Protect Children

Online Warning Messages for CSAM Prevention

Evidence and Practice Mapping Report

Protect Children
2025

Executive Summary

Online warning messages are a promising tool to tackle use of child sexual abuse material (CSAM). Evidence shows that well-designed messages can disrupt offending, deter CSAM seeking, and encourage help-seeking.

This report reviews current international literature and practice on deterrence and help-seeking messaging to map existing approaches. It finds that implementation remains uneven and research gaps persist, and identifies priorities for further evidence.

Contents

Cite this report

Contributors

This report was prepared for Protect Children by Tegan Insoll[1], Dr Joel Scanlan[2], Dr Caoilte Ó Ciardha[3], Dr Juha Nurmi[4], Valeriia Soloveva[1], Emily May[1], Anna Ovaska[1], and Nina Vaaranen-Valkonen[1].

[1] Protect Children
[2] CSAM Deterrence Centre, University of Tasmania
[3] University of Kent
[4] Faculty of Information Technology and Communication Sciences, Tampere University

Protect Children

1. Introduction

The proliferation of child sexual abuse material (CSAM) is increasingly recognised as a global epidemic with wide-reaching consequences. In tackling the use of CSAM, online environments offer opportunities for early intervention. Research shows that many CSAM offenders experience a desire to stop, yet few seek help to change their behaviour, often due to fear or a lack of awareness about available resources (Insoll et al., 2024). This gap highlights a critical opportunity for timely interventions at moments of potential risk.

Digital warning messages can intervene in real time by disrupting searches, prompting reflection, and linking users to support. Research suggests that well-designed messages can reduce engagement with CSAM and encourage help-seeking. However, implementation is inconsistent across platforms, and gaps in research remain.

We reviewed the current evidence base on digital CSAM prevention warning messages and investigated how warning messages are being implemented in practice. This report provides a practical and policy-oriented overview of the current landscape.

This report offers an overview of current evidence and practice on digital CSAM prevention warning messages.

2. CSAM Deterrence and Prevention

Deterrence Messaging

Deterrence messaging aims to **discourage harmful or illegal behaviour** by raising awareness of consequences, introducing barriers, or promoting alternatives. It has been widely applied in fields such as public health, fraud prevention, and tax compliance. Examples include graphic health warnings on cigarette packaging (Gallopel-Morvan et al., 2019), official letters deterring tax evasion (Bott et al., 2017), and home security stickers to reduce burglary risk (Chainey, 2021).

Increasingly, deterrence messaging has been deployed in digital spaces, and studies have tested its use to tackle cyberattacks, digital piracy, and engagement with harmful communities (see e.g., Cross & Kelly, 2016). A review by Prichard et al. (2022b) found that most studies on online warnings indicated that **behaviour could be modified by the use of a warning**. This evidence underpins the rationale for adapting deterrence strategies to CSAM prevention in online environments.

Online CSAM Preventive Interventions

Given the scale and complexity of the proliferation of CSAM online, and the limited capacity of law enforcement to investigate every instance (Quayle & Koukopoulos, 2019), there has been growing interest in preventive measures that can disrupt demand. One such approach is the use of online warning messages, which can be automatically triggered when users engage in high-risk behaviours.

These messages are displayed in real time, reaching individuals at critical moments of potential offending risk (Prichard et al., 2022b). The deployment of CSAM prevention warning messages reflects a broader policy and research shift toward preventive responses to online child sexual abuse.

How Online CSAM Warning Messages Work

Online CSAM prevention messages are typically deployed through automated systems that detect high-risk user behaviours in real time.

Common triggers include:
- Entering CSAM-related **keywords** into a search engine or content platform
- Attempting to access a **URL** previously identified and removed for hosting CSAM content (Hunn et al., 2023)
- Activation of **grooming detection** algorithms
- Uploading **images** that match **known CSAM hashes**

Once triggered, the system interrupts the user's action and displays a warning message. These can appear as **banner notices, pop-ups, or full-page blocks** requiring user acknowledgement before proceeding. Evidence suggests that active and interruptive formats are more effective at deterring behaviour (Prichard et al., 2022b). However, the design and implementation of these warnings vary significantly across platforms and jurisdictions, influencing both reach and consistency of impact.

What the Messages Communicate

CSAM warning messages are most often implemented by search engines or content platforms. While message framing varies, they typically communicate **legal and behavioural warnings**, highlight the **harm caused to children**, and may encourage **help-seeking** by linking to support services (Walsh et al., 2023). For instance, the UK's online safety regulator, Ofcom, recommends that large search services display warnings when users enter CSAM-related queries, including information on the illegality of CSAM and links to prevention resources (Ofcom, 2023).

These interventions are considered **secondary prevention** strategies, targeting individuals who may be at risk of offending but have not yet committed a known crime (Smallbone & Wortley, 2017; Wortley & Prichard, 2023).

In contrast, primary prevention broadly targets the entire population to stop offending before it begins (e.g., public awareness campaigns), while tertiary prevention focuses on preventing re-offending among those already detected (e.g., treatment programs) (Smallbone & Wortley, 2017; Knack et al., 2019).

Theoretical Foundations: Situational Crime Prevention

The rationale behind these interventions aligns with **situational crime prevention theory**, which suggests that altering the immediate environment can reduce offending opportunities.

In the CSAM context, warning messages can increase the perceived effort required to access illegal content and heighten the perceived risk of detection. They can also disrupt the sense of anonymity or inevitability that may facilitate harmful behaviour (Wortley, 2012).

Beyond Deterrence: Education and Support

Online CSAM prevention messages serve not only to deter offending but also to educate and support. They raise awareness of the legal and personal consequences of CSAM consumption, explain the harm to children, and challenge the normalisation of CSAM in certain online spaces. By encouraging help-seeking and providing access to anonymous support resources, these interventions can promote long-term behavioural and attitudinal change.

Scalability & Cost-Effectiveness

A key advantage of online prevention messaging lies in its **scalability**. For technology companies, delivering automated warnings at scale is exceptionally **low-cost**, as the digital infrastructure to reach millions of users already exists. The marginal cost of issuing additional warnings is virtually zero, while the societal costs of CSAM offending, including investigation, prosecution, incarceration, and the lifelong harm to victims, are immense.

This cost-effectiveness, combined with preventive potential, has driven increasing interest in online warning systems as part of a broader shift toward proactive, preventive responses to online child sexual abuse (Quayle & Koukopoulos, 2019).

3. Research Evidence on CSAM Prevention Messaging

A growing body of research on CSAM warning messages provides encouraging evidence that timely and well-designed messages can reduce engagement with CSAM and support users in seeking help. This section presents a concise review of relevant studies, mapping the current evidence base on the effectiveness and implementation of online prevention messaging.

Experimental RCTs on Warning Messages for CSAM Prevention
Prichard et al., 2021; Scanlan et al., 2022; Prichard et al., 2022a; Prichard et al., 2024.

Approach
This series of randomised controlled trials (RCTs) tested the effectiveness of warning messages designed to deter users from attempting to access websites related to "barely legal" pornography and deterring users from sharing sexual images. Across the studies, users who clicked on honeypot advertisements were randomly assigned to either a control condition (no message) or received deterrence or therapeutic-style messages before being given the option to proceed.

The studies varied the message content (e.g. harm to viewer, harm to victim, illegality, traceability by police), format (text-only, image), and scenario (e.g. viewing "barely legal" content, uploading a sexual image to gain access).

Findings
Across the trials, warning messages consistently dissuaded users from accessing the sites. Messages emphasising the illegality of the behaviour and the ability of police to track users' activity were the most effective deterrents.

Therapeutic and harm-based messages also reduced engagement, though to a lesser extent. There were minimal differences observed between text-based and image-based formats, suggesting that content may matter more than presentation.

The findings suggest that simple, targeted messages delivered at key moments of user intent can disrupt harmful behaviour and offer a cost-effective, scalable approach to online CSAM prevention.

reThink Chatbot Evaluation: Reducing the Number of Searches for CSAM on Pornhub UK
Scanlan et al., 2024 for the Internet Watch Foundation, Lucy Faithfull Foundation, and Aylo

Approach
The study evaluated the reThink Chatbot, which aimed to direct individuals attempting to search for CSAM on Pornhub in the UK to support services. When a user's search included one of the 28,000+ banned terms on Pornhub, the chatbot appeared as a pop-up, along with a warning message. The chatbot warned users about their illegal behaviour and signposted them to Lucy Faithfull Foundation's Stop It Now service to get help and support to stop their behaviour.

Findings
The evaluation found that the warning message and chatbot reduced the total number of searches for CSAM on Pornhub. Most sessions which triggered the warning and chatbot once do not appear to have searched for CSAM again. However, the effectiveness of the intervention seemed to decline over the length of the study. The authors suggest integrating video campaigns and changing the content for users who have previously seen the warning to add to the novelty of the warning page in order to increase engagement.

Message Content and Framing Influences Perceived Effectiveness of Warning Messages for Child Sexual Abuse Material

Ó Ciardha et al., 2025 (in preparation)

Approach

A large online experiment with four independent samples tested how different warning message framings were perceived in terms of deterring further CSAM searching, promoting help-seeking, and encouraging reporting. Participants included general population groups from several countries (US, UK, South Africa, Mexico) and individuals pre-screened for higher risk of problematic online sexual behaviour involving children. Messages included industry-standard themes (legality, harm to children) and alternative framings aimed at encouraging help-seeking by highlighting the accessibility of support, reducing stigma, or addressing psychological distress. In the two later samples, several messages were re-worded into more positive, "gain-framed" versions.

Findings

Messages emphasising legal consequences and harm to children were consistently rated most effective for deterring further CSAM searches and encouraging reporting. For motivating help-seeking, messages focusing on the attainability and confidentiality of support, or on reducing feelings of anxiety and distress, ranked highest, particularly when positively framed. Reframing legality and harm messages into positive terms improved their perceived help-seeking impact while maintaining their strong deterrence effect. These findings underline that message content matters and should be matched to the platform's primary goal, whether that is immediate deterrence or creating an opportunity to direct individuals toward help.

Adolescent Perceptions of Warning Messages to Counter Online Sexual Abuse

Garrett et al. 2025 (in press)

Approach

This mixed-methods study sought feedback from 298 Australian adolescents (aged 16-18) on the perceived utility of online warning messages. Participants read vignettes describing problematic online scenarios, including image-based abuse (IBA), sexual extortion, and cyberbullying, and then evaluated corresponding image- and text-based warning messages. The messages were designed for the different roles in the scenarios: victims, offenders, and bystanders. The study measured adolescent perceptions of the messages' ability to capture attention, encourage reflection or help-seeking, and motivate engagement with support services. It also assessed preferences for image-based versus text-only formats.

Findings

Participants generally agreed that the warning messages would attract the target's attention and prompt reflection or help-seeking. However, they perceived victims as more likely to click a link to access support services than offenders or bystanders. Perceptions of the messages were largely similar across male and female participants. Participants were divided on whether image-based or text-only warnings were more effective, suggesting a combined approach may be best. Qualitative feedback indicated that while the messages were generally well-received, their impact could be limited by factors like the user's pre-existing intent, and that stronger 'scare tactics' might be needed for offenders.

Designing Effective Digital Advertisements to Prevent Online Consumption of Child Sexual Exploitation Material

Henry, 2020

Approach

By discussing a set of test advertisements in focus groups and interviews with staff and clients of New Zealand's major treatment service providers for harmful sexual behaviour, the qualitative study sought to understand which images, text, and advertisement designs are likely to be most effective in directing persons struggling with CSAM use to seek support.

Findings

The study found that effective ads should avoid judgmental tones and instead offer messages that are informative and hopeful. The study found that advertisements should use direct yet non-confrontational language and evoke trust through assurances of anonymity and confidentiality.

Pathways and Prevention for Indecent Images of Children Offending: A Qualitative Study

Bailey et al., 2022

Approach

In a qualitative study, the authors interviewed 20 men in the UK who had viewed CSAM and been investigated by the police, asking about their offending and discussing methods they believe would have been more successful in deterring their illegal behaviour.

Findings

Most participants had never encountered deterrence messages, and those that had did not find them to be effective. The participants identified the following methods that they believe would be effective in deterring CSAM viewing: preventing online access to CSAM, receiving information about the illegality and consequences of offending, and being aware of available confidential help to stop.

4. Implementation of CSAM Prevention Messaging in Practice

We investigated how CSAM warning messages are implemented in practice. This section provides an overview of the current approaches.

Search Providers

Google: CSAM Prevention and Deterrence Intervention on Search

Google has had warning messages for CSAM searches in place since 2013, and has redesigned and expanded them a number of times. Current messages focus on the illegality of CSAM, provide instructions on how to report content to authorities, and offer direction to help resources. Google reports that their warning messages decrease follow-up CSAM searches and have directed hundreds of thousands of users to CSAM reporting hotlines (Google, 2021; Steel et al., 2015). Google reported earlier this year that its new OneBox warning message, deployed in 21 countries, was viewed 20 million times in one month (Roy & Hunt, 2025). This highlights the scale of searches occurring and the importance of such interventions on search provider platforms.

Microsoft: Child Abuse Search Pop-Ups on Bing

Microsoft's Bing search engine also introduced pop-up warnings for CSAM searches in 2013 (BBC News, 2013). However, limited information about the warnings is openly available.


Pornography Sites

Aylo: CSAM Deterrence Messages on Pornhub

In collaboration with relevant organisations, adult content site Pornhub implements one global and 11 localised deterrence messages. The deterrence messages emphasise the illegality of CSAM and offer support resources for controlling urges to engage in illegal behaviour (Pornhub, 2025). Between 2021 and 2023, the deterrence messages on the UK website were triggered over 4.4 million times (Scanlan et al, 2024).
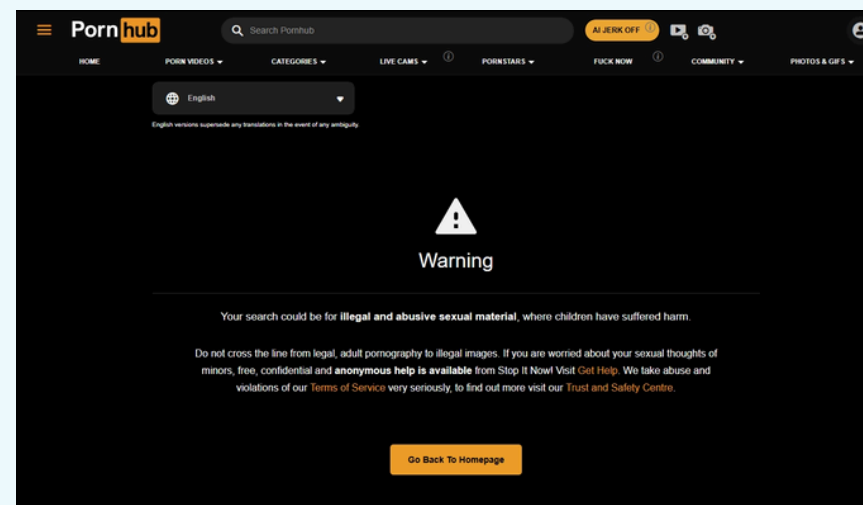
Social Media

Many social media sites deploy pop-up warnings that appear when users search for terms associated with CSAM. For example, Meta deploys such warnings on Facebook and Instagram that highlight the illegality and consequences of viewing CSAM and the harm it causes to children, and encourages users to get confidential help (Meta, n.d.). TikTok has introduced similar search-related warnings, while Snapchat issues warnings when CSAM-related content is removed. However, publicly available information about the presence and effectiveness of these warnings remains limited.

Facebook

TikTok

Snapchat

Instagram
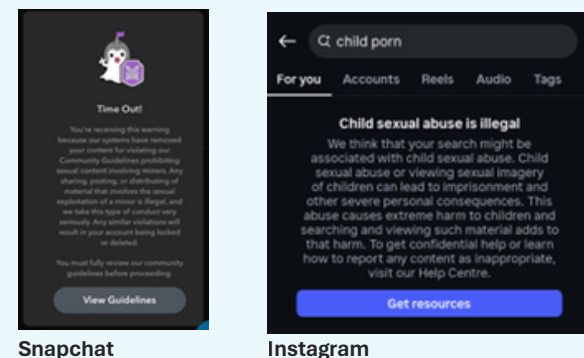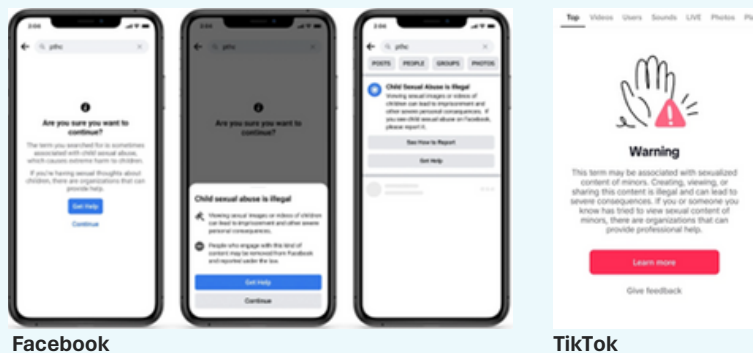
Law Enforcement

Police2Peer: Targeting File Sharing of Child Sexual Abuse Material
The program operates by deploying files that masquerade as CSAM on peer-2-peer file sharing networks, but are warning messages from the police. The messages, which direct viewers to https://helplinks.eu/, are photographs of police officers in uniform holding up a physical warning. As such, the messages are strongly law enforcement themed, but also direct users to therapeutic support services (Europol, n.d).

Collaborative Initiatives

Internet Watch Foundation: Splash Pages
The Internet Watch Foundation recommends that its members implement splash pages to replace blocked sites containing CSAM. These splash pages include warnings about the illegality of the content and encourage users to seek help (IWF, n.d.). Splash pages are used, for example, by internet service providers in the UK, in addition to Microsoft and Google (ECPAT International, 2021).
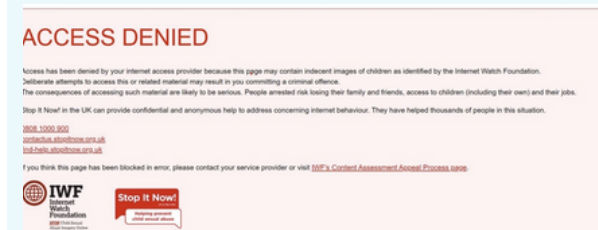
Project Intercept (Lucy Faithfull Foundation)

The Lucy Faithfull Foundation has been partnering with technology companies (such as Google) to include signposting links to the Stop It Now therapeutic helpline for over ten years (Walsh et al, 2023). These partnerships have since become a focus point in Project Intercept, funded by Nominet, an initiative seeking to roll out effective online warnings with a range of technology partners to prevent CSAM use. The project aims to establish new partnerships with the technology industry, to test different message types, and to develop new approaches to scale up warning messages. To date, the Lucy Faithfull Foundation has partnered with more than 10 mainstream technology service providers, which drive an average of 30,000 people a month from around the world to seek help at Stop It Now.

The CSAM Deterrence Centre

The newly established CSAM Deterrence Centre works with the technology industry and other partners that design interventions to prevent the accessing and sharing of CSAM and other harmful behaviours, including grooming and sexual extortion (CSAM Deterrence Centre, 2025). The Australian based Centre works with partners to implement safety-by-design, particularly through effective warning message design, to deter and disrupt CSAM offending. The Centre is a joint initiative between Jesuit Social Services (who operate Stop It Now! Australia) and the University of Tasmania.

5. Key Takeaways

The growing body of research on CSAM warning messages provides encouraging evidence that **timely, well-designed messages can deter engagement with CSAM** and **encourage help-seeking**.

Randomised controlled trials have demonstrated empirically that warning messages, especially those emphasising illegality and traceability, can interrupt harmful behaviour. Qualitative studies add further insight, underscoring the value of trust, tone, and confidentiality in encouraging help-seeking.

However, despite these promising findings, implementation across digital platforms remains uneven. Our mapping of current practice revealed that despite growing industry recognition of the role of deterrence messaging, there is limited publicly available information about how tech companies deploy prevention messages and how effective the measures are, delaying their adoption more broadly.

Our findings align with those of Price et al. (2025), whose review of digital interventions for the prevention or early intervention of online child sexual abuse found that limited application and evaluation of such measures restricts the ability to assess their effectiveness.

To close this gap, further research and cross-sector collaboration are needed to ensure that messaging strategies are **evidence-based, widely adopted, and evaluated** to facilitate a cultural shift towards proactive warning message adoption across the industry.

Initiatives such as **Project Intercept** and the newly established **CSAM Deterrence Centre** seek to achieve these aims, by enhancing collaboration between NGOs and the technology sector and generating evidence to inform regulators.

Specific gaps in research include understanding how to tailor warning messages to different types of CSAM offenders (Garrett et al., 2025 (in press); Ó Ciardha et al, 2025 (in preparation); Prichard et al., 2022b), and understanding how to ensure that the effectiveness of warning messages does not diminish over time due to habituation. Another gap, as noted by Anderson et al. (2023) is that the target audience is rarely involved in the design of crime prevention interventions.

Summary:
- Research shows that well-designed warning messages can **deter engagement with CSAM** and encourage help-seeking.
- We found **limited transparency** around how technology companies deploy prevention messages and measure their impact.
- New initiatives seek to **enhance collaboration** between civil society and the technology sector and generate evidence.

6. Good Practice in Message Design & Implementation

Research and practice indicate that online warning messages are **most effective when integrated into the design** of the overall intervention system rather than added as an afterthought. Messages developed later, or based on untested assumptions about the target audience, are often less effective (Wogalter et al., 2012). In the context of CSAM prevention, message design should be informed by behavioural science and adapted to the digital environment in which the message appears.

Studies across fields show that warnings are most effective when **attention-grabbing, clearly worded, believable, and issued by a credible source** (Prichard et al., 2022b; Lenorovitz et al., 2012; Riley et al., 2006).

Importantly, message effectiveness **depends heavily on context** and the same message design principles cannot be applied universally across settings. Online CSAM prevention messaging should be both highly specific and sensitive to the motivations, fears, and justifications of potential offenders.

Multi-format delivery may enhance the effectiveness of CSAM warning messages, although the evidence is mixed. Some studies suggest that combining **visual elements** (e.g. text and icons) with other formats (e.g. audio alerts, interactive elements) increases user attention and recall (Taylor & Wogalter, 2019). However, other studies found no significant difference in the effectiveness of text-only or image-based messages (Prichard et al., 2022a; Prichard et al., 2024). The addition of the reThink Chatbot increased the deterrent effect of the Pornhub warning message (Scanlan et al., 2024), however the authors could not separate whether this was due to the visual format or the chatbot interaction itself.

Core Themes in CSAM Warning Messages

Legal and moral deterrence: Messages that clearly highlight illegality and risk of prosecution are among the most empirically supported deterrence strategies, particularly when framed with official authority (Prichard et al., 2021). Although these messages tend to be less effective in motivating help-seeking, positively reframing them can increase their potential to encourage individuals to seek support (Ó Ciardha et al., 2025, in preparation).

Traceability and accountability: Messages implying or explicitly stating that a user's activity may be monitored (e.g. references to IP tracking or surveillance) can increase the perceived risk of detection, a key element of deterrence (Prichard et al., 2024). The impact of this approach on help-seeking has not yet been empirically tested.

Harm and victim awareness: Messages that describe the real world harm to children are designed to evoke empathy and disrupt cognitive distortions that justify CSAM use (Edwards et al., 2021). Harm-based messaging shows promise in in encouraging help-seeking, as it appeals to users' moral emotions and compassion.

Pathways to change and support: Offering information about and links to anonymous self-help or therapeutic resources can provide a non-punitive pathway toward behaviour change and plays a key role in long-term prevention (Wortley et al., 2024). Integrating help resources directly into deterrence messages improves accessibility at moments of heightened readiness to change.

Emotional appeals: Emotionally charged messages may be effective in deterring CSAM offending by provoking feelings such as guilt, empathy, or shame (Prichard et al., 2021). However, effectiveness depends heavily on tone, and excessively punitive or shaming language can trigger defensiveness and reduce help-seeking.

Reducing access: Introducing friction, such as by blocking search results, displaying interruptive pop-ups, or requiring active user acknowledgement can interrupt offending behaviours and prevent escalation among individuals early in the offending trajectory (Wortley, 2012; Wortley & Smallbone, 2006; Prichard et al., 2022b).

7. Conclusion

The proliferation of CSAM online presents an **urgent and complex challenge**. Online warning and support messages offer a **promising, scalable intervention** that can complement broader prevention efforts by disrupting the consumption of CSAM, deterring illegal behaviour, and promoting help-seeking at moments of risk. As this mapping paper has shown, evidence from experimental research and real-world implementation suggests that well-designed messages **can effectively reduce engagement with CSAM** and encourage behavioural change.

However, implementation across platforms remains **fragmented** and often opaque, with limited publicly available information about effectiveness or evaluation. While **some major platforms** have taken steps to integrate deterrence messaging, others lag behind. Research gaps remain, particularly around tailoring messages to diverse offender profiles, evaluating long-term impact, and involving the target audience in message development.

There is **significant potential** to refine and expand the use of digital messaging as part of a broader strategy **to prevent online child sexual abuse and exploitation**.

Digital messaging holds strong potential as a scalable, evidence-informed tool within a comprehensive CSAM prevention strategy.

8. References

Anderson, L., Bates, L., & Schaefer, L. (2023). Design and development of a digital intervention for crime prevention: a road safety intervention targeting young driver non-compliance. Safer Communities, 22(1), 56-71. https://doi.org/10.1108/SC-07-2022-0026

Bailey, A., Allen, L., Stevens, E., Dervley, R., Findlater, D., & Wefers, S. (2022). Pathways and Prevention for Indecent Images of Children Offending: A Qualitative Study. Sexual offending (Trier, Germany), 17, e6657. https://doi.org/10.5964/sotrap.6657

BBC. (2013). Microsoft's Bing introduces child abuse search pop-ups. BBC News. https://www.bbc.com/news/technology-23476089

Bott, K., Cappelen, A., Sørensen, E., & Tungodden, B. (2017). You've got mail: A randomized field experiment on tax evasion. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3033775

Chainey, S. (2021). A quasi-experimental evaluation of the impact of forensic property marking in decreasing burglaries. Security Journal, 35, 966-985. https://doi.org/10.1007/s11292-019-09409-7

Cross, C., Kelly, M. (2016). The problem of "white noise": examining current prevention approaches to online fraud. Journal of Financial Crime, 23(4), 806-818. http://dx.doi.org/10.1108/JFC-12-2015-0069

CSAM Deterrence Centre. (2025). https://www.csamdeterrence.com/

ECPAT International. (2021). Factsheet: What are Splash Pages? https://ecpat.org/wp-content/uploads/2021/05/IT-Factsheet-What-are-splash-pages.pdf

Edwards, G., Christensen, L., Rayment-McHugh, S., & Jones, C. (2021). Cyber strategies used to combat child sexual abuse material. Trends & issues in crime and criminal justice, 636. https://doi.org/10.52922/ti78313

Europol. (n.d.). Police2Peer. https://www.europol.europa.eu/partners-collaboration/police2peer

Gallopel-Morvan, K., & Pantin-Sohier, G. (2019). Dissuasive cigarettes: which cues are the most effective at deterring young people from smoking?. Public Health, 174, 22-30. https://doi.org/10.1016/J.PUHE.2019.05.034

Garrett, E., Sauer, J.D., Roehrer, E., Prichard, J. & Scanlan, J. 2025 (in press). Adolescent Perceptions of Warning Messages to Counter Online Sexual Abuse. (non-academic summary: https://www.csamdeterrence.com/adolescent-perceptions-of-warning-messages/)

Google. (2021). Our efforts to fight child sexual abuse online. https://blog.google/technology/safety-security/our-efforts-fight-child-sexual-abuse-online/

Henry, C. (2020). Designing Effective Digital Advertisements to Prevent Online Consumption of Child Sexual Exploitation Material. Journal of Child Sexual Abuse, 29(8), 877-899. https://doi.org/10.1080/10538712.2020.1841354

Hunn, C., Watters, P., Prichard, J., Wortley, R., Scanlan, J., Spiranovic, C., & Krone, T. (2023). How to implement online warnings to prevent the use of child sexual abuse material. Trends & issues in crime and criminal justice, 669. https://doi.org/10.52922/ti78894

Insoll, T., Soloveva, V., Díaz Bethencourt, E., Ovaska, A. K., Nurmi, J., Paju, A., Aaltonen, M., & Vaaranen-Valkonen, N. (2024). Factors Associated with Help-Seeking Among Online Child Sexual Abuse Material Offenders: Results of an Anonymous Survey on the Dark Web. Journal of Online Trust and Safety, 2(4). https://doi.org/10.54501/jots.v2i4.205

Internet Watch Foundation. (n.d.). URL Blocking: Good Practice. https://www.iwf.org.uk/our-technology/our-services/url-list/url-blocking-good-practice/

Knack, N., Winder, B., Murphy, L., & Fedoroff, J. P. (2019). Primary and secondary prevention of child sexual abuse. International review of psychiatry, 31(2), 181–194. https://doi.org/10.1080/09540261.2018.1541872

Lenorovitz, D. R., Leonard, S. D., & Karnes, E. W. (2012). Ratings checklist for warnings: a prototype tool to aid experts in the adequacy evaluation of proposed or existing warnings. WORK: A Journal of Prevention, Assessment & Rehabilitation, 41(S1), 3616-3623. https://doi.org/10.3233/WOR-2012-0114-3616

Meta. (n.d.). Analyzing the intention behind harm in order to prevent it. https://about.meta.com/actions/safety/onlinechildprotection/offenders/

Ó Ciardha, C., Ildeniz, G., & Ruisch, B.C. (in preparation). Message content and framing influences perceived effectiveness of warning messages for child sexual abuse material.

Ofcom. (2023). Protecting people from illegal harms online Annex 8: Illegal Content Codes of Practice for search services. https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/270826-consultation-protecting-people-from-illegal-content-online/associated-documents/annex-8-draft-illegal-content-codes-of-practice-for-search-services?v=330406

Pornhub. (2025). Trust and Safety Initiatives. https://help.pornhub.com/hc/en-us/articles/19159382668435-Trust-and-Safety-Initiatives

Price, S., McKillop, N., Scanlan, J., Rayment-McHugh, S., Christensen, L., & Prichard, J. (2024). A Review of Digital Interventions as Secondary Prevention Measures to Combat Online Child Sexual Abuse Perpetration. Journal of Child Sexual Abuse, 33(7), 869-890. https://doi.org/10.1080/10538712.2024.2415549

Prichard, J., Krone, T., Spiranovic, C., & Watters, P. (2019). Transdisciplinary research in virtual space: Can online warning messages reduce engagement with child exploitation material? In R. Wortley, A. Sidebottom, N. Tilley & G. Laycock (Eds.), Routledge handbook of crime science. London: Routledge: 309–319. https://doi.org/10.4324/9780203431405

Prichard, J., Wortley, R., Watters, P. A., Spiranovic, C., Hunn, C., & Krone, T. (2021). Effects of Automated Messages on Internet Users Attempting to Access "Barely Legal" Pornography. Sexual Abuse, 34(1), 106-124. https://doi.org/10.1177/10790632211013809 (Original work published 2022)

Prichard, J., Scanlan, J., Krone, T., Spiranovic, C., Watters, P. & Wortley, R. (2022a). Warning messages to prevent illegal sharing of sexual images: Results of a randomised controlled experiment. Trends & issues in crime and criminal justice, 647. https://doi.org/10.52922/ti78559

Prichard, J., Scanlan, J., Watters, P., Wortley, R., Hunn, C., & Garrett, E. (2022b). Online messages to reduce users' engagement with child sexual abuse material: a review of relevant literature for the reThink chatbot. University of Tasmania. Report. https://hdl.handle.net/102.100.100/502869

Prichard, J., Wortley, R., Watters, P., Spiranovic, C., & Scanlan, J. (2024). The effect of therapeutic and deterrent messages on Internet users attempting to access 'barely legal' pornography. Child Abuse & Neglect, 155: https://doi.org/10.1016/j.chiabu.2024.106955

Quayle, E. (2020). Prevention, disruption and deterrence of online child sexual exploitation and abuse. ERA Forum 21, 429–447. https://doi.org/10.1007/s12027-020-00625-7

Quayle, E., & Koukopoulos, N. (2019). Deterrence of online child sexual abuse and exploitation. Policing: A Journal of Policy and Practice 13(3): 345–362. https://doi.org/10.1093/police/pay028

Roy, A., & Hunt, G. (2025). Redesigning Google's CSAM prevention and deterrence intervention on Search [Presentation]. Talking for Change: Preventing Child Sexual Abuse & Use of CSAM, 13th of June 13, Toronto, ON, Canada. https://talkingforchange.ca/event/

Scanlan, J., Watters, P., Prichard, J., Hunn, C., Spiranovic, C., & Wortley, R. (2022). Creating Honeypots to Prevent Online Child Exploitation. Future Internet 14(4). https://www.mdpi.com/1999-5903/14/4/121

Scanlan, J., Prichard, J., Hall, L. C., Watters, P., & Wortley, R. (2024). 'reThink Chatbot evaluation', University of Tasmania, Hobart. ISBN 978–1–922708-67-0. https://www.lucyfaithfull.org.uk/files/reThink_Chatbot_Evaluation_Report.pdf

Smallbone, S. & Wortley, R. (2017). Preventing child sexual abuse online. In J Brown (ed), Online risk to children: Impact, protection and prevention. Wiley: 143. https://doi.org/10.1002/9781118977545.ch8

Taylor, J. R., & Wogalter, M. S. (2019). Specific egress directives enhance print and speech fire warnings. Applied ergonomics, 80, 57-66. https://doi.org/10.1016/j.apergo.2019.04.009

Ullman, J. R., & Silver, N. C. (2018). Perceived Effectiveness of Potential Music Piracy Warnings. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62(1), 1353-1357. https://doi.org/10.1177/1541931218621309

Walsh, M., Denis, D., & Findlater, D. (2023). Deterring online child sexual abuse and exploitation: lessons from seven years of campaigning. Lucy Faithfull Foundation. https://www.lucyfaithfull.org.uk/research/deterring-online-child-sexual-abuse-and-exploitation-lessons-from-seven-years-of-campaigning/

Wogalter, M.S., Laughery, K. R., & Mayhorn, C. B. (2012). Warnings and hazard communications (Chap. 29). In G. Salvendy (Ed.), Handbook of Human Factors / Ergonomics (4th ed.). New York: Wiley. https://www.safetyhumanfactors.org/wp-content/uploads/2020/07/342)Wogalter,Laughery,Mayhorn(2012).pdf

Wortley, R. (2012). Situational prevention of child abuse in the new technologies. In K. Ribisl & E. Quayle, Preventing Online Exploitation of Children. London: Routledge. https://www.researchgate.net/publication/316232775

Wortley, R., Findlater, D., Bailey, A., & Zuhair, D. (2024). Accessing child sexual abuse material: Pathways to offending and online behaviour. Child Abuse & Neglect, 154. https://doi.org/10.1016/j.chiabu.2024.106936

Wortley, R., & Smallbone, S. (2006). Applying situational principles to sexual offenses against children. https://www.researchgate.net/publication/294621 96_Applying_situational_principles_to_sexual_offe nses_against_children

Wortley, R., & Prichard, J. (2023). Online Messaging as a Cybercrime Prevention Tool in the Postpandemic Age. In R.G. Smith, R. Sarre, L.YC. Chang, & L.YC Lau (Eds.) Cybercrime in the Pandemic Digital Age and Beyond. Palgrave Studies in Cybercrime and Cybersecurity (209-232). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-29107-4_11

Protect Children

Cite this report

Contributors

This report was prepared for Protect Children by Tegan Insoll[1], Dr Joel Scanlan[2], Dr Caoilte Ó Ciardha[3], Dr Juha Nurmi[4], Valeriia Soloveva[1], Emily May[1], Anna Ovaska[1], and Nina Vaaranen-Valkonen[1].

[1] Protect Children
[2] CSAM Deterrence Centre, University of Tasmania
[3] University of Kent
[4] Faculty of Information Technology and Communication Sciences, Tampere University