry associations not want to do the same? Industry should collectively "level up" its protection of children online, not standardise the bare minimum in child safety.

The omission of first-generation CSAM from the requirement has significant negative ramifications for protecting children from online abusers. The aim of these Codes to protect children from violence can only be realised with inclusion of first-generation CSAM detection. To the extent that improvements in technology are needed, industry, through the 28-member Technology Coalition, has already publicly committed to doing so. In the words of Sean Litton, Tech Coalition Executive Director in their first annual report:

> **We are resolved to drive forward the improvements in technology and systems that will ultimately eradicate the online sexual abuse and exploitation of children on our platforms**.[24]

Industry can only eradicate online sexual abuse and exploitation of children by addressing first-generation CSAM (which in one month of CSAM received by NCMEC, accounted for 17%), and livestreamed CSAM, as discussed above. CSAM that has not previously been detected, which can sometimes accompany livestreamed child sexual abuse, can indicate new production or ongoing child sexual abuse taking place on online platforms. Victims of ongoing or recent CSAM production urgently need to be identified so that appropriate authorities can safeguard them from situations where there is heightened risk of repeated and future abuse and exploitation.

As noted in the *Explanatory Memorandum*, technology already exists to detect first-generation CSAM, as major tech companies routinely do so. While the draft Codes say that such technology is not as accurate as technology to detect known CSAM, industry has not said how accurate it is, nor how accurate it would need to be. The public deserves to know and children deserve everyone's best efforts to protect them from online sexual abuse.

Many actions that companies take to protect users or the public are imperfect. For instance, technology to detect spam sometimes inaccurately sends legitimate emails to the spam folder. Everyone has had the experience of their bank or financial institution freezing their credit or debit card for "suspicious activity" that actually was the user's own activity. We accept these inconveniences and frustrations to protect our emails from spam and our finances from theft because we know technology is not perfect and the harm prevented is worth it. Accepting some level of error in first-generation CSAM detection for the sake of protecting untold numbers of children from sexual abuse seems is the right balance, but such

[24] https://www.technologycoalition.org/annual-report

a conversation has not even happened because of a lack of transparency on the accuracy of such technology.

> **Recommendation:** Include a new compliance measure in the Social Media Services Online Industry Code, the Relevant Electronic Services Online Industry Code and the Designated Internet Services Online Industry Code that requires providers to detect, remove and report first-generation CSAM and livestreamed CSAM.

Survivor Feedback on first-generation CSAM and livestreamed CSAM detection:

Three adult survivors of online sexual exploitation as children, Joy*, Liberty*, and Ruby*, have provided feedback on these industry codes. A trained IJM social services professional with over 10 years of experienced sought consultation from these survivors, asking them these questions, which elicited the following answers:

1. Do you agree that industry should use technology to detect CSAM (Child Sexual Abuse Material) that has not been seen before (i.e., first-generation or "new" CSAM)?

2. Do you agree that industry should use technology to detect CSAM in livestreaming?

Joy*: "I think there should be a technology that will detect CSAM. Because in my experience, I was abused when I was still young but I was only rescued after several years after the abuse. It is better that children will be rescued earlier by early detection. With early detection, there will be less children that will be further abuse if perpetrators are detected or arrested early on. Foreigner pedophiles must also be detected and stopped early on because they create the demand for CSAM both on the production and livestreaming."

Ruby*: "It is really important to be vigilant and people who have power, to invest to create a system that will ensure safer community online for children. Most specially on present times that children are using online gadgets. There's a child that I personally knew whom I caught using a phone. What's worst was there a foreigner stranger that was waving on her and trying to communicate with her. It frightened me how these 'stranger's can easily access children online. This ignite my desire to advocate for safety of children online. To create a technology that will detect CSAM online and protect children from harm specially in livestreaming."

Liberty*: "I agree that there shall be a technology that will detect CSAM. To prevent these CSAM from being shared most specially in livestreaming. There's a platform I know that detects if there are 'harmful' material online, if other online platform is able to do it, I believe it can really be reported."

The recommendations and feedback of individuals with lived experience are critical for industry to hear since, after all, they are the ones most directly harmed by industry policy decisions and actions. These survivors, already safe, speak out for the sake of all children still being abused or at risk of online sexual abuse.

B. Extend requirement to use technological tools to proactively detect known CSAM beyond the Social Media Services Code and Designated Internet Services Code.

The minimum compliance measure requiring the use of technological tools to detect and remove known CSAM under the *Social Media Services Online Safety Code* [Compliance Measure #8, p. 10] and the *Designated Internet Services Online Safety Code* [Compliance

Measure #6, p. 9-10] is critical to protecting children from ongoing harm and victimisation. This measure should also be extended to Relevant Electronic Service providers, as a minimum compliance measure. One survivor, Ruby* notes that

> most online sexual exploitation of children occurs on mainstream tech platforms. We need tech companies to urgently prioritise the detection of this content – particularly in its most hidden forms, like livestreaming.
> We need governments in demand-side countries to be part of the solution.[25]

A 2022 study of online exploitation of children in the Philippines, *Disrupting Harm*, found that among children who experienced online sexual exploitation on social media, "Facebook or Facebook Messenger were the most common platforms where this occurred, accounting for over 90% of cases. Other platforms cited, to a much lesser degree, were TikTok, Twitter, Instagram, and Snapchat."[26] Note that under the draft Industry Codes, private messaging services would be covered under the *Relevant Electronic Services Online Safety Code*.

In the words of a Filipino survivor of online sexual exploitation,

> I am one of the survivors who is ready to talk about our experiences, I want to legislate the cessation of online sexual exploitation such as livestreams on Facebook or any app. I do not want women to experience more and even men get that kind of abuse because it's not a joke. I am asking for help so that I can process how to stop this abuse. Everyone needs protection. Every person performing this abuse must be stopped or monitored. – Diana*, 20-year-old survivor (13 years old at the time of abuse[27].

This statement reflects how survivors of online sexual exploitation view their harm and how they want service providers to respond to these gross violations of human dignity and platform terms of service. Detection, removal, and reporting should be part of the minimum compliance for Tier 1 and Tier 2 relevant electronic service providers, both to respond to the survivors of online sexual exploitation and to protect more children from suffering this form of abuse on your platforms.

> **Recommendation:** Compliance measure 9 (Use of technological tools to detect and remove known CSAM) under the *Relevant Electronic Services Online Safety Code* should not be an optional measure. As a minimum compliance measure, require Tier 1 relevant electronic service providers to use technological tools to detect and remove known CSAM.

C. Require ongoing investment in tools and personnel to detect and respond to Class 1A material, especially detection of first-generation CSAM and livestreamed CSAM.

The *Social Media Services Online Safety Code* (Compliance Measure #9) requires Tier 1 providers, as a minimum compliance measure, to:

> make ongoing investments in tools (for example, using hashing, machine learning, artificial intelligence, or other safety technologies) and personnel that support the capacity of the provider to detect, and take enforcement action concerning class 1A material, proportional to

[25] https://news.trust.org/item/20210922121204-xq3l0

*Pseudonym

[26] DH_Philippines_ONLINE_FINAL.pdf (end-violence.org)

[27] Alice, et al. (2022) https://www.ijmuk.org/stories/survivor-letter-to-uk-government-online-safety-bill

the incidence of class 1A material on the service and the extent class 1A materials are accessible to Australian end-users.

The *Designated Internet Services Online Safety Code* (Compliance Measure #7) has a similar requirement for Tier 1 providers to make ongoing investment in tools and personnel with respect to known class 1A material. These are measures are critical in order to keep abreast of continually evolving technologies and the scale of online harm and the sophistication of offenders when perpetuating such harm. These compliance measures should be strengthened by including an explicit requirement to invest in developing innovative technologies – such as "image and video classifiers" – to detect first-generation CSAM, including via client-side technology, in live video streams and stored on or transmitted through service providers' servers.

While PhotoDNA and similar technologies are widely used by service providers to identify known CSAM, they are not designed to find first-generation content, much less spot the livestreaming of these abuses. Where tools are used to detect and block the distribution of known CSAM, survivors will be protected from the ongoing distribution of illegal images and videos depicting their sexual abuse and exploitation. Where image classifiers or other innovative technologies are used to detect first-generation CSAM, children will be protected as perpetrators' attempts to memorialise and even sell abusive acts will be disrupted and frustrated by devices incapable of recording or capturing illegal, abusive crime scene.

Innovative detection technologies must be developed and deployed to identify and safeguard potentially thousands of children in urgent need of protection. This is especially true for technologies that can detect abusive livestreaming and other first-generation CSAM before they enter private messaging and encrypted services. Some examples of tools with this capability already currently exist *[see examples listed under Heading I.B above and Heading III.D, below]*.

For instance, Dr. Hany Farid, Professor at University of California, Berkley, who pioneered PhotoDNA, has shared about the accuracy and reliability of end-to-end encryption scanning software that is already in existence and scans malware and viruses[28]. This is an analogy to technology to include CSAM detection and maintain the same level of privacy that users currently enjoy. One example is Apple, which implemented client-side scanning that prevents the automatic viewing and sending of CSAM by encouraging either the child to seek help from a trusted adult or the sender to use a helpline for their at-risk behaviour.[29] *[See other examples listed under Heading D, below]*. As such, the requirement for ongoing investment in detection tools and personnel should also be a minimum compliance measure for Relevant Electronic Services providers. *[See Heading D, below, for recommendation.]*

> **Recommendation:** Include in Minimum Compliance Measure #9 under the *Social Media Services Online Safety Code* and under #7 under the *Designated Internet Services Online Safety Code*, as part of ongoing investment in tools, an explicit requirement to invest in developing innovative technologies to detect first-generation CSAM and livestreamed CSAM.
>
> This may include technology to detect and prevent/disrupt the production and distribution of the contact, not only technology to detect and report post-harm. Collaboration with safety tech companies should be encouraged.

[28] Paul Tang (2021) https://www.paultang.nl/en/event-csam/

[29] https://www.apple.com/child-safety/

D. Require encrypted services to use tools to detect behavioural signals, develop and use technological tools to detect class 1A material, and actively enforce violations of its policies.

The *Relevant Electronic Services Online Safety Code* provides only one measure promoting the use of technological tools by encrypted service providers to detect known CSAM on their platforms, and only as an optional measure (Compliance Measure #9, p. 14). This represents a significant gap in stopping the use of encrypted platforms for creation and distribution of CSAM and other class 1A material. Instead, the Code should require encrypted service providers, as a minimum compliance measure, to:

- Use technological tools designed to detect behavioural signals associated with the distribution of CSAM; and
- Make ongoing investments in tools and personnel that support the capacity of the provider to detect and take enforcement action concerning class 1A material.

One example of this at work is through Cyacomb Safety, a detection technology for end-to-end encryption that protects personal privacy while anonymously matching and detecting known CSAM with shared user content.[30]

Cyacomb CEO Ian Stevenson explains:

> As firm believers in privacy, it was important for us to develop a solution to the growing problem of online child sexual abuse, whilst respecting the right to privacy of users on social media and online messaging platforms. Our Contraband Filter technology is built on principles of privacy-by-design and we have received written advice from the ICO's Innovation Hub on our proof of concept work during the Safety Tech Challenge Fund. We believe there are no fundamental data protection barriers to deployment. I'm immensely proud of the team here at Cyacomb, especially their overriding determination to overcome the hurdle that it is technologically impossible to do anything to reduce the problem of online child sexual exploitation within an end-to-end encrypted messaging environment.[31]

Consistent with Cyacomb, UK government's Cyber Security experts have developed a paper endorsing client-side scanning and confirming that this can be done without breaking end-to-end encryption for services that have this enabled.[32] Privacy preserving client-side image detection technologies have been developed by not only Cyacomb but also SafeToNet[33], DragonflAI[34] and Apple.[35] Some of these technologies detect CSAM before it can enter an end-to-end encrypted environment, thus preserving both safety and privacy.[36]

---

[30] Cyacomb (2022) https://www.cyacomb.com/company/news/2022/september/first-line-of-defence-cyacomb-launches-online-safety-software-to-combat-child-sexual-abuse-whilst-protecting-privacy/?utm_source=ActiveCampaign&utm_medium=email&utm_content=News+and+opportunities+from+across+the+Alliance&utm_campaign=September+2022+Newsletter

[31] https://www.cyacomb.com/company/news/2022/september/first-line-of-defence-cyacomb-launches-online-safety-software-to-combat-child-sexual-abuse-whilst-protecting-privacy/?utm_source=ActiveCampaign&utm_medium=email&utm_content=News+and+opportunities+from+a

[32] Ian Levy, Crispin Robinson (2022) https://arxiv.org/abs/2207.09506

[33] https://www.youtube.com/watch?v=2xwKfsSwv1I

[34] https://www.dragonflai.co/

[35] https://www.apple.com/child-safety/

[36] A Positive Step for Child Protection and Privacy (IJM), https://www.ijm.org/news/positive-step-child-protection-privacy

Detecting Behavioural Signals

One tool for detecting behavioural signals is IJM's *Tech and Financial Sector Indicators of Livestreaming Online Sexual Exploitation of Children*, which lays out specific language, behaviours, and other signals indicative of the production of new CSAM, including via livestreamed video. These indicators reveal actions that are often not illegal on the surface, but when combined with each other, reveal a high likelihood of this form of abuse against children occurring. This tool is available upon request to endosec@ijm.org.

> **Recommendation:** Compliance measure 10 (Use of technological tools to detect behavioural signals associated with CSEM and pro-terror material) under the *Relevant Electronic Services Online Safety Code* should not be optional but should be a minimum compliance measure for all encrypted relevant electronic service providers.

Ongoing investments in detection tools:

Encrypted services providers should be required to explore, develop, and test tools that allow for the identification of CSAM even within applications and messaging services that utilise end-to-end encryption. Encryption is one tool to facilitate users', including children's, privacy. However, without appropriate tools that allow for scanning of messaging and file sharing in encrypted environments, either prior to encryption or through another process, a significant portion of CSAM will remain undetected.

For example, in 2019, when Meta (then Facebook) announced a plan to adopt end-to-end encryption throughout its platforms, the U.S. National Center for Missing and Exploited Children (NCMEC) estimated that the plan would result in loss of up to 70% of Meta's (then Facebook's) annual reports. With Meta continuing to be the leader in detecting and reporting CSAM found on its platforms, that would represent the vast majority of reports made to NCMEC each year.

Use and refinement of existing tools described above and others could support the continued detection of first-generation CSAM and livestreamed CSAM, while still allowing for the implementation of privacy protections.

> **Recommendation:** Include a new compliance measure under the *Relevant Electronic Services Online Safety Code* that requires relevant electronic service providers, including encrypted service providers, to make ongoing investments in tools and personnel that support the capacity of the provider to detect and take enforcement action concerning class 1A material and make explicit requirement to invest in developing innovative technologies to detect first-generation and live-streamed CSAM.

E. Support the privacy and security of victims and survivors.

The Head Terms, under 6. Limitations and lawful conduct, allows for privacy considerations to limit a service provider's adoption of a particular compliance measure:

> *Note: In considering whether it would be reasonable for an industry participant to adopt a particular compliance measure under this Code, it will be relevant for the industry participant to take into account the desirability of not intruding upon, and otherwise maintaining the privacy and integrity of, private communications between end-users. However, where indicated in the Schedule, it may be appropriate for an industry participant to adopt measures that involve analysis of behavioural signals and other data or trends to prevent, detect and address harmful activity. (emphasis added; p. 13)*

The *Explanatory Memorandum* also raises privacy and security considerations in stating:

> The industry considered whether the Code should include measures that would require providers to proactively monitor or scan users' private file storage and private communications (for example, emails and text messages). The industry concluded that the extension of proactive detection measures could have a negative impact on the privacy and security of end-users of private communications and file storage services, including services used by businesses and government enterprises. (p. 8)

Detection tools can be deployed without compromising privacy

As noted above, companies already deploy virus and malware detection technology in emails and that has never compromised user privacy. User privacy cannot be relied upon as a blanket argument against proactive scanning by industry service providers. The technology exists – and can be further refined by industry – to scan for and remove illegal content without infringing on the personal privacy of users. Cyacomb's technology, referenced above, scans the movement of content across domain borders without interfering with encryption, thus maintaining user privacy.[37] For over a decade, major tech companies have been using PhotoDNA to identify known CSAM through hashing, without compromising the privacy of non-contraband content.[38] Even to the extent that some child safety actions may implicate user privacy, ECPAT International's study in Europe found "76% of adults have indicated a willingness to allow automated technology tools that specifically scan for and detect child sexual abuse material online – even if this means giving up some of their privacy. Most agree that regulating online spaces with the best interest of children is essential to ensuring their safety online."[39]

With the increase of client-side scanning and accurate detection technology, the error rate of detection technology decreases and individuals maintain their privacy online while service providers foster safer online environments. When it comes to ensuring privacy, Dr. Hany Farid appropriately explains in a 2022 podcast with IWF:

> We routinely scan on our devices, on our email, on our cloud services for everything including spam and malware and viruses and ransomware and we do that willingly because it protects us. It protects our devices and, without that, without the ability even within end-to-end encryption, to scan for harmful content to our devices, we would be dead in the water. I don't think it is hyperbolic to say that, if we are willing to protect ourselves, then we should be willing to protect the most vulnerable among us. It is the same basic core technology, and I reject those that say this is somehow giving something up. I would argue this is, in fact, exactly the balance that we should have in order to protect children online and protect our privacy and our rights.[40]

Allowing offenders to create and share CSAM online constitutes a gross violation of the privacy and safety rights of children. The abuse that was committed against the child to create the material in the first place is a clear violation of bodily integrity. Each time that depiction of abuse is shared and viewed constitutes a re-victimisation of the survivor. On top of that, the sharing and viewing of these materials is a severe breach of the privacy of the child, who has not consented to the images being created, let alone distributed.

[37] https://www.weforum.org/agenda/2022/09/technology-shaping-the-future-of-digital-safety/
[38] https://www.salesforce.org/blog/international-justice-mission-data-privacy/
[39] ECPAT International (2022) https://ecpat.org/project-beacon
[40] Dr. Hany Farid (2022) https://www.iwf.org.uk/news-media/blogs/encryption-vs-privacy-in-conversation-with-professor-hany-farid/

Privacy, security and safety are not mutually exclusive

We affirm the eSafety Commissioner's comments that privacy, security and safety are not mutually exclusive but rather "mutually reinforcing".[41] Investing in advanced scanning technologies is the right way to manage the healthy tensions between these three imperatives. Further investment can be done by affording legal opportunities for tech companies to train AI on CSAM datasets to improve its accuracy through the retention of identified CSAM or collaboration with agencies.

In fact, our society already widely accepts the following measures precisely for online safety:

(a) Virus scanning on computer files and hard drives
(b) Biometric scanning for identify verification, eg. Apple's FaceID
(c) Profile verification on online dating and gaming platforms

F. Ensure a safer online environment for everyone.

Objective 1 in the Codes

The draft Codes explicitly limit the goal of providing a safer online environment to a safer environment for Australian end-users. Objective 1 in each of the Codes is "Industry participants will take reasonable proactive steps to create and maintain a safe online environment for end-users in Australia." This is in contrast to the eSafety Commissioner's *Position Paper on Code Development* which does not include the last four words, setting out the objective as "Industry participants will take reasonable proactive steps to create and maintain a safe online environment." (p. 66)

Victims of online harms and internet misuse on Australian platforms are not confined to Australian end-users; many Australians are involved in exploiting and causing online harm to others outside of Australia. For example, IJM's 2020 Study on the Online Sexual Exploitation of Children found that Australians accounted for nearly 1 in 5 offenders who engaged in livestreamed sexual abuse of children in the Philippines. None of the child victims were Australian end-users, yet online platforms available in and used by Australians were weaponised for that harm. The compliance measures in the Codes should protect children globally from online harm committed through the use of Australian digital platforms or committed by Australia-based users.

Enforcement of policies concerning class 1A and 1B material; reporting and complaints mechanisms

The *Social Media Services Online Safety Code*, the *Relevant Electronic Services Online Safety Code* and *Delegated Internet Services Online Safety Code* all contain minimum compliance measures relating to systems, processes and technologies that enable the provider to take appropriate actions for enforcement of policies prohibiting class 1A and 1B material, a requirement to provide clear reporting and complaint mechanisms about class 1A and 1B material available on the service and steps for providers to respond effectively to complaints about class 1A and 1B material.

The clear guidelines and steps in these provisions are critical to providing a safer online environment; however, these provisions entitle only Australian end-users to request and receive support from providers in removing CSAM depicting themselves or someone else. The children depicted in CSAM on Australian platforms who are not themselves Australian

[41] https://www.weforum.org/agenda/2022/09/technology-shaping-the-future-of-digital-safety/

end-users are equally deserving of assistance in regaining their privacy and the removal of these materials. For example, the U.S. National Center Missing and Exploited Children made available to the Philippine Government 3.2 million CyberTipline reports in 2021. Of the 193 rescue operations conducted by the Philippines Internet Crimes Against Children Center between February 2019 and July 2022, 35% of offenders arrested or charged resulted from referrals from Australian-based investigations or intelligence leads.[42] Those reports related to all forms of online sexual exploitation of children, including the distribution of child sexual abuse materials (photos, videos) and other forms of abuse such as sextortion, online enticement or grooming and livestreaming abuse. Under the existing language, these children outside of Australia would be unable to request assistance for the removal of these materials, as they are unable to make reports. See Annex A for exemplary stories of Australian offenders convicted of livestreamed child sexual abuse production and distribution.

> **Recommendation:** For compliance measures relating to systems, processes and technologies that enable a provider to respond to policy violations of policies prohibiting class 1A and class 1B material, and measures under Outcome 8 (Industry participants provide clear and effective reporting and complaints mechanisms for class 1A and class 1B material) and Outcome 9 (Industry participants effectively respond to reports and complaints about class 1A and class 1B material), ensure that enforcement action and service provider responses to reports are not limited to Australian end-users and Australian reporters of class 1A and class 1B material.

[42] Adelaide man jailed and 15 young victims rescued after international child abuse investigation | Australian Federal Police (afp.gov.au)

IV. Recommendations for Individual Codes

This section looks at the *Social Media Services Online Safety Code*, the *Relevant Electronic Services Online Safety Code* and the *Designated Internet Services Online Safety Code* and provides recommendations on relevant compliance measures under each Code, organised by theme. The comments make reference to observations in the corresponding paragraphs in Section III (General Observations and Recommendations).

Social Media Services Online Safety Code

<table>
<tr><th colspan="2">Detection of first-generation CSAM:</th></tr>
<tr><td>No provision in draft Code</td><td>IJM Comments
See Section III.A.</td></tr>
<tr><td colspan="2">Recommendation: Include a new compliance measure that requires providers of social media services to detect, remove and report first-generation CSAM and livestreamed CSAM.</td></tr>
<tr><th colspan="2">Ongoing investment in detection tools and personnel:</th></tr>
<tr><td>Provision in draft Code
9) Ongoing investment in tools and personnel by Tier 1 social media services
A provider of a Tier 1 social media service must make ongoing investments in tools (for example, using hashing, machine learning, artificial intelligence, or other safety technologies) and personnel that support the capacity of the provider to detect, and take enforcement action concerning class 1A material, proportional to the incidence of class 1A material on the service and the extent class 1A materials are accessible to Australian end-users.</td><td>IJM Comments
See Section III.C.</td></tr>
<tr><td colspan="2">IJM Recommendation: Include, as part of the ongoing investment in tools, an explicit requirement to invest in developing innovative technologies to detect first-generation CSAM and livestreamed CSAM.

This may include technology to detect and prevent/disrupt the production and distribution of the contact, not only technology to detect and report post-harm. Collaboration with safety tech companies and organisations like Thorn should be encouraged.</td></tr>
<tr><th colspan="2">Enforcement of policies; reporting and complaint mechanisms</th></tr>
<tr><td>Provision in draft Code
2) Systems, processes and technologies for enforcement of policies prohibiting class 1A material
A provider of a Tier 1 or Tier 2 social media service must implement systems, processes and technologies that enable the provider to take appropriate enforcement action against end-users who violate terms and conditions, community standards, and/or acceptable use policies prohibiting class 1A material. At a minimum, social media service providers must have standard operating procedures that:</td><td>IJM Comments
See Section III.F.</td></tr>
</table>

a) specify the role of personnel in reviewing and responding to reports of class 1A materials by Australian end-users (more detail under measure 4); b) include clear internal channels for personnel in escalating, prioritising and assessing reports of class 1A material by Australian end-users; and c) provide operational guidance to personnel in relation to steps that should be taken when the service receives reports of class 1A materials by Australian end-users, including the steps that must be taken concerning the removal of class 1A materials in accordance with measure 3.	
IJM Recommendation: For Compliance Measure 2 (above), as well as the measures under Outcome 8 (Industry participants provide clear and effective reporting and complaints mechanisms for class 1A and class 1B material) and Outcome 9 (Industry participants effectively respond to reports and complaints about class 1A and class 1B material), ensure that enforcement action and service provider responses to reports are not limited to only Australian end-users and Australian reporters of class 1A and class 1B material.	
Annual Reporting:	
Provision in draft Code **32) Annual reporting by providers of a Tier 1 social media service** A provider of a Tier 1 social media service must submit a Code report which as a minimum contains the following information: a) details of the risk assessment it has carried out pursuant to clause 3, together with information about the risk assessment methodology adopted. b) the steps that the provider has taken to comply with the applicable minimum compliance measures; and c) an explanation as to why these measures are appropriate. The first Code report must be submitted to eSafety 12 months after this Code comes into effect. Subsequent Code reports must be submitted annually.	**IJM Comments** Annual reporting that includes the tools service providers are using to detect online CSEM and data on livestreamed abuse and first-generation CSAM detected or blocked on their platform or service can be a critical opportunity to develop better detection technologies.
IJM Recommendation: Include in the minimum information required in the Code report the tools service providers are using to detect online sexual exploitation of children and data on livestreamed abuse and first-generation CSAM detected or blocked on their platform or service.	

Relevant Electronics Services Online Safety Code

Guidance on Risk Assessment:	
Provision in draft Code Relevant Electronic Service providers (other than encrypted services and other listed under 5(d)) are required to undertake a risk assessment of their service to determine whether they fall within Tier 1, Tier 2 or Tier 3.	**IJM Comments** The length of time that material lasts or is displayed is not an accurate indication of the level of risk of the service being used for harmful online content. Material that is ephemeral can be captured through a recording or screenshot, and stored, distributed and shared online, causing harm,

Under **Table under 6. (c)** – Potential for virality (functionality) – one factor that could categorise a service as Tier 2 or 3 vs. Tier 1 is whether: The relevant electronic service only enables sharing of: (a) material on a 1:1 basis between end-users, or within a defined group of end-users; or (b) ephemeral material (material that lasts or is displayed only for a short time)	trauma and revictimisation to the survivor for many years. Recorded livestream contributes substantially to the amount of CSAM available online. Livestreamed child sexual abuse is by nature ephemeral but is an egregious form of online harm where the abuse happens in real time and the victims are subject hands-on sexual abuse repeatedly – but each livestream is itself "ephemeral" by nature. The ongoing harm that continues to be perpetrated on platforms where the material is ephemeral indicates a high risk that class 1A material will be accessed, distributed or stored on the service.
IJM Recommendation: In carrying out a risk assessment, "ephemeral material" should be removed as a factor that can categorise a service as Tier 2 or 3.	
Detection of first-generation CSAM:	
No provision in draft Code	**IJM Comments** See Section III.A.
IJM Recommendation: Include a new compliance measure that requires providers of relevant electronic services to detect, remove and report first-generation CSAM and livestreamed CSAM.	
Detection of known CSAM:	
Provision in draft Code **9. Use of technological tools to detect and remove known CSAM** A provider of a relevant electronic service may consider the availability and appropriateness of technological tools designed to detect, flag and/or remove instances of known CSAM the particular relevant electronic service, for example, through the use of hashing, machine learning, artificial intelligence or other safety technologies, and may implement such tools where available and appropriate for the relevant service.	**IJM Comments** This measure is an optional compliance measure under the draft Code. The compliance measure of using technological tools to detect and remove known CSAM should be a minimum compliance measure for relevant electronic service providers. See Section III.B.
IJM Recommendation: Compliance measure 9 (Use of technological tools to detect and remove known CSAM) should not be an optional measure. As a minimum compliance measure, require Tier 1 relevant electronic service providers to use technological tools to detect and remove known CSAM**.**	
Encrypted Services:	
Provision in draft Code **10. Use of technological tools to detect behavioural signals associated with CSEM and pro-terror material** Where it holds data that can be used for such an analysis, a provider of an encrypted relevant electronic service may deploy technological tools designed to detect behavioural signals associated with the distribution of CSEM or pro-terror material, and may implement such tools where available and appropriate for the relevant service.	**IJM Comments** See Section III.D.

<table>
<tr><td colspan="2">IJM Recommendation: Compliance measure 10 (Use of technological tools to detect behavioural signals associated with CSEM and pro-terror material) should not be optional but should be a minimum compliance measure for all encrypted relevant electronic service providers.</td></tr>
<tr><td>Ongoing investment in detection tools and personnel
No provision in draft Code</td><td>IJM Comments
See Section III.D.</td></tr>
<tr><td colspan="2">IJM Recommendation: Include a new compliance measure that requires relevant electronic service providers, including encrypted relevant electronic service providers, to make ongoing investments in tools and personnel that support the capacity of the provider to detect and take enforcement action concerning class 1A material and make explicit requirement to invest in developing innovative technologies to detect first-generation CSAM and livestreamed CSAM.</td></tr>
<tr><td colspan="2"><u>Enforcement of policies; reporting and complaint mechanisms</u></td></tr>
<tr><td>Provision in draft Code
3) Systems and processes for responding to violation of policies prohibiting CSEM and pro-terror material
A provider of a Tier 1 or Tier 2 relevant electronic service must implement systems and processes that enable the provider to take appropriate action in response to violations of terms and conditions, community standards, and/or acceptable use policies prohibiting CSEM and pro-terror material, including at a minimum, systems and process that:
a) enable the review by the provider of reports by Australian end-users of CSEM and pro-terror materials (more detail under “Trust and Safety function” below) and appropriate action to be taken in response; and
b) enable the prioritisation and, where necessary, escalation of reports of CSEM and pro-terror material by Australian end-users.

A provider of:
a) a closed communication relevant electronic service; or
b) an encrypted relevant electronic service,
must have standard operating procedures that either:
i. refer Australian reporters of CSEM and pro-terror materials to eSafety resources; or
ii. enable the review of reports by Australian end-users of CSEM and pro-terror materials (more detail under “Trust and Safety function” below and appropriate action in response).</td><td>IJM Comments
See Section III.F.</td></tr>
<tr><td colspan="2">IJM Recommendation: For Compliance Measure 3 (above), as well as the measures under Outcome 8 (Industry participants provide clear and effective reporting and complaints mechanisms for class 1A and class 1B material) and Outcome 9 (Industry participants effectively respond to reports and complaints about class 1A and class 1B material), ensure that enforcement action and service provider responses to reports are not limited to only Australian end-users and Australian reporters of class 1A and class 1B material.</td></tr>
<tr><td colspan="2"><u>Annual Reporting:</u></td></tr>
<tr><td>Provision in draft Code</td><td>IJM Comments</td></tr>
</table>

26) Annual reporting by providers of a Tier 1 relevant electronic service A provider of a Tier 1 relevant electronic service must submit a Code report which as a minimum contains the following information: a) details of the risk assessment it has carried out pursuant to clause 3, together with information about the risk assessment methodology adopted. b) the steps that the provider has taken to comply with the applicable minimum compliance measures; and c) an explanation as to why these measures are appropriate. The first Code report must be submitted to eSafety 12 months after this Code comes into effect. Subsequent Code reports must be submitted annually.	Annual reporting that includes the tools service providers are using to detect online CSEM and data on livestreamed abuse and first-generation CSAM detected or blocked on their platform or service can be a critical opportunity to develop better detection technologies.
IJM Recommendation: Include in the minimum information required in the Code report the tools service providers are using to detect online sexual exploitation of children and data on livestreamed abuse and first-generation and livestreamed CSAM detected or blocked on their platform or service.	

Designated Internet Services Online Safety Code

Detection of first-generation CSAM:	
No provision in draft Code	**IJM Comments** See Section III.A.
IJM Recommendation: Include a new compliance measure that requires providers of designated internet services to detect, remove and report first-generation CSAM and livestreamed CSAM.	
Ongoing investment in detection tools and personnel:	
Provision in draft Code	**IJM Comments**
7) Ongoing investment in tools and personnel by Tier 1 designated internet services A provider of a Tier 1 designated internet service must make ongoing investments in tools (for example, using hashing, machine learning, artificial intelligence, or other safety technologies) and personnel that support the capacity of the provider to detect, and take appropriate action concerning known class 1A material, proportional to the incidence of class 1A material on the service and the extent class 1A materials are accessible to Australian end-users.	See Section III.C.
IJM Recommendation: Include, as part of the ongoing investment in tools, an explicit requirement to invest in developing innovative technologies to detect first-generation and livestreamed CSAM, not just for known class 1A material.	
Enforcement of policies; reporting and complaint mechanisms	
Provision in draft Code	**IJM Comments**
2) Systems, processes and technologies for enforcement of policies prohibiting CSEM and pro-terror material A provider of a Tier 1 or Tier 2 designate internet service must implement systems and processes that	See Section III.F.

enable the provider to take appropriate enforcement action for violation of terms and conditions, community standards, and/or acceptable use policies prohibiting CSEM and pro-terror material. At a minimum, a provider of a Tier 1 designate internet service must: a) remove instances of CSEM and pro-terror materials identified by the provider on the service as soon as reasonably practicable unless otherwise required to deal with unlawful CSEM and pro-terror materials by law enforcement b) terminate an Australian end-user's account as soon as reasonably practicable in the event the Australian end-user is: i) distributing CSEM or pro-terror materials to Australian end-users with the intention to cause harm; ii) known to be using the account in breach of age restrictions concerning use of the service by an Australian child; or iii) has repeatedly violated terms and conditions, community standards and/or acceptable use policies prohibiting CSEM and pro-terror materials on the service; and c) take reasonable steps to prevent Australian end-users that repeatedly violate terms and conditions, community standards and/or acceptable use policies prohibiting CSEM and pro-terror material who have had their user account terminated from creating a new account.	
IJM Recommendation: For Compliance Measure 2 (above), as well as the measures under Outcome 8 (Industry participants provide clear and effective reporting and complaints mechanisms for class 1A and class 1B material) and Outcome 9 (Industry participants effectively respond to reports and complaints about class 1A and class 1B material), ensure that enforcement action and service provider responses to reports are not limited to only Australian end-users and Australian reporters of class 1A and class 1B material.	
Annual Reporting:	
Provision in draft Code **30) Annual reporting by providers of a Tier 1 designated internet service** A provider of a Tier 1 designated internet service must submit a Code report which as a minimum contains the following information: a) details of the risk assessment it has carried out pursuant to clause 4, together with information about the risk assessment methodology adopted; b) the steps that the provider has taken to comply with the applicable minimum compliance measures; and c) an explanation as to why these measures are appropriate. The first Code report must be submitted to eSafety 12 months after this Code comes into effect. Subsequent Code reports must be submitted annually.	**IJM Comments** Annual reporting that includes the tools service providers are using to detect online CSEM and data on livestreamed abuse and first-generation CSAM detected or blocked on their platform or service can be a critical opportunity to develop better detection technologies.
IJM Recommendation: Include in the minimum information required in the Code report the tools service providers are using to detect online sexual exploitation of children and data on livestreamed abuse and first-generation and livestreamed CSAM detected or blocked on their platform or service.	

V. About IJM

International Justice Mission (IJM) is a global organisation that protects people in poverty from violence. As the largest anti-slavery organisation in the world, IJM partners with local authorities in 29 program offices in 17 countries to combat slavery, violence against women and children, and other forms of abuse against people who are poor. IJM works with local authorities and governments to rescue and restore survivors, hold perpetrators accountable, and help strengthen public justice systems so they can better protect people from violence.

VI. About IJM's Center to End Online Sexual Exploitation of Children

IJM's Center to End Online Sexual Exploitation of Children supports IJM Philippines' mission in protecting children from online sexual exploitation of children and scaling protection globally through: (1) improving technology and financial sector detection and reporting of livestreamed child sexual exploitation, (2) increasing accountability in demand-side countries, and (3) sharing IJM's model to end online sexual exploitation of children with governments, NGOs, and other stakeholders, resulting in sustainable protection for children and accountability for perpetrators.

Annex A: Headlines of Australian Offender Production, Distribution, and Receiving of Livestreamed Child Sexual Abuse

- *AFP helps Filipino authorities arrest three women in livestream child abuse sting*[43]
- *Sydney man jailed for livestreaming sexual abuse of a Filipino child*[44]
- *Australian convicted for livestreaming sexual abuse of Filipino children*[45]
- *Australians are paying as little as $18 to watch child sexual abuse live streamed directly from The Philippines*[46]
- *Western Australian man jailed after ordering livestreamed child sex abuse from the Philippines*[47]
- *256 Australians spend more than $1.3 million watching child sexual abuse online*[48]
- *The live streaming of child sexual abuse in the Philippines has skyrocketed during the COVID-19 pandemic with perpetrators in Australia accounting for nearly a fifth of offenders*[49]
- *Four years' jail for Victorian man who paid Filipina to livestream sexual abuse of children in her care*[50]
- *Malone was sexually abused online aged eight. Many perpetrators are in Australia*[51]
- *Australian accused of child sex tourism arrested in the Philippines*[52]
- *Jail for man who exploited girls in the Philippines*[53]
- *Former public servant Ian Ralph Schapel jailed for abusing children from the Philippines online*[54]
- *Child sex tourist jailed for 'depraved' acts*[55]
- *How police zeroed in on vile teacher's child abuse live streams*[56]
- *Phillip John Ryan: Geraldton paedophile jailed over online abuse of dozens of children in the Philippines*[57]
- *'A danger to the community': Child sex pest's mega jail term*[58]
- *Man, 61, guilty of live-streaming child abuse walks free from court*[59]

[43] https://www.abc.net.au/news/2017-05-12/afp-assists-philippines-livestream-child-abuse-sting/8521820

[44] https://www.afp.gov.au/news-media/media-releases/sydney-man-jailed-livestreaming-sexual-abuse-filipino-child

[45] https://www.ijmuk.org/news/australian-convicted-for-livestreaming-sexual-abuse-of-filipino-children

[46] https://www.theaustralian.com.au/business/technology/australians-comprise-over-20-per-cent-of-a-growing-audience-paying-for-livestreamed-child-sex-abuse/news-story/b718b3a705e63f6c9f97f9a9f44498b5

[47] https://ijm.org.au/news/western-australian-man-jailed-after-ordering-livestreamed-child-sex-abuse-from-the-philippines/

[48] https://www.abc.net.au/news/2020-02-19/australians-paying-to-watch-child-sex-abuse-online/11979844

[49] https://www.sbs.com.au/news/article/malone-was-sexually-abused-online-aged-eight-many-perpetrators-are-in-australia/xc8epp10a

[50] https://osec.ijm.org/news-and-insights/news-updates/four-years-jail-for-victorian-man-who-paid-filipina-to-livestream-sexual-abuse-of-children-in-her-care/

[51] https://www.sbs.com.au/news/article/malone-was-sexually-abused-online-aged-eight-many-perpetrators-are-in-australia/xc8epp10a

[52] https://www.smh.com.au/world/australian-accused-of-child-sex-tourism-arrested-in-the-philippines-20160901-gr6x8x.html

[53] https://www.theage.com.au/national/victoria/jail-for-man-who-exploited-girls-in-the-philippines-20211006-p58xq6.html. Also https://7news.com.au/news/crime/vic-man-who-preyed-on-filipino-kids-jailed-c-4164514

[54] https://www.abc.net.au/news/2022-08-03/paedophile-ian-schapel-jailed-for-abusing-children-online/101294978. Also https://www.dailymail.co.uk/news/article-10107455/Retired-South-Australian-public-servant-Ian-Schapel-67-sexually-exploited-kids-Philippines.html

[55] https://www.couriermail.com.au/questnews/john-joseph-power-child-sex-tourist-sentenced-in-brisbane-supreme-court/news-story/a072ac217e0176ffc873a9761b870ed3

[56] https://www.cairnspost.com.au/truecrimeaustralia/police-courts-cairns/how-police-zeroed-in-on-russell-owen-lorbacks-payperview-child-abuse-habit/news-story/af00dae1f98a313f4e63de0ecfa7d86e

[57] https://thewest.com.au/news/crime/phillip-john-ryan-geraldton-paedophile-jailed-over-online-abuse-of-dozens-of-children-in-the-philippines--c-6985189

[58] https://www.geelongadvertiser.com.au/truecrimeaustralia/police-courts-geelong/batesfords-roger-allan-rivo-jailed-for-12-years-on-child-sex-offences/news-story/7f485b4ee4919495bf765e0b93eba0b5

[59] https://australianseniorsnews.com.au/news/man-guilty-of-live-streaming-child-abuse-walks-free-from-court/

- *Greens candidate who ran against Prime Minister at 2019 election charged with child sex offences*[60]
- *Australian businessman sentenced in Philippines for child sex crimes*[61]
- *WA man charged with 111 child abuse related offences*[62]
- *Victorian man jailed for child abuse offences, after his arrest led to rescue of children in the Philippines*[63]
- *Victorian man jailed for live distance child abuse offences*[64]
- *Brisbane man jailed for 8 years for abusing children overseas*[65]
- *Melbourne man, 61, sentenced for child abuse offences*[66]
- *Adelaide man jailed and 15 young victims rescued after international child abuse investigation*[67]
- *Sydney man jailed and child rescued in the Philippines*[68]
- *71-year-old man jailed for live distance child abuse*[69]

[60] https://www.abc.net.au/news/2020-06-10/greens-candidate-jonathan-doig-chaged-with-child-sex-offences/12338828

[61] https://www.smh.com.au/world/asia/australian-businessman-sentenced-in-philippines-for-child-sex-crimes-20180221-p4z11m.html

[62] https://www.afp.gov.au/news-media/media-releases/wa-man-charged-111-child-abuse-related-offences

[63] https://www.afp.gov.au/news-media/media-releases/victorian-man-jailed-child-abuse-offences-after-his-arrest-led-rescue

[64] https://www.afp.gov.au/news-media/media-releases/victorian-man-jailed-live-distance-child-abuse-offences

[65] https://www.afp.gov.au/news-media/media-releases/brisbane-man-jailed-8-years-abusing-children-overseas

[66] https://www.afp.gov.au/news-media/media-releases/melbourne-man-61-sentenced-online-child-abuse-offences

[67] https://www.afp.gov.au/news-media/media-releases/adelaide-man-jailed-and-15-young-victims-rescued-after-international-child

[68] https://www.afp.gov.au/news-media/media-releases/sydney-man-jailed-and-child-rescued-philippines

[69] https://www.afp.gov.au/news-media/media-releases/71-year-old-man-jailed-live-distance-child-abuse