safer **BUILT BY THORN**

Protect your platform from hosting new CSAM

Detect child sexual abuse images, video and text

Empower your trust and safety team with Safer Predict, the state-of-the-art software solution designed to help protect your content-hosting platform from the risks of hosting child sexual abuse material (CSAM) and text-based interactions that could lead to child sexual exploitation (CSE). Gain access to powerful AI technology that proactively detects these risks, creating a safer digital environment for your users.

Why choose Safer Predict?

Enhanced protection

Safer Predict's predictive AI technology identifies new or previously unreported instances of CSAM and CSE, ensuring that harmful content is swiftly detected and can be acted upon.

Trusted training data

Thorn's machine learning classification models were trained using high-quality data from trusted sources – in part using trusted data from the National Center for Missing and Exploited Children's (NCMEC) CyberTipline.

Efficient moderation

Streamline your content moderation processes by empowering moderators to conduct in-depth investigations of pertinent results, prioritize actionable content, and report harmful material more efficiently.

Customizable workflows

Tailor Safer Predict to fit your platform's specific needs with highly customizable workflows. Develop strategic detection plans, prioritize high-risk accounts, and expand CSAM and CSE detection coverage.
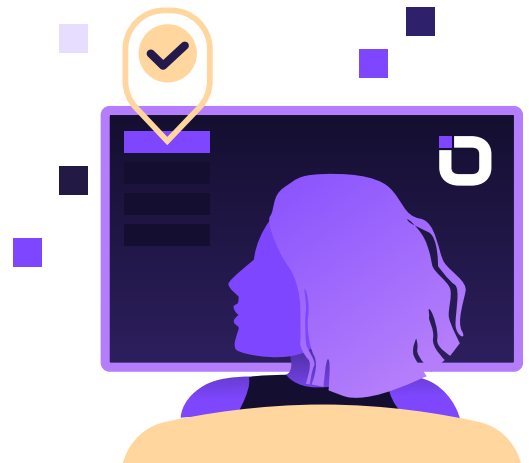
Deployment options

Thorn offers flexibility and control as you scale your CSAM detection whether you choose a self-hosted or Thorn-hosted solution.

WHY THORN

While other detection software providers address many types of harm, Thorn is singularly focused with the world's largest team solely dedicated to building technology to defend children from sexual abuse.
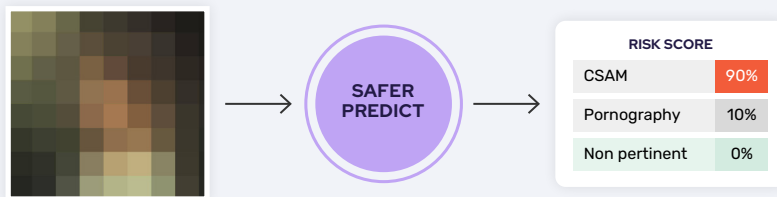
Safer Predict classification models are trained on actual CSAM and abuse text. Models that are trained simply on pornography and face/age are far less effective in detecting CSAM and child sexual exploitation.
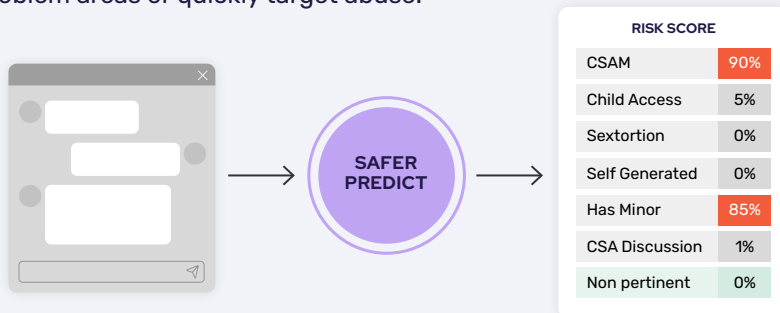
POWERED BY PREDICTIVE AI

Safer Predict leverages machine learning (ML) classification models (a.k.a. "classifiers") to detect new or previously unreported instances of CSAM and CSE. Send images, video, or text for analysis to our classifiers. A risk score is generated for each item, signifying how likely the content is to be violative, providing indicators and a description of the identified violations to make human decisions easier and faster.

The **Safer Predict CSAM classifier** identifies new or previously unreported CSAM by analyzing thousands of attributes within images and videos to predict the presence of CSAM. Content is sorted into three classes: CSAM, pornography, and benign. Potential CSAM is queued for human review.

RISK SCORE

CSAM	90%
Pornography	10%
Non pertinent	0%

The **Safer Predict CSA text classifier** identifies abuse based on the context of the conversation. The classifier is built and trained on verified CSE data and allows users to 'stack' multiple labels to narrow down problem areas or quickly target abuse.

RISK SCORE

CSAM	90%
Child Access	5%
Sextortion	0%
Self Generated	0%
Has Minor	85%
CSA Discussion	1%
Non pertinent	0%

MAXIMIZE YOUR INVESTMENT

Safer Predict provides highly customizable workflows that allow you to develop a strategic detection plan, focus in on high-risk accounts, and expand child sexual abuse detection coverage. Prioritize and escalate content by setting a precision level for detection, and leverage the provided label for pertinent content. In addition to detecting new content, Safer Predict can help you zero in on areas of high risk.

- ☑ Scan high-risk accounts.
- ☑ Conduct more in-depth investigations.
- ☑ Clear out moderator queues more efficiently.
- ☑ Reduce visibility of potential abuse content while waiting for manual review.
- ☑ Create custom workflows based on platform policies.
- ☑ Allocate technical resources as needed.

Get started today

Safer Predict provides a layer of protection for your company's brand, reputation, and user experience — with the tools your trust and safety team needs to effectively combat CSAM and CSE.

Ready to take your trust and safety efforts to the next level? Contact us today to learn more. Together, we can build a safer digital future for all.

safer. BUILT BY THORN