

Deterrence and help- seeking

Learning brief | 2026 updated version

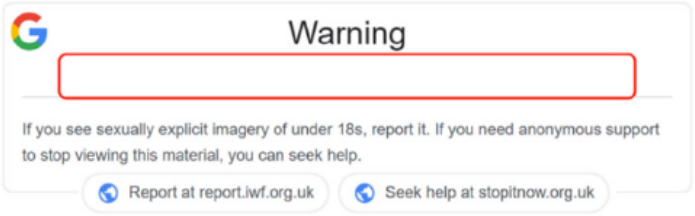
Safe Online | Tech Coalition

What is deterrence and help-seeking?

There are a range of approaches and strategies to deter potential and current offenders from engaging or further engaging with child sexual abuse material (CSAM) online and seeking help to stop them doing this in the future.

Deterrence messages come in a number of forms: from splash pages to search engine warning messages to helpline campaigns, and are a key strategy and are used to change or influence online user behaviours around searching for CSAM. Search engines, social media platforms, and adult content sites now commonly display warning messages when users enter terms associated with CSAM. This offers an efficient and potentially effective entry point, not only to deter people from searching for CSAM online, but also to seek help to stop them doing so in the future.

Help-seeking, in parallel to deterrence messaging, goes beyond stopping a search in the moment, towards people actively seeking support to stop in the future. Some work is being done to promote help seeking through deterrence messaging, such as the examples here, however persistent gaps still remain in the evidence.



Over the last few years, **the Research Fund has supported research to address evidence gaps in relation to deterrence and help seeking through several research projects**. Projects are focused on:

- Understanding the risk factors of people engaging with CSAM and identifying early warning signs
- Testing different types of deterrence messaging to support behaviour change and help seeking behaviours
- Understanding barriers to help seeking behaviour and signposting to help seeking resources

Since this brief was first published in February 2024, the evidence base has moved from perception to practice. **Research Fund grantees have completed programmes of work testing which messages people say would move them, and the first large-scale evaluations of warning messages on live platforms have now been reported.** Over the same period, generative AI has introduced a deterrence problem that did not exist when the first brief was written: material that requires no contact with a child to produce, and a claim circulating among some users that this makes it harmless. This 2026 update covers both developments.

This learning brief draws on findings from five **Tech Coalition Safe Online Research Fund** grantees introduced below.

Protect Children focuses on two research areas: **tech pathways of CSAM users to CSAM in the dark web, and barriers and enablers in help-seeking behaviour of CSAM users.** Its anonymous survey of people searching for CSAM on dark web search engines has passed 30,000 responses in 21 languages, and in 2026 it ran a large-scale warning message experiment on a dark web search engine with University of Kent.

Medical School Berlin (MSB)'s research is focused on **understanding risk factors and identifying early warning signs for people transitioning from viewing legal to illegal materials online, such as CSEM.** Its RAPPID consortium surveyed general population samples in six countries alongside darknet, fanfiction, and help-seeking samples, then took profile-based warning messages into live testing with Aylo.

This project tests and improves an AI system for detecting criminal patterns related to child sexual exploitation live streaming in Latin America. It **analyses conversations between offenders on dark web forums to detect how abuse, including live-streamed abuse, is planned and discussed**. Its AI tool, ARTEMISA, is being expanded across six Latin American countries.

The University of Kent's completed programme ran ten studies with roughly 25,000 participants in the UK, US, Mexico, and South Africa to **identify the factors that lead people to view CSAM online and to test the effectiveness of deterrence messages in encouraging at-risk individuals to seek help**. Its findings informed a redesigned search warning now evaluated on live traffic.

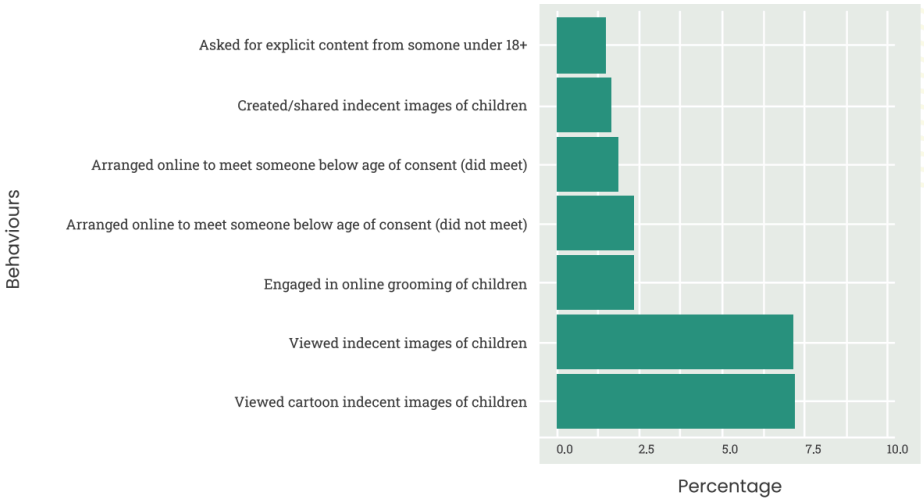
University of Kent's second project examines **AI-generated CSAM and what it means for prevention**, combining stakeholder interviews, analysis of around 1,600 clearweb forum posts, and surveys of people with a sexual interest in children and of roughly 4,500 members of the public, **including an experiment testing harm-based messages**.

In this project, SaferNet Brazil **maps the misuse of generative AI to create and share CSAM and non-consensual sexual images of children and adolescents in Brazil** through a national mapping of school-based cases, a poll of 679 young people, and forensic analysis of reported URLs, showing who is now producing this material.

What have we learned so far about people who view CSAM and their practices?

More men offend but women do too. Both University of Kent and MSB asked people what their current problematic (including illegal) practices regarding CSAM were, and while men across every type of behaviour tested in both studies reported higher online CSAM practices than women, a significant number of women did too in each study. MSB's completed Study 1 puts a figure on this: across its samples, most respondents reporting problematic sexual interests were men, but between one and three in every ten women also reported them.

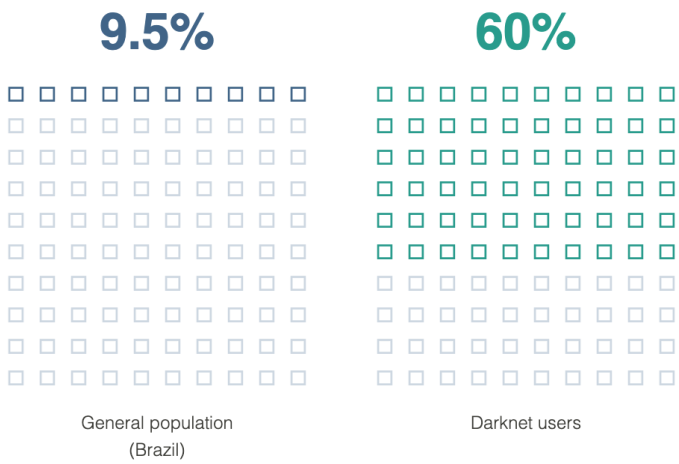
Substantial numbers report risky or illegal behaviour: across University of Kent's completed programme in the UK, US, Mexico, and South Africa, approximately 14% of participants were classed as risky or higher need. In the largest study, between 1.8% and 10.8% self-reported engaging in illegal behaviour such as viewing CSAM, with substantial variation between countries. A more detailed breakdown is given below.



Minimisation is not confined to offender communities. Between 9.5% of MSB's Brazilian general population sample and 60% of its darknet sample did not fully reject the claim that chatting online about sex with children does no harm. Fanfiction community members strongly opposed sexual depictions of children but were noticeably less opposed to sexual chat with them. On dark web forums these beliefs are actively taught and reinforced, with offenders sharing techniques for grooming and evading detection alongside material that rationalises abuse (Universidad de los Andes).

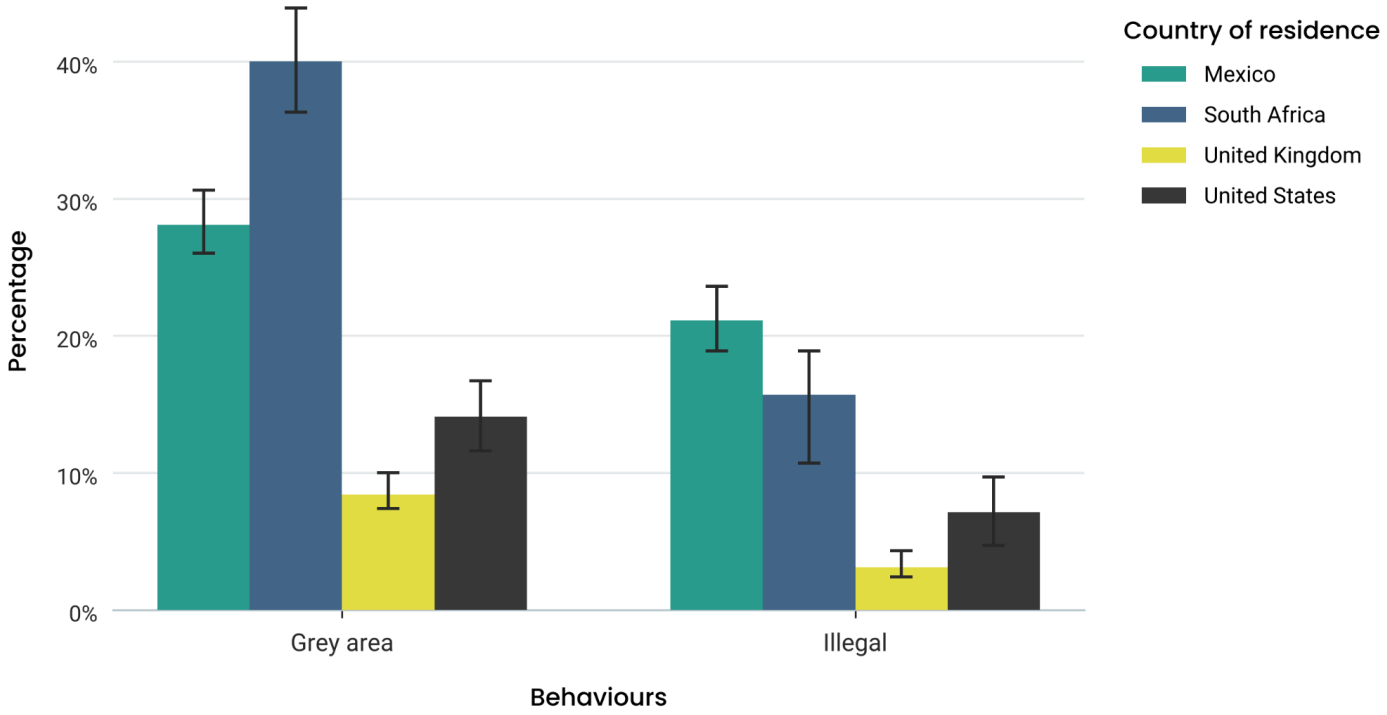
Normalisation of harm across populations

Proportion who did not completely disagree that chatting online about sex with children causes no harm



Source: Medical School Berlin, RAPPID (2025). Each icon represents 1 in 100 respondents; 9.5% approximated to 10 icons for display.

Variations in reported behaviours across geographies: Using University of Kent data, there are clear differing self reported practices across contexts. Substantial numbers of participants self-reported behaviours as adults indicative of risky or illegal online sexual behaviour relating to children or reported some degree of attraction to OCSEA. The University of Kent team observed substantially higher rates in two non-representative studies in Mexico and South Africa. While this is important to identify, why this is the case is not clear.



Interest in pornography focused on young people is higher than that focused on children. Not only does the type / seriousness of the CSAM being viewed impact on self reporting viewing rates, but the age of the children being viewed also plays a role. In MSB's work, 2% reported an interest in viewing CSAM. However, rates of CSAM interest were 5 to 15 times higher if respondents were asked about their CSAM interests in young people rather than children across various behaviours (e.g., engaging in sexual behaviour, viewing pornography, etc.).

Perpetrators who view CSAM are likely to seek direct contact with a child afterwards: Protect Children's survey has now passed 30,000 responses, and the updated figures sharpen the picture. 77% of respondents had encountered CSAM on the surface web and 32% had used social media specifically to search for, view, or share it. 48% had attempted to contact a child via social media and 37% via a messaging app.

The pathway from dark web CSAM consumption to contact-seeking runs largely through the mainstream platforms children use every day. Protect Children's 2026 survey of more than 20,000 CSAM seekers adds that search engines (27%) and pornography sites (22%) were the open web platforms most commonly used to search for CSAM, two environments where warning messages already appear.

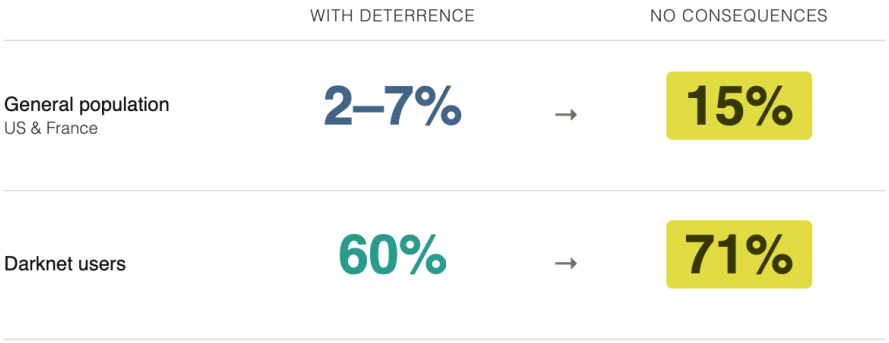
What drives it, and what holds people back from help

People are more likely to break laws when they feel they won't get caught. MSB's completed Study 1 shows how large this effect is. Asked to imagine a scenario with no negative consequences, about 15% of the US and French samples and 71% of the darknet sample reported an interest in CSEM consumption.

In the general population, the rise in interest under the no-consequences scenario was accounted for almost entirely by the perceived risk of being caught. MSB's own reading of this is that the **internet is widely seen as a lawless space**, and that prevention messaging should therefore make clear that detection is both possible and likely. Protect Children's dark web experiment adds a caveat: **messages that lean on detection are less likely to be credible to privacy-aware users**, so how well this framing travels depends on where a message appears.

The impunity effect

Self-reported interest in CSAM consumption, with and without perceived legal consequences



Source: Medical School Berlin, RAPPID (2025). General population N = 1,000 per country; darknet N = 2,331 valid responses.

Sexual interest in children is only one driver among several: the pilot finding that loneliness, boredom, and overlapping social and sexual needs play a role has now been developed into a set of distinct risk profiles. MSB describes psychologically stable, antisocial, withdrawn, and high sex drive types among people who engage with this material, alongside emotional and relational functioning, online-specific risk markers, and antisocial attitudes as independent correlates. Additional risk factors examined include:

- Sexual frequency
- Diversity of pornography use, and preference for pornography over partnered sex
- Excitement at breaking taboos online, and a view of the internet as a lawless space
- Childhood adversity
- Seeing children as sexual being

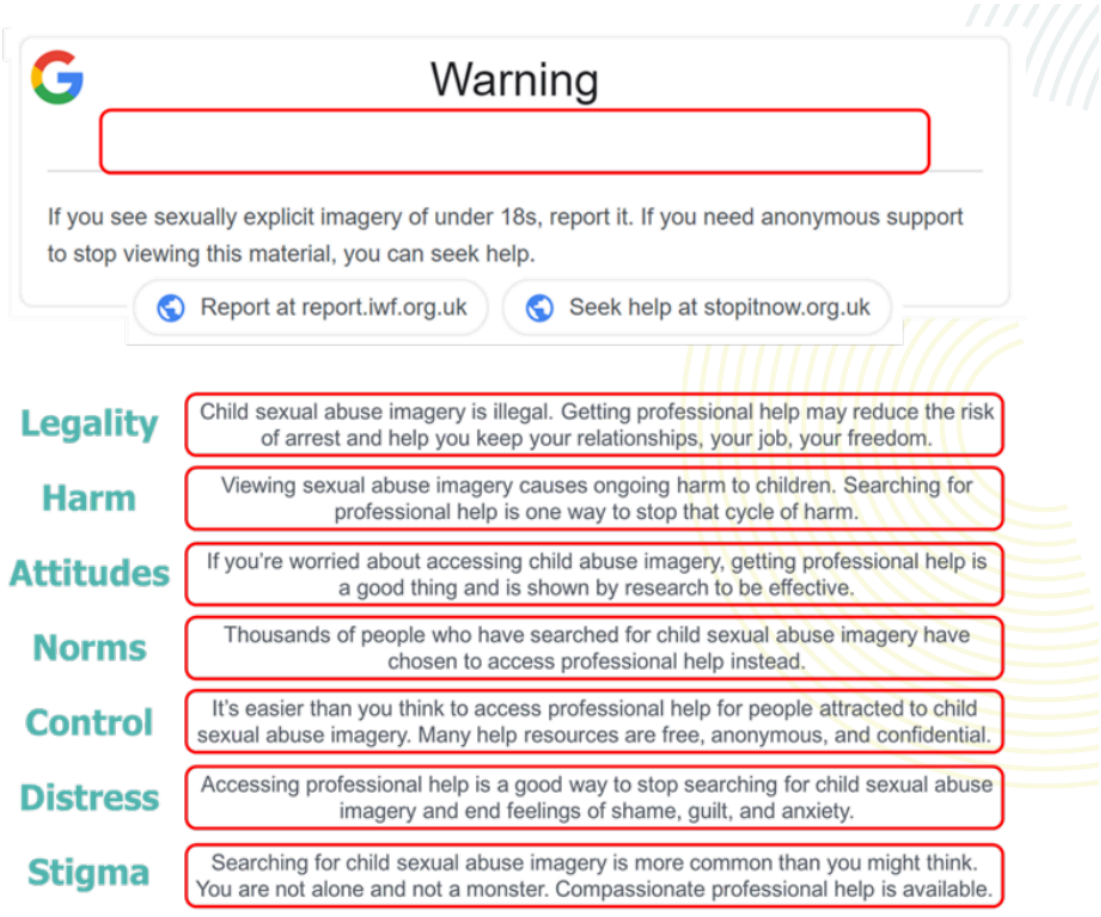
Help-seeking depends on more than distress. University of Kent found people were more inclined to seek help when they felt anxious about their online sexual behaviour, believed people around them would seek help in their position, felt able to access help, and perceived less stigma in doing so.

Protect Children's dark web survey adds who is hardest to reach: respondents who were less likely to seek help included those seeking weekly contact with children (31.6%) and having weekly thoughts about contacting children (25%), those in contact with and influenced by other CSAM users (31%), and those viewing CSAM depicting toddlers and infants (24%). **This puts weight on messaging that lowers the barrier rather than assuming motivation to seek help is already there.**

How can we promote deterrence and help-seeking?

Help-focused messages are seen as more likely to prompt help-seeking. Across ten studies, University of Kent found that the illegality and harm messages typical of industry were rated most effective for deterring searches and encouraging reporting, but weaker at prompting help-seeking. Messages about reducing distress and increasing control over getting help were rated most likely to move people towards support. One gain-framed message ranked consistently high:

"Accessing professional help is a good way to stop searching for child sexual abuse imagery and end feelings of shame, guilt, and anxiety."



In studies 7-10 participants were asked to rank hypothetical Google OneBox messages in terms of which would be most effective in encouraging deterrence, help-seeking, and reporting. The figure shows messages used for studies 9 & 10.

Testing messages in a live environment. University of Kent and other researchers joined the expert panel that redesigned Google's warning for CSAM-seeking search queries, shifting it from a reporting message to one that leads with confidential help while keeping the warning on illegality. Evaluated across nine countries, the new warning reduced follow-on CSAM searches within the same session by 3.8 percentage points compared with the old one, and helplines saw an immediate rise in traffic. The effects are modest and measured within a

single session, so displacement to other services cannot be ruled out, but they are among the first large-scale behavioural evidence that reorienting a warning towards help changes what people do next.

A help-led search warning, measured on real behaviour

3.8 pts

fewer follow-on CSAM searches per session than the old message

5.6 pts when only European countries are compared

0.73%

of sessions clicked straight through to a help service

Helpline traffic rose immediately in all five regions that shared data

0.5% to 1.7%

UK sessions clicking through to help, month before vs month after

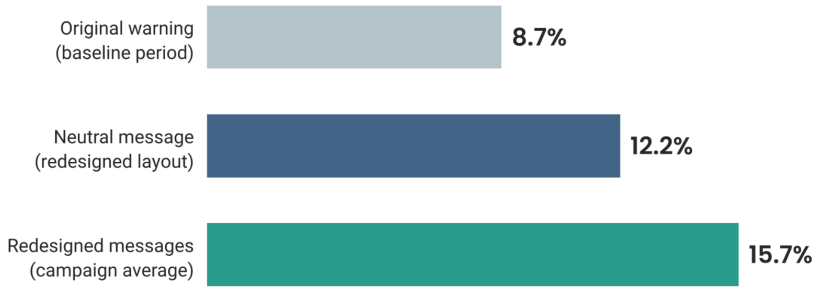
The UK warning already linked to help before the redesign

Source: Umbach et al. (2026), with University of Kent. Redesigned warning launched February 2025; data January to May 2025.

A second evaluation tested message content directly, at scale. Protect Children, working with University of Kent and partners, ran a 140-day field experiment in 2025 on Ahmia.fi, a dark web search engine that blocks CSAM-related searches and shows a warning linking to Protect Children's anonymous ReDirection self-help programme. The team tested eight redesigned messages against a neutral one across more than 3 million blocked searches. **Every message beat the neutral one and clicks through to help rose from 8.7% to 15.7%** (equivalent to over 115,000 additional clicks), but the neutral message alone also beat the old baseline, so the redesigned layout did part of the work. Search volume did not change significantly, so the gain was in help-seeking rather than deterrence.

Redesigned warnings nearly doubled clicks to help

Share of warnings on a dark web search engine that led to a click through to an anonymous self-help programme



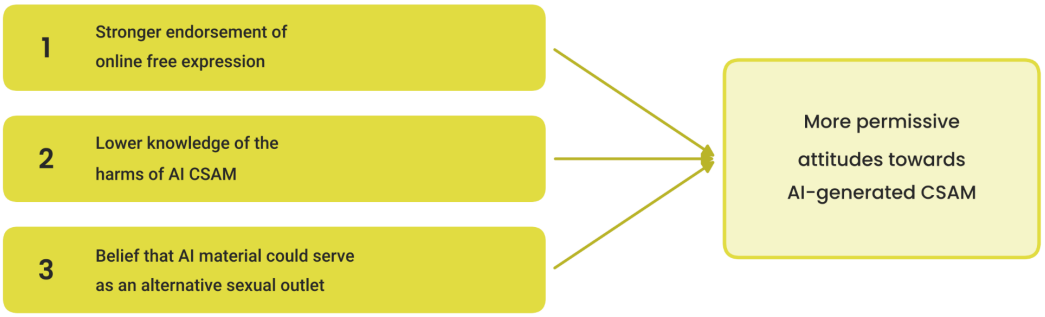
Source: Protect Children (2026), Preventive Pathways. 140-day study, more than 3 million blocked CSAM searches.

What people say would persuade them is a weak guide to what does. In a follow-up survey of over 6,000 searchers, 17.3% of those who recalled a warning said they clicked a help link at the time, and over time 39.4% said it made them reflect on their behaviour, 15.2% reported stopping in the long term, and 6.4% sought help. Most respondents picked a different message as most persuasive from the type they said they preferred in principle, a reminder that stated preferences are a weak guide to what works.

AI-generated CSAM and the deterrence challenge

Generative AI has created a form of CSAM that requires no contact with a child to produce. **With it has come a claim that circulates widely among people who use this material: if no real child was involved, no harm was done.** That claim is a deterrence problem in its own right, because it gives people a reason to continue that the message types tested in the original programme were never designed to answer. University of Kent's second Research Fund project examines how far this reasoning has spread and what would counter it.

What predicts permissive attitudes is belief rather than sexual preference. University of Kent surveyed 100 people with a sexual interest in children alongside a general population sample of roughly 4,500. After accounting for attitudes toward camera-taken CSAM, three factors were associated with more permissive attitudes toward AI CSAM (see on the right).



Source: University of Kent, survey of 100 people with a sexual interest in children. Emerging findings.

In the general population sample, attitudes toward the illegality of both forms were overwhelmingly prohibitive, though endorsement of illegality was slightly weaker for AI-generated material.

The people producing it are not who warnings were written for. Stakeholders interviewed by University of Kent described AI tools lowering the technical barriers to producing abusive imagery and **drawing in people outside established CSAM communities**, **including adolescents, while expanding the contexts in which abusive imagery is produced to include peer nudification, harassment, and sexual extortion.** In SaferNet Brasil's school cases, offending was collective and peer-driven, and often treated as bullying or a prank. In its poll of 679 young people, 86.8% of girls but 61.9% of boys said creating a non-consensual sexual deepfake of a classmate is definitely a crime, and 12.4% of young men said they had made one.

The cross-cutting conclusion is about harm literacy. University of Kent's reading of its own findings is that the risk sits in the narrative of harmlessness itself, and that the response is to make the harms of AI-generated material understood rather than assumed. That is a different task from the seven message types tested in the original programme, which addressed illegality, harm, distress, control, norms, and stigma in relation to camera-taken material. University of Kent has begun to test which framings work.

In its general population survey, most harm-based framings of AI CSAM produced little change in how people prioritised prevention resources, but two produced small shifts: linking AI CSAM to sexual extortion, and emphasising that it can depict real children. What made a framing work was credibility rather than novelty, as people moved when they already found the argument legitimate, while those already familiar with AI CSAM moved least. These are attitudinal results, and whether harm literacy messaging changes behaviour has not yet been tested.

Tools emerging from the research

Several of the tools flagged in the 2024 brief now exist, and the table below sets out what has been built as well as what is still to come.

Grantee	Tool or output	Where it stands
Medical School Berlin	**Risk factors, profile-based warning messages, a validated image set, and survey guidance**	Risk factors for moving from legal to illegal material have been identified, and profile-based messages are in live testing. A new AI-generated, age- and ethnicity-diverse image set replaces stimulus material from the early 2000s, and guidance on surveying hard-to-reach populations has been published.
Protect Children	**Recommendations for industry, the ReDirection self-help programme, and warning message guidance**	Recommendations for the tech industry on surface web platforms and messengers. ReDirection has been visited over 100,000 times. The 2026 Preventive Pathways report adds recommendations on evaluating what warnings say, rotating messages, and making help the dominant action.
Universidad de los Andes	**ARTEMISA, an AI alert system for CSEA live-streaming**	Rebuilt around a continuum model that tracks how cases move between grooming, sextortion, and live-streaming, using language analysis to identify offender interactions. Being expanded across six Latin American countries.
University of Kent	**Optimised deterrence messages, and harm communication on AI CSAM**	Moved from testing into deployment, informing a redesigned search warning that leads with help in 11 regions and a large-scale dark web message experiment. Next: an article in AI & Society and a series of insight briefs.

Sources

This learning brief is based on evidence and interviews from five research projects funded by the **Tech Coalition Safe Online Research Fund**. More information on each grantee's research is available below.

Grantee	Publications	Contact
Protect Children Finland	CSAM Perpetrator Research Report: Findings from a Survey of CSAM Perpetrators on Digital Platform Use and Design (Tell Me More About Tech) Tech Platforms Used by Online Child Sexual Abuse Offenders Preventive Pathways: How warning messages interrupt and prevent online child sexual abuse at scale (2026) Factors associated with help-seeking among online CSAM offenders (2024)	**Nina Vaaranen-Valkonen** nina.vaaranen-valkonen@suojellaanlapsia.fi **Tegan Insoll** tegan.insoll@suojellaanlapsia.fi
Medical School Berlin Germany	*Publication incoming*	**Robert Lehmann** robert.lehmann@medicalschool-berlin.de **Laura Quinten** laura.quinten@medicalschool-berlin.de
Universidad de los Andes Colombia	Guarding the Guardians: Automated Analysis of Online Child Sexual Abuse Artificial Intelligence for the Prevention of Online Sexual Abuse of Children and Adolescents in Colombia (book chapter)	**Lina María Saldarriaga** l.saldarriaga@aulasenpaz.org **Viviana Quintero** v.quintero@aulasenpaz.org **Yuly Calderón** y.calderon@aulasenpaz.org
University of Kent United Kingdom	Deterring searches for child sexual abuse material on Google Search and promoting help-seeking (2026) AI-generated child sexual abuse material: what's the harm? (AI & Society, 2026)	**Caoilte Ó Ciardha** c.c.ociardha@kent.ac.uk



SaferNet Brasil
Brazil

Publication incoming

Juliana Cunha
julianacunha@safernet.org.br

Sofia Schurig
sofia@safernet.org.br

Thiago Tavares
thiagotavares@safernet.org.br

Acknowledgements

This learning brief was prepared by Neira Budiono (Safe Online), Natalie Shoup (Safe Online) and Nicholas Wells (Tech Coalition).

It synthesises research by the Tech Coalition Safe Online Research Fund grantees and their collaborators. We are grateful to them for their work, and for reviewing the findings presented here.

Protect Children: Eva Díaz Bethencourt, Tegan Insoll, Anna K. Ovaska, Valeriia Soloveva and Nina Vaaranen-Valkonen

University of Kent: Caoilte Ó Ciardha

Medical School Berlin: Robert Lehmann and Laura Quinten

Universidad de los Andes: Diana Agudelo, Pablo Arbeláez, Yuly Calderón, Alejandro Castañeda, Ángela Castillo, Laura Hernández, Wilmar Osejo, Carolina Paz, Juanita Puentes, Viviana Quintero and Lina María Saldarriaga

SaferNet Brasil: Juliana Cunha, Sofia Schurig and Thiago Tavares

Co-authors and collaborators: Mikko Aaltonen, John Buckley, Donald Findlater, Ainslie Heasman, Griffin Hunt, Elizabeth Letourneau, Juha Nurmi, Arttu Paju, Rebecca S. Portnoff, Ethel Quayle, Abhishek Roy, Joel Scanlan, Chad Steel, Rebecca Umbach, Maximilian von Heyden and Richard Wortley

How to cite this resource

Safe Online and Tech Coalition (2026). *Deterrence and help-seeking: learning brief, 2026 edition*. Tech Coalition Safe Online Research Fund.

Safe Online | Tech Coalition

The Tech Coalition Safe Online Research Fund is investing in knowledge and research to end online child sexual exploitation and abuse (CSEA). The Tech Coalition and Safe Online joined hands in 2020 to collectively work to support knowledge and research towards ending digital harms against children. The Fund is continuing its focus on innovative research that produces actionable insights to impact product and policy development, with a priority given to research that can help inform the technology industry's approach to combating online child sexual exploitation and abuse.