

American University Washington College of Law

Digital Commons @ American University Washington College of Law

Joint PIJIP/TLS Research Paper Series

2022

Protecting Children in the Age of End-to-End Encryption

Laura Draper

American University Washington College of Law, ldraper@wcl.american.edu

Follow this and additional works at: <https://digitalcommons.wcl.american.edu/research>



Part of the [Intellectual Property Law Commons](#), [International Trade Law Commons](#), and the [Internet Law Commons](#)

Recommended Citation

Draper, Laura, "Protecting Children in the Age of End-to-End Encryption" (2022). *Joint PIJIP/TLS Research Paper Series*. 80.

<https://digitalcommons.wcl.american.edu/research/80>

This Article is brought to you for free and open access by the Program on Information Justice and Intellectual Property and Technology, Law, & Security Program at Digital Commons @ American University Washington College of Law. It has been accepted for inclusion in Joint PIJIP/TLS Research Paper Series by an authorized administrator of Digital Commons @ American University Washington College of Law. For more information, please contact DCRepository@wcl.american.edu.

Protecting Children in the Age of End-to-End Encryption

Laura Draper

Senior Project Director
Tech, Law & Security Program

Washington College of Law
American University

This project received funding support from Meta and the Silicon Valley Community Foundation. TLS maintains strict intellectual independence and sole editorial direction and control over its intellectual property, ideas, projects, publications, events, and other research activities. Consistent with TLS policy, the content of this report reflects the views of its author alone.



TECH, LAW & SECURITY
PROGRAM

TABLE OF CONTENTS

EXECUTIVE SUMMARY	i
ACKNOWLEDGMENTS	iii
INTRODUCTION	1
PART 1: UNDERSTANDING THE ISSUE	4
PART 2: UNPACKING THE HARM	9
<i>I. CSAM Production and Distribution</i>	10
<i>II. Perceived First Person (PFP) Material Production and Distribution</i>	12
<i>III. Internet-Enabled Domestic Child Sex Trafficking</i>	16
<i>IV. Live Online Child Sexual Abuse & Sexual Exploitation of Children in the Context of Travel and Tourism</i>	17
PART 3: INTERVENTIONS	21
<i>I. Prevent</i>	23
<i>II. Detect</i>	26
<i>III. Deter</i>	37
<i>IV. Refer</i>	38
<i>V. Disrupt</i>	39
RECOMMENDATIONS	42
<i>I. Improve the Report-to-Prosecution Pipeline</i>	42
<i>II. Engage Relevant Stakeholders</i>	45
<i>III. Prioritize Upstream Efforts</i>	47
APPENDIX 1: TERMINOLOGY	49
APPENDIX 2: RELEVANT U.S. STATUTES	55
<i>I. CSAM Production and Distribution</i>	55
<i>II. Perceived First Person Material Production and Distribution</i>	57
<i>III. Internet-Enabled Domestic Child Sex Trafficking</i>	59
<i>IV. Live Online Child Sexual Abuse & Sexual Exploitation of Children in the Context of Travel and Tourism</i>	62
APPENDIX 3: SENTENCING FACTORS RECOMMENDED BY U.S. SENTENCING COMMISSION	64

EXECUTIVE SUMMARY

The sexual exploitation and abuse of children did not start with the internet, but advances in digital and mobile technology have enabled and facilitated related crimes. Social media platforms make it easier for offenders to find vulnerable children. Cell phone cameras and webcams make it easier for offenders to record the exploitation and abuse. Those same technologies enable offenders to live stream the sexual abuse of a child from the other side of the globe. Cloud storage makes it easier for offenders to hoard large caches of illicit images.

Now, end-to-end encryption allows offenders to trade these images with less fear of detection. “End-to-end encryption” has moved from a buzzword among tech-savvy individuals to a default setting in many communications systems. Whereas offenders previously needed the technical know-how to access the dark web in order to find and trade child sexual abuse material (CSAM),¹ they can now download to their cell phones any number of free end-to-end encrypted messaging apps, which provide the same level of security and privacy as the dark web. This greatly reduces the risk to the offender when sharing such material.

It is easy to let conversations about online child sexual exploitation and abuse devolve into no-win arguments about the merits (or lack thereof) of end-to-end encryption. Law enforcement officials bemoan that this technology leads to criminals “going dark,” meaning they can communicate in places that police cannot access, even with a warrant. Privacy advocates cry that the technology is necessary and dismiss law enforcement’s “but the children” arguments as fearmongering. This report sidesteps this debate by simply assuming, without judgment, that end-to-end encryption is here to stay, and asks, **how are we going to combat online child sexual exploitation and abuse?**

This report examines this problem in depth.

First, it discusses trends in online child sexual exploitation and abuse, it explains the impact of end-to-end encryption, and it outlines the current process of detection, investigation, and prosecution. The existing system, in which reports of these crimes flow from tech companies to a national clearinghouse to law enforcement, is filled with multiple checks and balances, including human reviewers. The universal adoption of end-to-end encryption across communications platforms will make it more difficult for tech companies to detect these crimes, especially the trading of CSAM.

Second, this report details the most common patterns of harm. “Online child sexual exploitation and abuse” is often treated as one specific category of crime: recording the sexual abuse of children and sharing those recordings. This narrow definition fails to account for the multitude of harms that

¹ Previously known as “child pornography,” “child sexual abuse material” is now the more widely accepted term to describe this category of imagery. To clarify, CSAM is only one element of child sexual abuse and exploitation.

fall under this umbrella term, including sexually explicit images of children in which the children themselves appear to record the images, domestic child sex trafficking, live streams of child sexual abuse, and the sexual exploitation of children in the context of travel and tourism. Because the problem is broad in scope and varied in nature, the interventions necessary to combat it must be similarly broad and bespoke.

Third, this report explores the range of ways to intervene in these patterns of harm. There are methods focused at preventing the abuse and exploitation, including targeted education campaigns and in-person supports and programming, as well as ways to detect the harm once it has occurred and to minimize the negative effects. Because the internet facilitates or enables these crimes, it is easy to focus on technical solutions. While the report presents and considers technical options, including metadata analysis and artificial intelligence, the report also explores interventions that fall within the ambit of civil society and/or local government, including clinical treatment and undercover police operations.

The report ends with recommendations aimed at decision-makers who are committed to combating online child sexual exploitation and abuse. The recommendations include:

- **Improve the Report-to-Prosecution Pipeline**, by setting industry standards for data collection and retention; creating uniform criteria for lawful data requests; streamlining law enforcement's reporting; establishing guidelines for triaging reports; and addressing victims' needs throughout the process. This process is central in the fight against online child sexual exploitation and abuse, and must therefore operate efficiently.
- **Engage Relevant Stakeholders**, by identifying who can intervene; investing in partnerships; leveraging incentive structures; and exploring alternative sources of signals suggesting malicious conduct. Addressing this problem requires a whole-of-community approach; policymakers must think broadly to assemble the right team.
- **Prioritize Upstream Efforts**, by investing in community-based opportunities and researching tech-based interventions. End-to-end encryption creates a black box around communications in which all parties want to keep private; effective responses therefore must occur earlier in the process.

The underlying offenses are not new, but advances in mobile and digital technology have changed the way they manifest, and technology itself is constantly evolving. To combat online child sexual exploitation and abuse, we must also evolve.

ACKNOWLEDGMENTS

This project was first conceived of by Jennifer Daskal, who then served as the faculty director of the Tech, Law & Security Program (TLS) at the Washington College of Law, American University. Great debts are also owed to Gary Corn, TLS program director, and Alex Joel, senior project director and TLS team leader, for their wisdom and feedback throughout the life of this project. Jenna Ruddock and Kady Hammer provided invaluable support, both in researching related topics and in discussing how to process and organize the vast amount of material. TLS senior fellow Melanie Teplinsky also provided feedback throughout the project and planted the idea of visual depictions of the patterns of harm. J. Peter Scoblic’s editorial expertise was necessary to make this report readable.

This report would not exist but for the immense amount of time graciously provided by subject matter experts from a wide variety of stakeholders. These interviews—around 70 in total—generated the core content of the report. Without the gracious gift of their time, this report would not be as robust, nuanced, or simply as smart. Those interviewees who agreed to be identified are listed here, along with the entity with which they were affiliated at the time of the interview(s).²

Kendra Albert, *Harvard Law School*

Dan Gomez, *TacLogix*

Arif Alikhan, *TacLogix*

Maggie Goodrich, *TacLogix*

Jacqueline Beauchere, *Microsoft (currently Snap Inc.)*

Victoria Green, *Marie Collins Foundation*

Emma Hardy, *Internet Watch Foundation*

Austin Berrier, *Homeland Security Investigations, U.S. Department of Homeland Security*

Laura Harmon, *King County Prosecuting Attorney’s Office*

Kelly Crouch, *Seattle Police Department*

Carol Hepburn, *Hepburn Law Firm*

Kate D’Adamo, *Reframe Health and Justice*

Laura Higgins, *Community Safety, Roblox*

James Duke, *New York City Police Department*

Denton Howard, *InHope*

Hany Farid, *University of California, Berkeley*

Stacy Irving, *Delaware Valley Intelligence Center, Philadelphia Police Department*

Tom Farrell, *SafeToNet*

Brendan Ittelson, *Zoom*

Chris Fisher, *Seattle Police Department*

Janani Iyengar

Mary Anne Franks, *University of Miami School of Law*

Brandon James, *Seattle Police Department*

Guillermo Galarza Abizaid, *International Centre for Missing & Exploited Children*

Brandon Kaopuiki, *International Justice Mission*

Debbie Garner, *Georgia Bureau of Investigation*

Konstantinos Komaitis

Daniel Kahn Gillmor, *American Civil Liberties Union*

Chelsea Komlo, *University of Waterloo*

² Affiliations are provided for identification purposes only, and do not indicate the individual was speaking on the entity’s behalf.

Sean Litton, *Tech Coalition*

Vicki McDermott, *New York City Police Department*

Lianna McDonald, *Canadian Centre for Child Protection*

Fallon McNulty, *National Center for Missing & Exploited Children*

Tanya Meisenholder, *New York City Police Department*

Alec Muffett

Beda Mohanty, *Meta*

Jonathan Lee, *Meta*

Karuna Nain, *Meta*

Blake Norton, *Philadelphia Police Department*

Chris O'Connell, *New York City Police Department*

Marcos Ortiz, *Seattle Police Department*

Josh Parecki, *Zoom*

Riana Pfefferkorn, *Stanford Internet Observatory*

Lloyd Richardson, *Canadian Centre for Child Protection*

Jon Rouse, *Queensland Police Service, Task Force Argos*

Jenna Ruddock, *Tech, Law & Security Program, Washington College of Law*

Chloe Setter, *WeProtect Global Alliance*

Dan Sexton, *Internet Watch Foundation*

John Shehan, *National Center for Missing & Exploited Children*

Justin Sherman, *Tech, Law & Security Program, Washington College of Law*

Walter Smith, *Philadelphia Police Department*

Thomas Sullivan, *Metropolitan (D.C.) Police Department*

Dhanaraj Thakur, *Center for Democracy and Technology*

INTRODUCTION

Child sexual exploitation and abuse—rape of a child, child sex trafficking, child sexual abuse material—have forever been a scourge on society. The internet did not create these crimes, but it did make it easier to commit these offenses: easier for offenders to identify and target vulnerable children, easier for offenders to find and create child sexual abuse material, easier for offenders to trade these images among themselves, and easier for offenders to hoard these images.

As technology evolves, so does the problem. A major evolution in information and communication technologies (ICTs)³—such as social media and messaging apps—is the steady trend toward the adoption of end-to-end encryption. Whereas offenders once needed to understand how to access the recesses of the dark web to anonymously find and trade child sexual abuse material (CSAM),⁴ they can now simply download to their cell phone a free messaging application with end-to-end encryption or log on to an end-to-end encrypted live stream to view child sexual exploitation and abuse (CSEA). These technological changes have profound implications for ways to combat the problem. This report examines these implications and provides necessary context to identify the best ways to intervene.

Discussions of how to combat online child sexual exploitation and abuse often morph into debate over the wisdom of end-to-end encryption, which is a method of secure communication that prevents third parties from accessing content while it is transferred from one system or device to another.⁵ End-to-end encryption is necessary to protect users' privacy, including that of human rights activists and journalists located in authoritarian states, *and* end-to-end encryption creates a black hole where offenders can trade illicit images of abused children with impunity. Both statements are true, which is why the debate on the propriety of end-to-end encryption is ongoing (and possibly unsettle-able). In the context of online CSEA, the debate often treats privacy and child safety as mutually exclusive concepts and pits them against each other.

This report avoids this debate by acknowledging that end-to-end encryption is or will become the default across information and communication technologies. The question motivating this project is how can stakeholders—tech companies, law enforcement, civil society—be held accountable for their commitment to combat online CSEA given the increasing adoption and proliferation of end-to-end encryption?

This project seeks to identify and assess the challenges in combating online CSEA, as well as the technological tools, partnerships, and other innovative approaches that can be employed to intervene.

3 Throughout this report, “information and communication technologies” is the general term used to describe user-facing technologies—i.e., social media platforms, messaging apps.

4 “Child sexual abuse material,” or CSAM, is the term typically used to replace what has otherwise been known as “child pornography,” due to the widely accepted argument that “sexualised material that depicts or otherwise represents children is indeed a representation, and a form, of child sexual abuse.” Interagency Working Group, *Terminology Guidelines for the Protection of Children from Sexual Exploitation and Sexual Abuse* (herein *Luxembourg Guidelines*), 2016, 38. For a more nuanced discussion of this term, see Appendix 1.

5 Ben Lutkevich and Madelyn Bacon, “Definition: end-to-end encryption (E2EE),” SearchSecurity, last modified June 2021, <https://www.techtarget.com/searchsecurity/definition/end-to-end-encryption-E2EE>. For a more technical and nuanced explanation of end-to-end encryption, see Mallory Knodel, Sofia Celi, Fred Baker, Olaf Kolkman, and Gurshabad Grover, “Definition of End-to-end Encryption,” IETF Datatracker, last modified September 28, 2022, <https://datatracker.ietf.org/doc/draft-knodel-e2ee-definition>.

Although end-to-end encryption protects data and communications in transit, it does *not* prevent exploitation at the end point. For example, if a third party accesses a user's device, either lawfully (i.e., with a search warrant) or unlawfully (i.e., via hacking), that third party can access the content sent and received by that device, regardless of the encryption standards applied to the communication method itself. Similarly, if either the sender or the intended recipient elects to self-report the content voluntarily, that is another means by which a third party may access the content of an end-to-end encrypted communication.

The report is divided into three primary sections.

1. **Understanding the Issue** provides background on CSEA and the impact of the internet. This section also provides a primer on how tech companies, law enforcement, and the public collaborate, and it describes the central role of the National Center for Missing & Exploited Children (NCMEC), a statutorily created, nonprofit clearinghouse for matters related to child protection.
2. **Unpacking the Harm** explains the scope of the problem and identifies different ways children are harmed through online CSEA.
3. **Interventions** outlines a list of interventions that can help reduce overall harm by preventing the abuse or exploitation in the first instance, or by detecting it once it has occurred and then either deterring escalation and further abuse, referring the offender or victim to support services, and/or disrupting the pattern of harm.

The report closes with a series of recommendations on ways to improve the data flow among tech companies, law enforcement, and NCMEC; how to engage relevant stakeholders; and the need to prioritize upstream interventions.

There are also three appendixes. The first appendix defines relevant terms and explains the reason for their use over other similar phrases. The second appendix connects particular sections of the U.S. Criminal Code that criminalize different actions discussed herein. The list is not exhaustive but provides some guidance as to the relevant statutory language. The third appendix describes what the U.S. Sentencing Commission considers relevant when imposing a criminal sentence for someone convicted of one of these offenses.

A Note on Language

Terminology around CSEA is particularly fraught, and even experts working on this issue may disagree about the best way to describe particular elements of the offenses. There are several reasons for this.

First, child sexual exploitation and abuse is a very difficult and emotionally laden topic. While a dry, clinical approach to language may at times be helpful to provide some emotional distance, there is a risk of ignoring the very real harms experienced by very real victims. It is important to balance both these considerations so that people who work in this space can maintain their mental health and also remember why this work is so important.

Second, this is a global problem. It is not uncommon for victims, offenders, and the electronic service provider (ESP)⁶ to be based in three separate countries. As a result, multiple legal jurisdictions are implicated, and with that, multiple sets of legal terms are used.

Third, a lot of terms that have been used historically are now understood to stigmatize, minimize, or blame the victim. Contemporary understanding of the impact of this language must also be taken into account.

Ultimately, this report seeks to balance all these considerations in selection and use of language, and undoubtedly fails in many instances. The language used throughout

⁶ In this report, “electronic service providers” (ESPs) is the general term used for companies that provide information and communication technologies—e.g., Facebook, WhatsApp, Google, Kik.

this report is intended to be broadly accessible so that readers can stay engaged while appreciating the nuances in the different issues. At first use of a term, a definition will be provided, and Appendix 1 provides additional context and explanation for readers interested in learning more about the choice of that term.

This report primarily focuses on U.S. law and policy, although lessons from other countries are incorporated and suggested interventions may be applicable elsewhere. It is intended for anyone who cares about this issue and wants to do more to combat it—decision-makers and policymakers at tech companies, in civil society, and at law enforcement agencies, as well as officials in federal, state, and local government.

PART 1: UNDERSTANDING THE ISSUE

The internet did not create child sexual exploitation and abuse, but it has exacerbated it.⁷ The growth of digital and mobile technology has resulted in roughly a fivefold increase in the number of people convicted for creating child sexual abuse material (CSAM) over a 15-year period.⁸ Technology has also made it easier for offenders to find vulnerable children to exploit: more than one-third of people sentenced in 2019 for producing CSAM met their victims through an online platform, more than double the proportion 10 years earlier.⁹

Before the internet, possession of CSAM was space limited; offenders could physically possess only what they could covertly store. Now, with advances in electronic storage, that limit does not exist, and as a result, nonproduction offenses (i.e., possession, receipt, and distribution) increasingly involve large numbers of videos and images. Nonproduction offenses in 2019 involved a median of more than 4,000 images, and some offenders possessed and distributed millions of images and videos.¹⁰ To be clear, these images show the abuse of the most vulnerable victims: over half of nonproduction offenses in that same time frame included images of infants or toddlers, and nearly every offense (99.4%) included prepubescent victims.¹¹

Measuring the global scope of child sexual exploitation and abuse with accuracy is extraordinarily difficult due to variations in definitions and data collection practices across jurisdictions.¹² Here is what can be stated with certainty: in 2021, NCMEC¹³ received over 29 million reports of suspected child sexual exploitation, up from 21.7 million reports in 2020, and more than 10 times the number received a decade prior.¹⁴

This dramatic increase in reports is not exclusively due to a concomitant increase in hands-on abuse or distribution of CSAM (although increases in these activities may account for some of the change). Instead, this increase is likely due, in part, to ESPs adopting highly efficient detection tools.¹⁵ PhotoDNA, a CSAM detection tool that relies on “perceptual hashing” technology, was developed in

7 Although CSEA occurred prior to the advent of the internet, there are certain forms of abuse and exploitation that are possible only because of current technology, such as live-streamed abuse, discussed in greater detail below. For a discussion on cyber-enabled versus cyber-dependent crimes, see ECPAT International, *Trends in Online Child Sexual Abuse Material*, April 2018, 6, <https://ecpat.org/wp-content/uploads/2021/05/ECPAT-International-Report-Trends-in-Online-Child-Sexual-Abuse-Material-2018.pdf>.

8 U.S. Sentencing Commission, *Federal Sentencing of Child Pornography: Production Offenses*, October 2021, 3, <https://www.ussc.gov/research/research-reports/federal-sentencing-child-pornography-production-offenses>.
The U.S. Sentencing Commission’s report notes that “child pornography production cases comprise a small percentage of the overall federal caseload,” and so the 422% increase is calculated based on 98 offenders sentenced in fiscal year 2005 rising to 512 offenders sentenced in fiscal year 2019.

Under U.S. law, CSAM-related offenses are often bucketed in the following way, with increasing penalties attached: (1) possession, (2) receipt and/or distribution, and (3) production. While possession and receipt have commonly understood definitions, distribution and production are legally broader than one might expect. For more thorough discussions of these terms, see Appendix 1.

9 U.S. Sentencing Commission, *Federal Sentencing of Child Pornography: Production Offenses*, 5. “[O]ver one-third (35.4%) of production offenders sentenced in fiscal year 2019 were internet strangers who met their victims through an online platform, more than double the proportion of offenders sentenced in fiscal year 2010 who met their victims online (14.3%).”

10 U.S. Sentencing Commission, *Federal Sentencing of Child Pornography: Non-Production Offenses*, June 2021, 4, <https://www.ussc.gov/research/research-reports/federal-sentencing-child-pornography-non-production-offenses>.

11 U.S. Sentencing Commission, *Federal Sentencing of Child Pornography: Non-Production Offenses*, 4.

12 See, e.g., ECPAT International, *Trends in Online Child Sexual Abuse Material*.

13 According to its website, the National Center for Missing & Exploited Children (NCMEC) “is a private, non-profit 501(c)(3) corporation whose mission is to ... reduce child sexual exploitation, and prevent child victimization.” U.S. law mandates that electronic service providers report CSEA detected on their platforms. This organization is described in greater detail in Appendix 1

14 “Our 2021 Impact,” National Center for Missing & Exploited Children, accessed September 24, 2022, <https://www.missingkids.org/content/ncmec/en/ourwork/impact.html>.

15 *The EARN IT Act: Holding Tech Industry Accountable in the Fight Against Online Child Sexual Exploitation*, U.S. Senate Committee on the Judiciary, 4 (2020) (testimony of John Shehan, vice president, Exploited Children Division, NCMEC) <https://www.judiciary.senate.gov/imo/media/doc/Shehan%20Testimony.pdf>.

2009 and has been widely adopted throughout the tech industry.¹⁶ ESPs use this program, as well as programs like it, to automatically scan content on their platforms to detect CSAM. This method has proved extremely accurate, reliable, and fast.¹⁷ It is this process that Meta has used with Facebook Messenger and that is credited for the large amount of reporting by Meta to NCMEC; in 2021, Facebook made 22.1 million reports (out of the 29.2 million received total from ESPs) to NCMEC.¹⁸

In end-to-end encrypted environments, ESPs cannot detect and report CSAM using perceptual hashing techniques. As more and more tech companies implement end-to-end encryption, the volume of reports to NCMEC will likely drop dramatically. Regulations in Europe provided a real-world test case to examine the impact on reporting when hash-matching methods are disrupted. In December 2020, the ePrivacy Directive in the European Union (EU) went into effect and, for a time, limited ESPs' ability to use hash-scanning technology to detect CSAM. NCMEC examined EU-related reports submitted by ESPs in the weeks before and after the directive went into effect and found that the number of reports decreased by 51% in the first six weeks.¹⁹

There is an oft-repeated myth that when an ESP receives notice of a "positive match" alerting to CSAM on its platform, the account holder affiliated with that match is immediately subject to criminal penalties.²⁰ This is false. The process is far more complicated. Understanding the complexities of how information flows from ESPs to NCMEC to police to prosecutors is necessary to understand the broader issue.

Starting with NCMEC, it receives tips in one of four ways:

1. from an ESP upon detection of online CSEA,
2. from a member of the public,
3. from proactive investigative efforts by law enforcement, or
4. from an ESP that identifies the online CSEA once law enforcement serves the provider with legal process (i.e., a subpoena or search warrant).

Currently, NCMEC receives the overwhelming majority of tips from ESPs.²¹ However, as discussed above, as end-to-end encryption continues to proliferate, tech companies' ability to confidently detect CSEA on their systems is expected to diminish, such that the other interventions will become increasingly important.

16 PhotoDNA was developed in 2009 by Microsoft in partnership with Dartmouth College. "PhotoDNA," Microsoft, accessed September 24, 2022, <https://www.microsoft.com/en-us/photodna>.

This program, which has since been donated to NCMEC, relies on "perceptual hashing" of images. Other companies have leveraged this same technique to build similar detection methods, such as Apple's NeuralHash system or Google's CSAI Match. Perceptual hashes and other detection methods are discussed in greater detail in part 3. For a more technical discussion of the various permutations of automated hash scanning, see Ian Levy and Crispin Robinson, *Thoughts on Child Safety on Commodity Platforms*, July 21, 2022, <https://arxiv.org/pdf/2207.09506.pdf>.

17 Levy and Robinson, *Thoughts on Child Safety on Commodity Platforms*, 50–51.

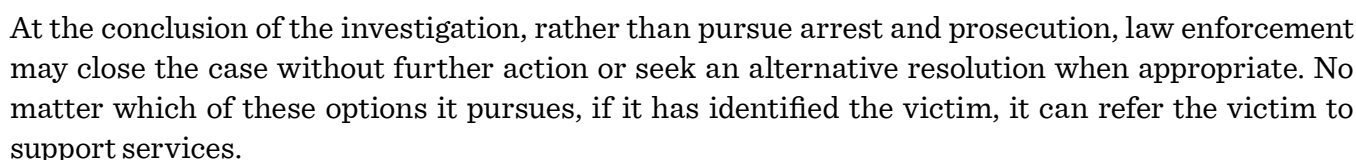
18 NCMEC, *2021 CyberTipline Reports by Electronic Service Providers (ESP)*, 2022, <https://www.missingkids.org/content/dam/missingkids/pdfs/2021-reports-by-esp.pdf>.

19 "European Union CyberTipline Data Snapshot: Reports Submitted by Technology Companies," A Blog Update and EU CyberTipline Data Snapshot, NCMEC, February 17, 2021, <https://www.missingkids.org/blog/2020/we-are-in-danger-of-losing-the-global-battle-for-child-safety>.

20 See, e.g., Kashmir Hill, "A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as a Criminal," *New York Times*, August 21, 2022, <https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html>.

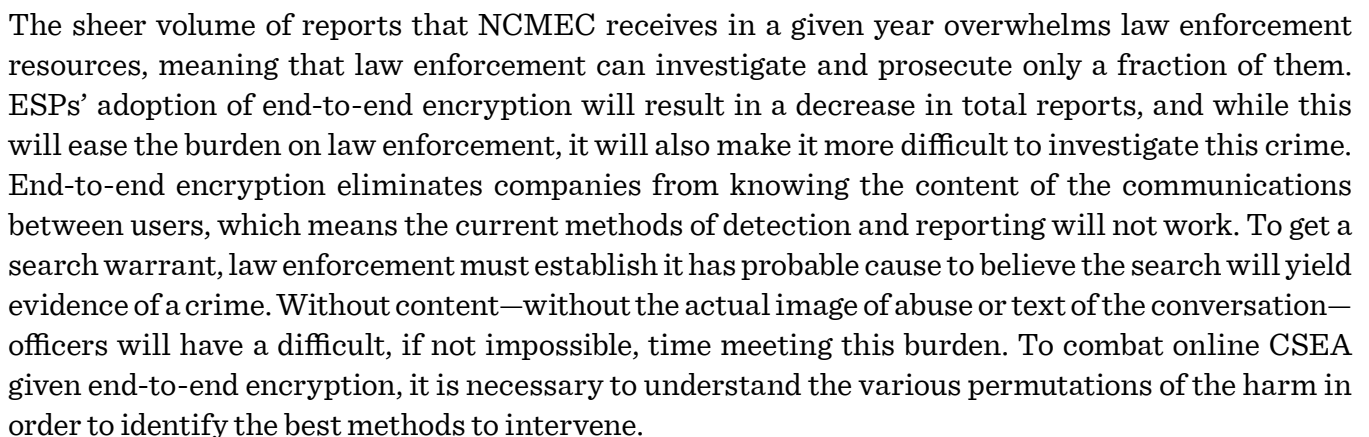
21 NCMEC received 29,397,681 reports in 2021. Of those, 240,598 (less than 1%) were from the public, and the remaining 29,157,083 were from electronic service providers. NCMEC, "Our 2021 Impact."

FIGURE 2: THE ROLE OF LAW ENFORCEMENT²⁴



Protecting Children in the Age of End-to-End Encryption | Fall 2022

FIGURE 3: FROM REPORT TO PROSECUTION



PART 2: UNPACKING THE HARM

Online child sexual exploitation and abuse is often treated as a single problem—the sharing of CSAM among offenders. However, online CSEA extends far beyond offenders sharing CSAM with one another; there are many different ways the internet contributes to and/or facilitates the sexual exploitation and abuse of children. To effectively combat this problem, it is necessary to both zoom out—to grasp the full scope of the problem—and zoom in, to better understand the variations and nuances within the different patterns of harm.

As explored in greater detail in part 3, different interventions are effective at combating online CSEA in different ways at different points in different patterns of harm. This is especially true when discussing the problem in the context of end-to-end encryption, which has the greatest impact at the tail end of most patterns of abuse—specifically, the trading of CSAM among offenders. With the use of end-to-end encryption treated as a given, it is necessary to move upstream in the process, which requires comprehension of the actions that lead to the production and sharing of CSAM.

Online CSEA generally falls into one or more of four patterns (represented graphically and discussed in greater detail below).

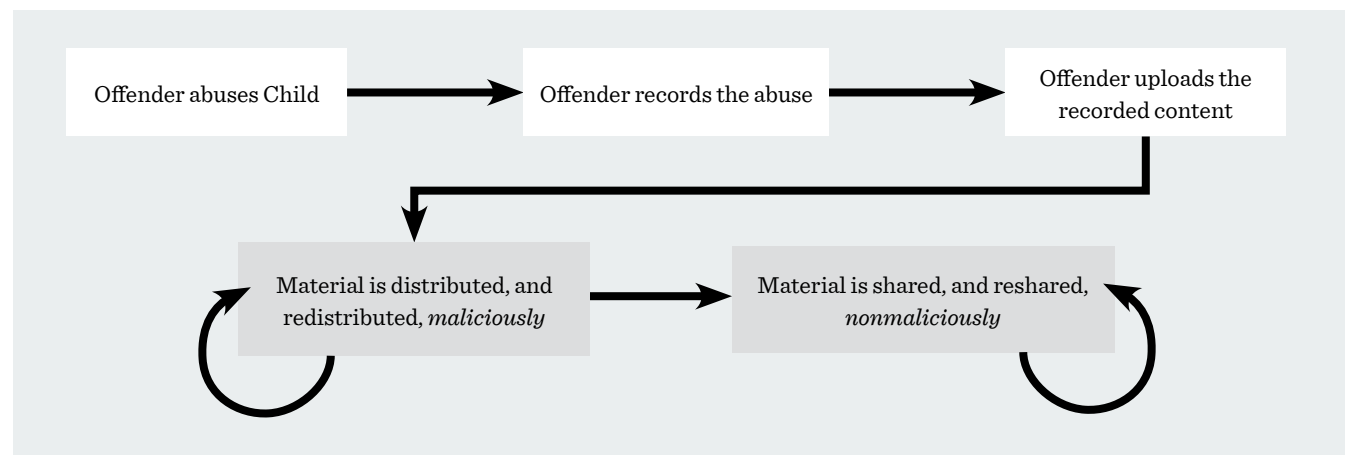
- I. **CSAM production and distribution** is the recording the sexual abuse of a child and then sharing it online. This is what is most commonly (but erroneously) understood to encapsulate online CSEA.
- II. **Perceived first person material production and distribution** refers to the newer trend of explicit images that appear to be created by the child themselves, although the child may have created the images unwittingly or unwillingly and the child may not have known about or intended broader distribution of the imagery.
- III. **Internet-enabled domestic child sex trafficking** describes the ways in which traffickers and demand-side offenders use the internet in the sexual exploitation of children.
- IV. **Live online child sexual abuse and sexual exploitation of children in the context of travel and tourism** outlines two related but distinct patterns of harm in which the child is often located outside the United States and the sexual abuse may be live streamed for paying customers around the world or paying customers may travel to those locations to abuse the child.

These patterns are not mutually exclusive, nor do they represent an exhaustive description of the problem. Instead, they are presented here in an effort to disaggregate the larger issue.

I. CSAM Production and Distribution

This diagram represents what is most commonly understood as online CSEA—the recording of child sexual abuse and the distribution of that recording. Most often, when people speak about CSAM, it is in reference to this pattern. That said, even this pattern is not a monolith and actually represents several distinct harms that present distinct opportunities for intervention.

FIGURE 4: CSAM PRODUCTION AND DISTRIBUTION PATTERN²⁵



This pattern begins with the hands-on sexual abuse of the child and the recording of that abuse, in either photos or videos. Once the abuse has occurred and been recorded, the offender uploads the content and distributes it.²⁶ Once uploaded and initially distributed, there is the redistribution of CSAM.²⁷ Redistribution can be further subdivided into sharing among offender populations and sharing for nonmalicious purposes.²⁸

²⁵ For this diagram and those subsequent in this section, each text box represents an action taken, and the actions in the gray text boxes are those that could occur in end-to-end encrypted environments—for example, through the use of an encrypted chat or video application. The arrows show directionality, indicating how actions feed into one another, and the circular arrows illustrate actions that may continue to repeat. Within any given pattern of abuse, there may be multiple possible starting points, and the actions may interweave or repeat, resulting in revictimization and exploitation.

²⁶ It is possible that the offender produces CSAM and maintains it in a cache entirely disconnected from the internet, which is illegal, *see* 18 U.S.C.A. § 2252A(a)(5)(B), but outside the scope of this project as it is unimpacted by end-to-end encryption.

However, even if the offender does not send the content to others, if the offender elects to store the content on a third-party server, it will have entered the internet ecosystem insofar as it has relied upon means or facilities of interstate commerce, establishing federal jurisdiction. 18 U.S.C.A. § 2252A(a)(5)(B): Any person who “knowingly possesses, or knowingly accesses with intent to view, any book, magazine, periodical, film, videotape, computer disk, or any other material that contains an image of child pornography that has been mailed, or shipped or transported using any means or facility of interstate or foreign commerce or in or affecting interstate or foreign commerce by any means, *including by computer*, or that was produced using materials that have been mailed, or shipped or transported in or affecting interstate or foreign commerce by any means, *including by computer*; [shall be punished according to this statute]” (emphasis added).

²⁷ The U.S. Sentencing Commission published two reports in 2021, *Federal Sentencing of Child Pornography: Production Offenses* (October 2021) and *Federal Sentencing of Child Pornography: Non-Production Offenses* (June 2021), that are available at www.ussc.gov/topic/child-pornography. These reports identify and discuss factors the commission recommends that judges consider when sentencing people convicted of these crimes. To learn more about these factors, see Appendix 3.

²⁸ Facebook conducted a sample analysis of accounts reported to NCMEC for sharing CSAM over the course of three months in 2020 and estimated that approximately 75% of those accounts did not share the material with “malicious intent” (i.e., intent to harm a child); instead, the reported material had been shared in “outrage” (e.g., “can you believe the horrible things that happen to children?!”) or “poor humor” (e.g., a photograph of a child’s genitals being bitten by an animal). Facebook also acknowledges this is not a precise measurement and states that it is continuing its work to understand intent. Antigone Davis, “Preventing Child Exploitation on Our Apps,” Meta, February 23, 2021, <https://about.fb.com/news/2021/02/preventing-child-exploitation-on-our-apps/>.

This finding has not been independently verified, but it is still relevant to note because a sender’s motivation will impact what intervention is most successful in curtailing the behavior, and strategies that target people who share in poor humor or outrage will not be effective in stopping offenders who share for their own gratification.

“Known” CSAM presents a different issue from a mitigation perspective than “new” CSAM. Detecting known content, primarily through programs such as PhotoDNA, is a well-established and effective process. “New” material—i.e., imagery that has not previously been reported and therefore is not included in databases—cannot be detected using this method, and an equally effective method to detect “new” content has not yet been developed. This therefore creates an important distinction when evaluating interventions.

Whether the content is new or known, the continued harm to the victim is immense. No matter the intent of the person who shared the content, many child victims report that the knowledge that these images continue to circulate online causes them to feel revictimized.²⁹ And the scale of the problem is vast. According to NCMEC, in 2021 about 58% of image files and 89% of video files reported to the CyberTipline by ESPs were nonunique (i.e., they had been reported multiple times).³⁰ Facebook found that in a sample two-month period in 2020, copies of just six videos accounted for over one-half of all content reported.³¹

Once an offender has uploaded and shared the content, end-to-end encryption effectively creates a black box around the affiliated activity, preventing ESPs from accessing the content and preventing law enforcement from lawfully retrieving it from the provider with a search warrant. Communications apps that use end-to-end encryption currently rely on, and continue to refine, a set of signals and indicators intended to identify malicious users engaged in the sharing and redistribution of CSAM (see part 3 for a more in-depth discussion of this technique). But these indicators alone are unlikely to meet the necessary threshold for law enforcement to obtain a search warrant for the account holder’s home or devices. Therefore, for this pattern, more than any other, successful interventions must occur earlier in the pattern of harm due to the impact of end-to-end encryption eliminating visibility at the tail end.

29 In a survey conducted by the Canadian Centre for Child Protection, survivors of online CSEA were asked how the creation and distribution of the images of their abuse impacted them differently from the hands-on abuse itself. Of those who responded, 67% “pointed to the permanence of the images and the fact that if the images are distributed, their circulation will never end,” and 70% of respondents said they “worry constantly about being recognized by someone who has seen images of their abuse.” Canadian Centre for Child Protection, *Survivors’ Survey: Executive Summary 2017*, 28, https://protectchildren.ca/pdfs/C3P_SurvivorsSurveyExecutiveSummary2017_en.pdf.

30 “CyberTipline 2021 Report,” NCMEC, accessed September 24, 2022, <https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata>.

31 Antigone Davis, “Preventing Child Exploitation on Our Apps.”

II. *Perceived First Person (PFP) Material Production and Distribution*³²

This chart represents the relatively new trend of perceived first person (PFP) material—that is, explicit imagery of a child that appears to have been taken by the child in the image. The child may be enticed, induced, or exploited into taking the images by someone they either know in person or met online, or they may take the images voluntarily and independently.

After taking the photo or video, or in the course of live streaming a sexual act, the child may:

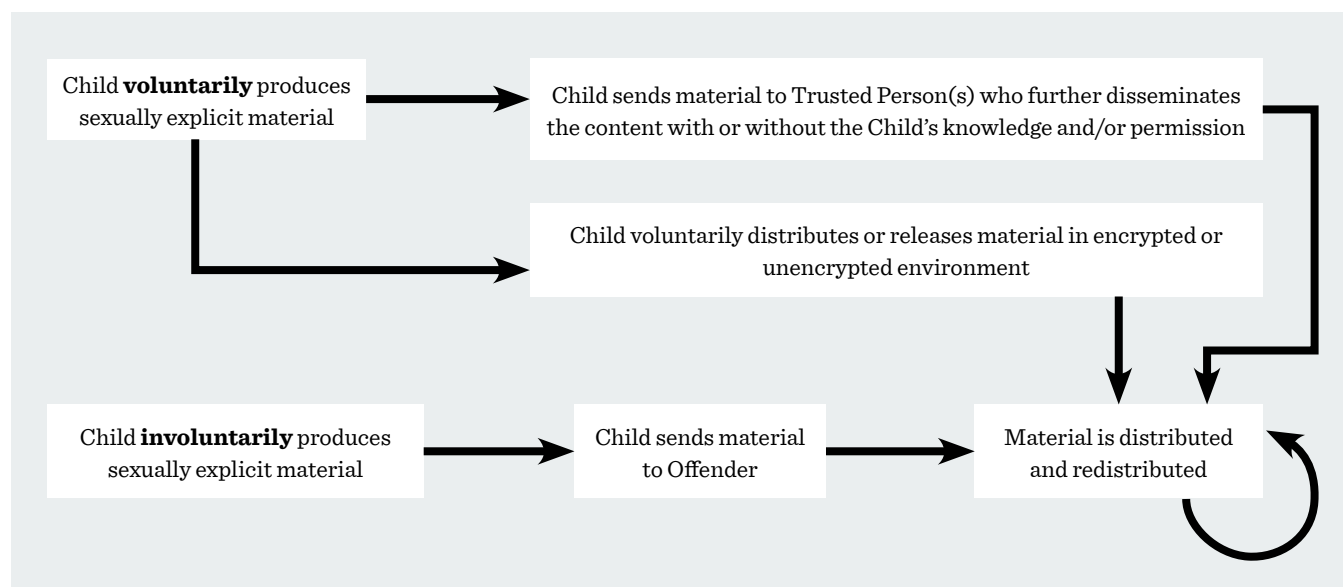
1. knowingly share the imagery with someone who does not distribute it further;

FIGURE 5: KNOWINGLY SHARING THE IMAGERY WITHOUT FURTHER DISTRIBUTION



2. knowingly share the imagery with someone who distributes it further with or without the child's knowledge, either to other like-minded individuals for malicious purposes, for a sense of revenge and/or control, or for other motivations;

FIGURE 6: KNOWINGLY SHARING THE IMAGERY WITH FURTHER DISTRIBUTION



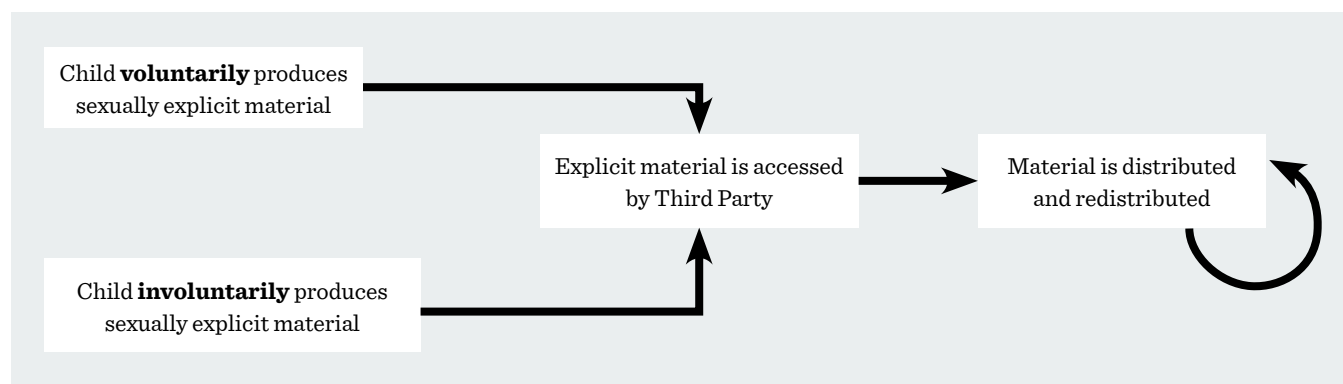
³² "Perceived first person CSAM" is the term advocated by the Tech Coalition, as opposed to the widely adopted "self-generated CSAM." It argues that the term "self-generated" implies agency by the child and therefore may infer blame while also failing to capture coercion and enticement, which are common in this pattern. See Tech Coalition, *Self-Generated Indecent Imagery Featuring Youth: Challenges & Opportunities*, April 2021, 3, <https://www.techcoalition.org/knowledge-hub/this-is-test-knowledge-2>.

Furthermore, while many of the images created are taken by the child, offenders may also remotely access the child's webcam and record images without the child's authorization or knowledge. For a more detailed discussion of this term, see Appendix 1.

For the figures in this section, italicized text refers to actions that do not contribute to online CSEA but that are necessary to understand the full range of related behaviors.

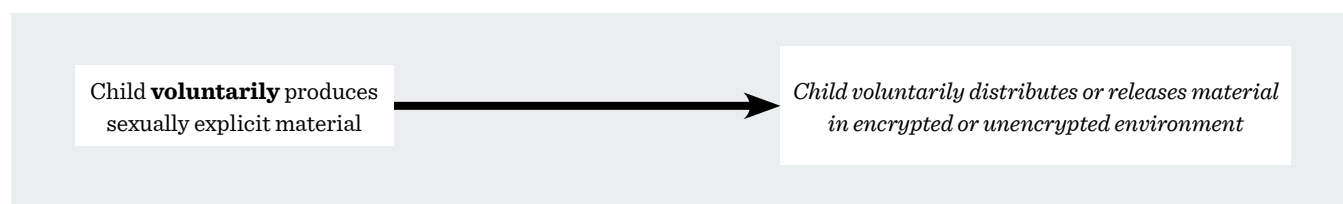
3. unwittingly share the images with an individual who has gained unauthorized access to the child's files; and/or

FIGURE 7: UNWITTINGLY SHARED IMAGES THROUGH UNAUTHORIZED ACCESS



4. distribute the image voluntarily via online platforms.

FIGURE 8: VOLUNTARY DISTRIBUTION

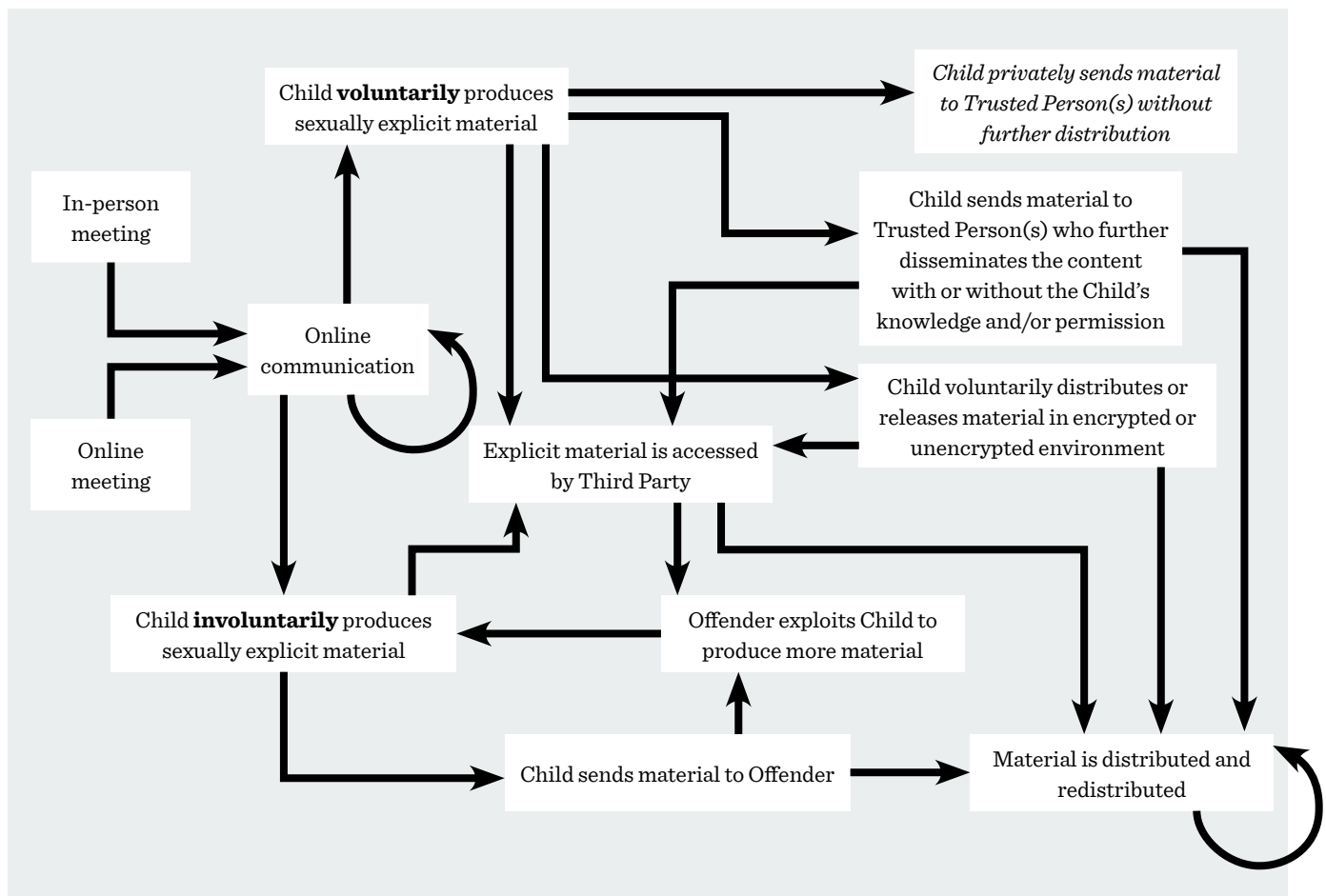


In some instances, the possession of one explicit image may result in the recipient exploiting or blackmailing the child for additional, often increasingly explicit, imagery (i.e., sextortion).

When the child is enticed, coerced, or blackmailed into producing this content, in-person sexual abuse may also be involved, either between the child and offender or between the child and another child they may be exploited into abusing. In either case, images and videos may be taken of the hands-on abuse, which could feed back into the CSAM production and distribution pattern described above.

Collectively, this pattern can be represented as follows.

FIGURE 9: PERCEIVED FIRST PERSON (PFP) MATERIAL PRODUCTION AND DISTRIBUTION



Although this is a relatively new phenomenon, it represents a large portion of the CSAM in the internet ecosystem. The Internet Watch Foundation reported that in 2020, 44% of confirmed CSAM contained perceived first person images, and in 2021, seven out of ten reports received by the Internet Watch Foundation included this type of content, representing dramatic increases from previous years.³³

³³ Six in ten of the Internet Watch Foundation's "actioned reports specifically show the sexual abuse of an 11-13 year old girl who has been groomed, coerced or encouraged into sexual activities via a webcam." See Internet Watch Foundation, *The Annual Report 2021, "2021 Trends & Data,"* accessed September 24, 2022, <https://annualreport2021.iwf.org.uk/trends>.

Irrespective of the morality or wisdom of the practice, sharing explicit photographs with romantic partners and even friends is increasingly part of adolescents' interpersonal and sexual development.³⁴ Research from 2020 found that about 20% of children ages 9 to 12 agreed that "it's normal for people my age to share nudes with each other," and 34% of 13-to-17-year-olds agreed with that statement.³⁵ The same study found that among children who have shared nudes, half reported that they had shared a nude photograph or video with someone they had not met in real life, and 41% reported that they had shared a nude photo or video with someone age 18 or older.³⁶ Beyond seeing a higher percentage of PFP content, researchers have identified the troubling trend of PFP content being "harvested" from their original upload location and posted to other publicly accessible sites.³⁷

Every action within this pattern of abuse can occur within an end-to-end encrypted environment, including the initial contact between the child and offender, as popular social media platforms adopt end-to-end encrypted messaging.³⁸

Unlike the typical production and distribution pattern described above, where communication is among offenders, the communications here typically involve the child victim. This is an important distinction when evaluating interventions to online CSEA. When adults share and receive CSAM for their own gratification, they are likely aware they are engaging in illegal behavior and would therefore not choose to report the content of their communications to a third party (either the company or law enforcement). However, a child victim may elect to self-report, and because the content is unencrypted on the child's device, so long as the child provides authorization, the content can be accessed by a third party even if it was sent and received in an end-to-end encrypted environment.

34 Law enforcement officials should *not* prioritize criminal investigation of these children. Instead, they should focus on individuals who exploit, coerce, and/or blackmail children and those who access and/or distribute the explicit images without authorization.

35 Thorn, *Self-Generated Child Sexual Abuse Material: Youth Attitudes and Experiences in 2020: Findings from 2020 quantitative research among 9–17 year olds*, November 2021, <https://www.thorn.org/resources-and-research/>.

36 This report relied on findings from a 20-minute online survey that collected self-reported data from minors 9 to 17. The sample included 742 9-to-12-year-olds and 1,260 13-to-17-year-olds. However, within that sample, respondents were split into two groups—one responding to questions about perceived hurdles to disclosure and one on the experience of and attitudes about perceived first person CSAM (relevant here). For this latter group, 351 9-to-12-year-olds and 651 13-to-17-year-olds participated. Data was weighted to age, gender, race, and geography to yield a representative nationwide sample. Thorn, *Self-Generated Child Sexual Abuse Material: Youth Attitudes and Experiences in 2020*.

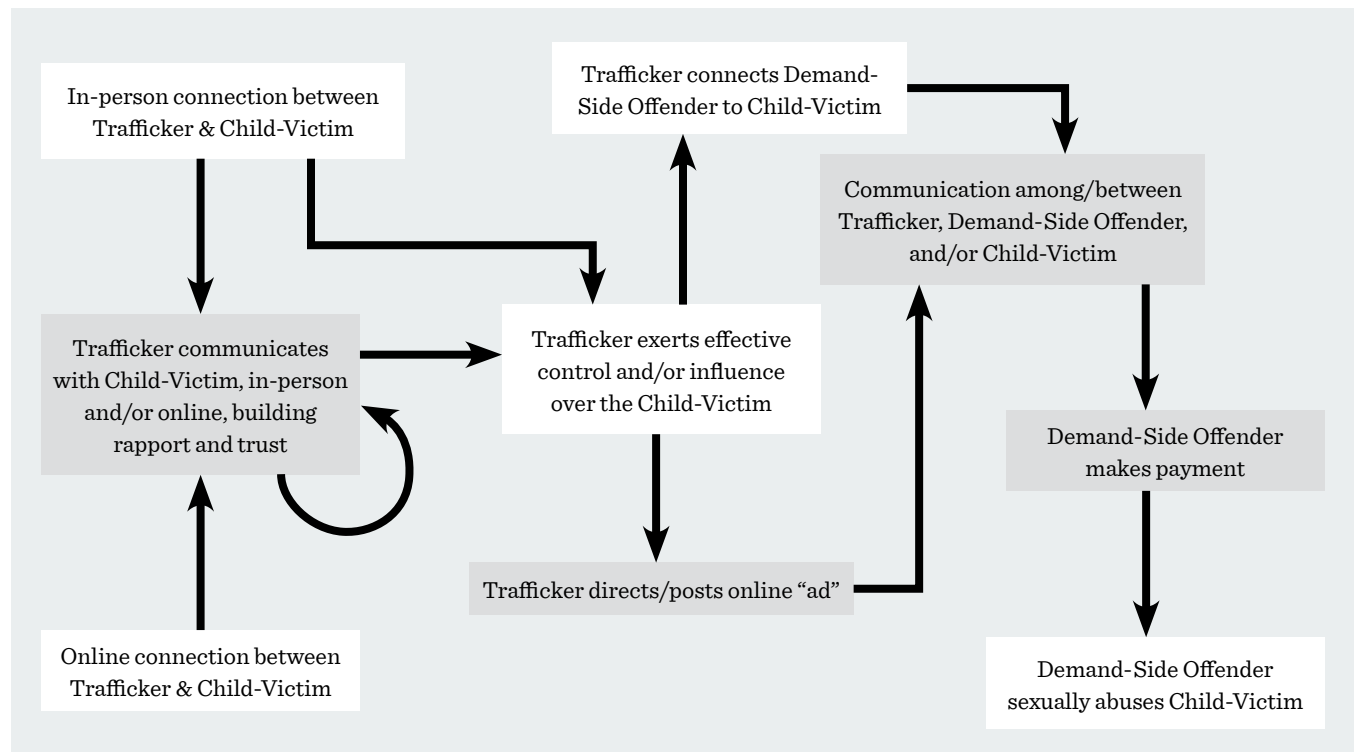
37 In 2015, researchers from the Internet Watch Foundation, with support from Microsoft, analyzed 3,803 photos and videos that were taken and supposedly shared by young people. Of these images, 89.9% had been "harvested" and posted to other public sites, and 100% of the images analyzed depicting children 15 and younger had been harvested and posted elsewhere. Internet Watch Foundation, *Emerging Patterns and Trends Report #1: Online-Produced Sexual Content*, March 10, 2015, 11, https://www.iwf.org.uk/media/2saninlk/online-produced_sexual_content_report_100315.pdf.

38 One law enforcement official reported that approximately 50% of content traded in end-to-end encrypted environments is "first generation"—i.e., new, perceived first person content. Interview by author, September 15, 2021.

III. Internet-Enabled Domestic Child Sex Trafficking³⁹

Unlike the patterns described above, much of this pattern of abuse occurs outside the confines of the internet. However, research suggests that technology plays an increasingly important role in grooming⁴⁰ and controlling child victims of sex trafficking, even while most victims have in-person contact with their traffickers.⁴¹

FIGURE 10: INTERNET-ENABLED DOMESTIC CHILD SEX TRAFFICKING PATTERN



After establishing trust with the child, the trafficker is able to control the child victim and arranges for demand-side offenders to sexually exploit the child.⁴² Unlike in the previously described patterns, there is a commercial element to child sex trafficking. The demand-side offender sexually exploits the child and, in exchange, provides money or other things of value (such as shelter or food). In-person abuse, by either the trafficker or demand-side offender, can occur throughout this cycle and be recorded, and PFP images can be created as part of online “advertisements.” Regardless of how

39 “Child sex trafficking” is defined in U.S. law as knowingly “recruit[ing], entic[ing], harbor[ing], transport[ing], provid[ing], obtain[ing], advertis[ing], maintain[ing], patroniz[ing], or solicit[ing] by any means” a person who “has not attained the age of 18 years and will be caused to engage in a commercial sex act.” 18 U.S.C. § 1591. It is typically treated as synonymous with “sexual exploitation of children in/for prostitution” or “child prostitution,” but as with “child pornography,” “prostitution” connotes consent, which is necessarily absent when a child is involved. For a longer discussion of this term, see Appendix 1.

40 “Grooming” is defined in the *Luxembourg Guidelines* as “the short name for the solicitation of children for sexual purposes. ‘Grooming/online grooming’ refers to the process of establishing/building a relationship with a child either in person or through the use of the Internet or other digital technologies to facilitate *either online or offline* sexual contact with that person” (italics in the original). Interagency Working Group, *Luxembourg Guidelines*, 51. For a longer discussion of the term, see Appendix 1.

41 Thorn and Vanessa Bouché, *Survivor Insights: The Role of Technology in Domestic Minor Sex Trafficking*, January 2018, <https://www.thorn.org/survivor-insights/>.

42 “Demand-side offender” is the term recommended by interviewees to describe the offender who engages in child sexual abuse and/or exploitation for the offender’s own sexual gratification. In contrast, the term “trafficker” describes an individual who aids, facilitates, and/or coordinates the hands-on abuse or CSAM production, often with commercial motivation. That is not to say that the trafficker may not also sexually abuse the child, but that their role is typically distinct. Interview by author, February 10, 2022; interview by author, April 14, 2022.

or when the explicit material is produced in this cycle, this content could feed back into the CSAM production and distribution pattern described above.

A survey of survivors of sex trafficking found that the median age of victims when first trafficked was 14 years old. For children younger than 10, the trafficker was almost exclusively a family member. Survey respondents who were older than 11 when first trafficked were “most likely to be trafficked by strangers, followed by people in their social network.”⁴³ Many survivors “experienced some form of childhood abuse and neglect, reporting high rates of verbal, physical, or sexual abuse.”⁴⁴ This volatility may have contributed to the children’s susceptibility to traffickers’ offers of food and shelter and promises of love and wealth.

End-to-end encryption obscures the content of communication among all parties—the child, the trafficker, and the demand-side offender. When the child is party to the communication, that may provide an opportunity to obtain the content. However, a strong majority (88%) of respondents to the aforementioned survivor survey reported they would not want their trafficker prosecuted,⁴⁵ which means they may not be willing to cooperate with law enforcement. In-person interventions may therefore provide some of the more promising solutions to combat this pattern of harm.

IV. Live Online Child Sexual Abuse⁴⁶ & Sexual Exploitation of Children in the Context of Travel and Tourism⁴⁷

The in-person components of this offense pattern most commonly occur outside the United States, but demand-side offenders—both those who pay for and view the live-streamed abuse and those who travel to sexually exploit children—come from all over the world, including the United States.⁴⁸ In countries such as the Philippines, nearly every referral about CSEA from foreign law enforcement agencies involves production of new imagery, often with a commercial component. These cases typically take the form of either live online child sexual abuse or sexual exploitation of children in the context of travel and tourism.

For live online child sexual abuse, demand-side offenders initiate contact with possible traffickers using the surface web (not the dark web),⁴⁹ looking for people who fit the common profile of a trafficker—for example, someone from a particular geographic region who posts suggestive photos of children and uses flirtatious language. If the prospective trafficker responds in kind, they engage in

43 Thorn and Vanessa Bouché, *Survivor Insights: The Role of Technology in Domestic Minor Sex Trafficking*.

44 Thorn and Vanessa Bouché, *Survivor Insights: The Role of Technology in Domestic Minor Sex Trafficking*.

45 Thorn and Vanessa Bouché, *Survivor Insights: The Role of Technology in Domestic Minor Sex Trafficking*.

46 Live online child sexual abuse “often represents a dual abuse of the child. She/he is coerced to participate in sexual activities, alone or with other persons—an act that already constitutes sexual abuse. The sexual activity is, at the same time, transmitted live through ICT and watched by others remotely.” Interagency Working Group, *Luxembourg Guidelines*, 46. For a more detailed explanation of this term, see Appendix 1.

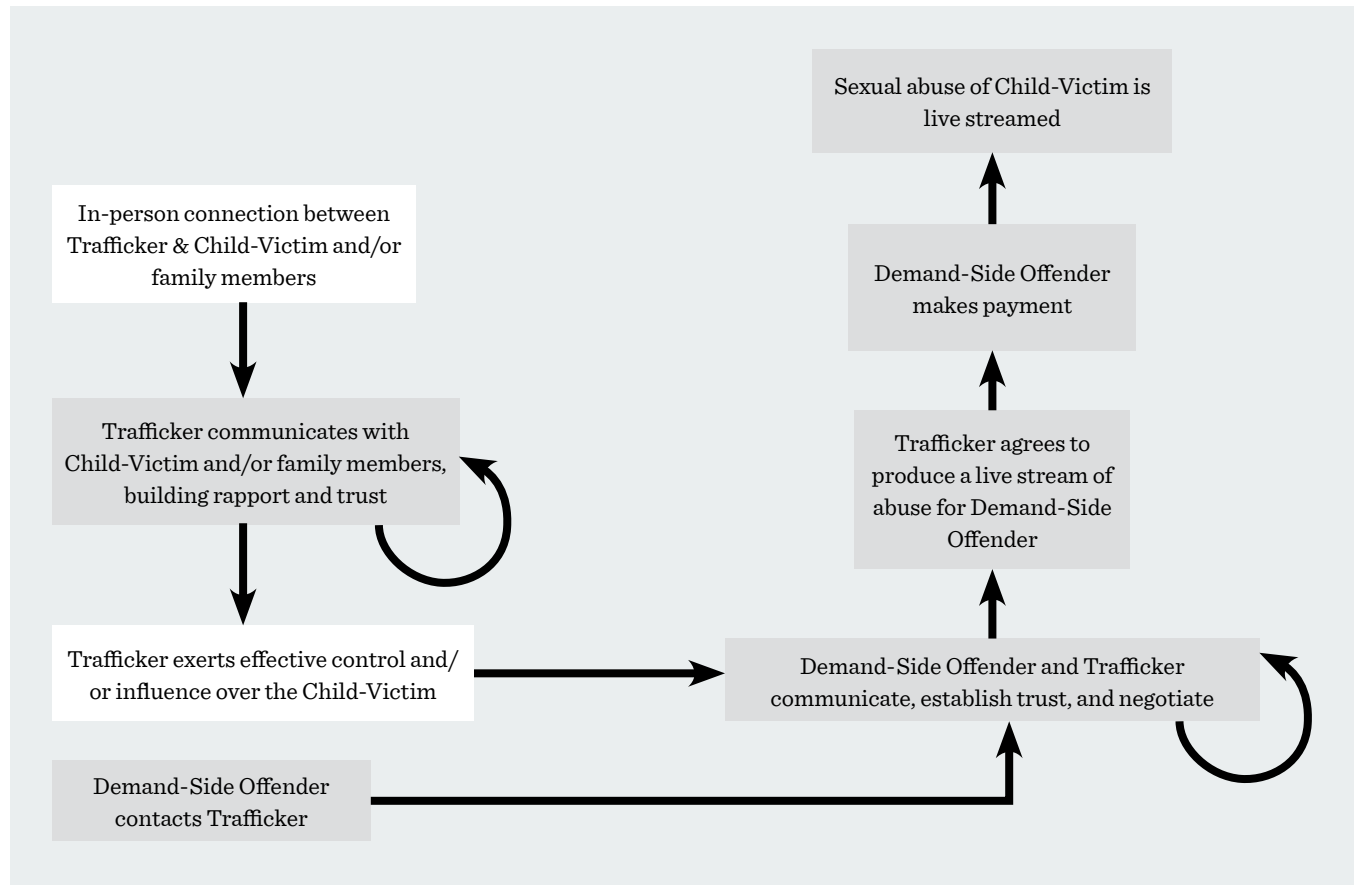
47 Sexual exploitation of children in the context of travel and tourism “refers to sexual exploitation of children that is embedded in a context of travel, tourism, or both. The offence can be committed by either foreign or domestic tourists and travellers and longer-term visitors. ... The term ... is used as an alternative to the broadly used term ‘child sex tourism.’” Interagency Working Group, *Luxembourg Guidelines*, 55. For a more complete definition, see Appendix 1.

48 International Justice Mission, *Online Sexual Exploitation of Children in the Philippines: Analysis and Recommendations for Governments, Industry, and Civil Society (Summary Report)*, 2020, <https://www.ijm.org/studies>.

49 The “clear web,” or “surface web,” consists of all publicly accessible websites that are indexed by search engines (e.g., <https://www.wcl.american.edu>). The “deep web” consists of the parts of the internet *not* indexed by traditional search engines (e.g., the landing page seen when logged in to one’s bank account). The “dark web” refers to a part of the internet *not* indexed by traditional search engines and that is *only* accessible through specific technical tools. (See, e.g., Federal Bureau of Investigation, “FBI and Partners Target Online Drug Markets,” accessed September 24, 2022, <https://www.fbi.gov/video-repository/jcode-102621.mp4/view>.) The “open web” refers to the publicly accessible areas of the clear and dark webs—i.e., those that do not require a password to access.

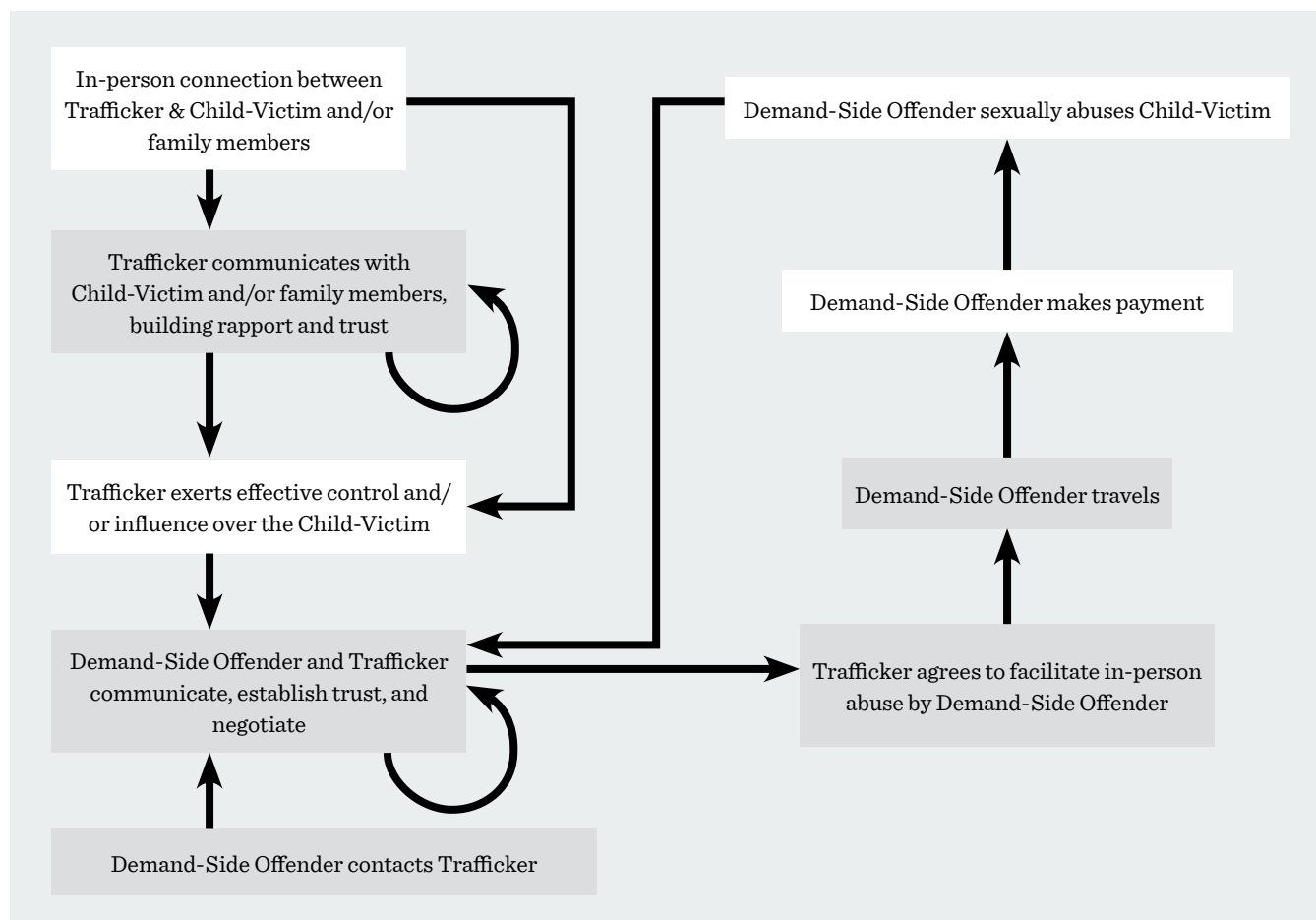
back-and-forth trust-building dialogue, often moving to encrypted messaging apps, resulting in the negotiation of an exchange of money for different levels of abuse/exploitation (e.g., still image, video, etc.).

FIGURE 11: LIVE ONLINE CHILD SEXUAL ABUSE



In the case of sexual exploitation of children in the context of travel and tourism, demand-side offenders and possible traffickers communicate using the same pattern as in the case of live online abuse—identify likely traffickers, initiate contact, and engage in trust-building. The difference is that the intended outcome of the demand-side offender in this scenario is to travel to exploit the child in person.

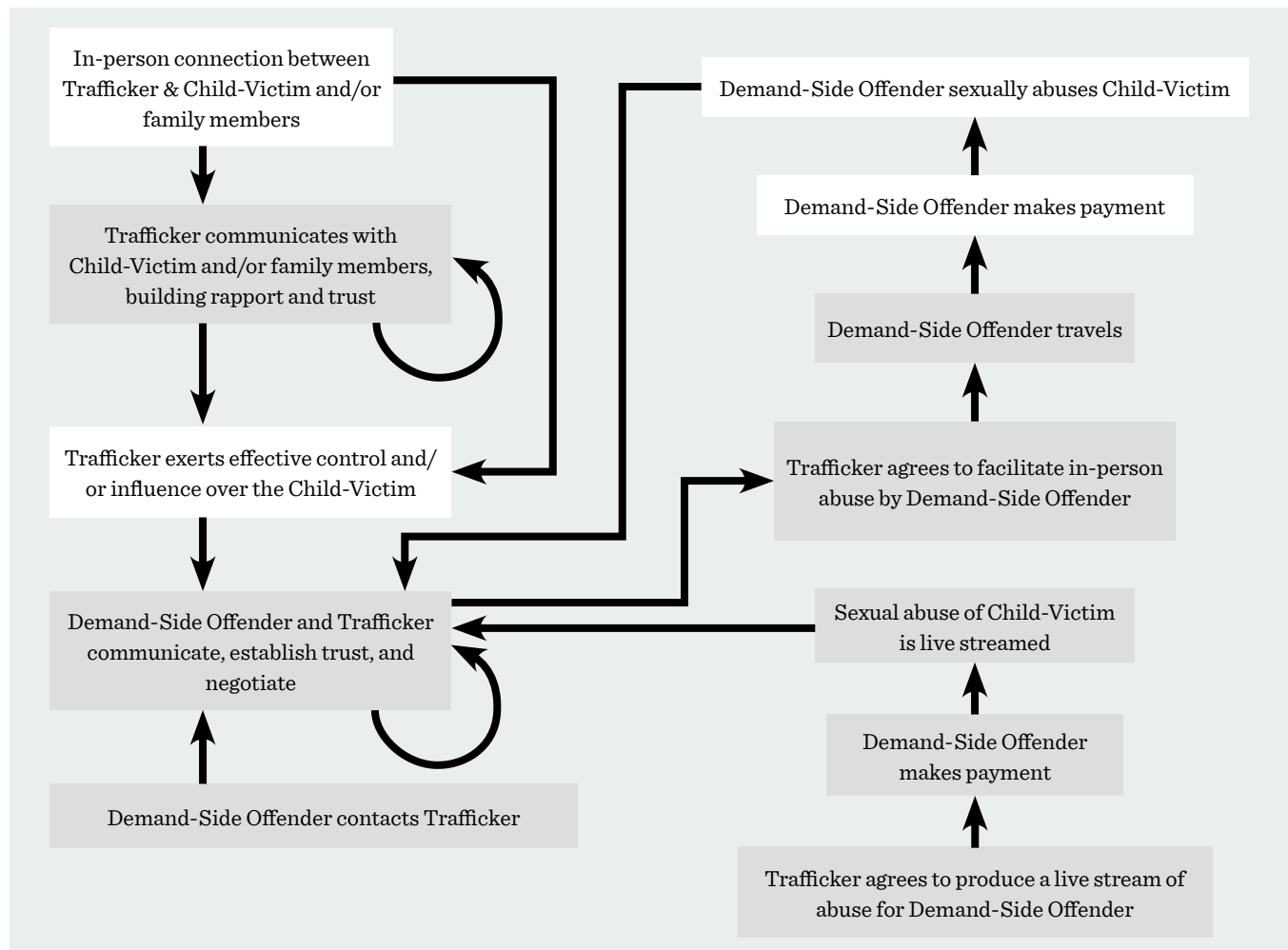
FIGURE 12: SEXUAL EXPLOITATION OF CHILDREN IN THE CONTEXT OF TRAVEL AND TOURISM



For both categories, there is typically a familial relationship between the trafficker and the child, and the trafficker builds and ultimately wields influence over the child.⁵⁰

⁵⁰ In research conducted by the International Justice Mission (IJM) about online sexual exploitation of children (OSEC) in the Philippines, 41% of victims were trafficked by their biological parents, and 42% were trafficked by other relatives. The IJM report defines OSEC as “the production, for the purpose of online publication or transmission, of visual depictions (e.g. photos, videos, live streaming) of the sexual abuse or exploitation of a minor for a third party who is not in the physical presence of the victim, in exchange for compensation.” The report is focused on OSEC originating from the Philippines in part because of IJM’s work with the government of the Philippines and in part because of the prevalence of OSEC in the Philippines. “Live-streaming OSEC is not unique to the Philippines, but it is believed to be more prevalent in the Philippines than in other countries due to numerous enabling factors such as widespread, inexpensive access to internet, a robust money transfer infrastructure, widespread English language proficiency, and the country’s historic commercial sex industry impacting its reputation as a sex trafficking source country.” International Justice Mission, *Online Sexual Exploitation of Children in the Philippines: Analysis and Recommendations for Governments, Industry, and Civil Society (Summary Report)*.

FIGURE 13: HOW LIVE ONLINE ABUSE OVERLAPS WITH EXPLOITATION IN THE CONTEXT OF TRAVEL AND TOURISM



This entire pattern is not a closed loop—every trafficker engages with multiple demand-side offenders, and demand-side offenders are typically in contact with multiple traffickers. Payment for live-streamed content is typically handled through a money transfer agency that can conduct international transfers, whereas payment for in-person abuse by traveling demand-side offenders is more often provided in cash. The scope of this problem is difficult to calculate, although one organization estimates that worldwide, about one million children are in “forced sexual exploitation” on any given day.⁵¹

As with the first pattern of harm, most steps in this pattern occur in end-to-end encrypted environments. Because the individuals in communication (i.e., the trafficker and demand-side offender) are engaged in illegal actions, they are adversarial to law enforcement and will not voluntarily self-report the content to a third party. Successful interventions in this pattern will therefore need to focus on any unencrypted surfaces and upstream nodes.

⁵¹ International Labour Office, Walk Free Foundation, and International Organization for Migration, *Global Estimates of Modern Slavery: Forced Labour and Forced Marriage*, 2017, referenced at <https://www.ecpatusa.org/statistics>.

PART 3: INTERVENTIONS

Online child sexual exploitation and abuse is a complex problem, and combating it requires a wide complement of efforts. There is no silver bullet. The ideas discussed in this section are presented as “interventions” rather than “solutions” because though they can aid in mitigating and reducing the harm, none of them solves the problem.

Because the unifying thread in online CSEA is the internet, the temptation is to look exclusively at technological solutions. However, the underlying issue—child sexual exploitation and abuse—would exist even if the internet did not. While there are some promising technological interventions, they must be considered within the larger context of the issue and should be explored concurrently with in-person, nontechnical solutions for a holistic approach to address the problem.

Table 1 (below) lists the interventions identified throughout the course of this project. (They are described in greater detail later in this section.) They are organized by the purpose of the intervention—prevent, detect, deter, refer, and disrupt—to provide an overarching structure, but these categories are neither perfect nor exhaustive. They also cannot be considered in isolation; for example, any intervention intended to “detect” online child abuse or exploitation must be paired with one or more interventions that seeks to deter, refer, and/or disrupt.

TABLE 1: OVERVIEW OF POSSIBLE INTERVENTIONS

<i>Purpose</i>	<i>Intervention Method</i>
Prevent	Community spaces for marginalized youth
	Comprehensive sex education
	Online safety education for children (embedded in the ICT)
	Perpetration prevention programs
	Safety by design
	Tech security education for children (in person)
Detect	Artificial intelligence image content analysis
	Perceptual hashing
	Metadata analysis
	Text-based analysis
	Undercover operations
	User reports
Deter	Issue warning notices to user
Refer	Redirect likely offenders
	Redirect likely victims
Disrupt	Investigate and arrest
	Deplatform abusive account
	Prevent creation and/or dissemination of CSAM

These interventions reflect the current environment and technical capabilities (and ambitions). Offender behavior morphs. Apps and platforms ebb and flow in popularity and use. Policies and

practices must be sensitive to these shifts, and decision-makers must be willing to make changes and updates as needed.

Each arrow in the patterns described in part 2 represents a potential intervention point. At each point, there could be multiple strategies, techniques, and/or solutions, technical or nontechnical, that could be implemented by a variety of stakeholders (i.e., social networking platforms, messaging apps, device manufacturers, safety tech firms, law enforcement, legislators). The immediate goal of the intervention may be prevention, detection, deterrence, referral, disruption, or some combination thereof, so long as the overarching purpose is to combat online CSEA.

Balancing costs and benefits

Every intervention has pros and cons, and decision-makers should carefully evaluate them in light of their intended and incidental impacts. Every intervention also imposes a cost, and there are limitations and barriers to implementation and/or scalability, including resources (money and/or time), technology (the technical capability is not ready for implementation or is not scalable), regulation/legislation (laws and/or regulations that prohibit or stymie a particular action), design (the capability exists, but its current design limits its broad use), social (there is a social stigma attached to a particular endeavor), and privacy (implementation may undercut some protections otherwise afforded through end-to-end encryption). Policymakers must consider whether the cost of an intervention is worth it, while keeping in mind that the cost of doing nothing is to allow online CSEA to continue unabated.

The European Commission outlined a set of five assessment criteria that present a useful framework for evaluating any intervention.⁵²

1. **Effectiveness:** how well does the solution detect and report known and unknown [CSAM]?
2. **Feasibility:** how ready is the solution and how easily can it be implemented, in terms of cost, time and scalability?
3. **Privacy:** how well does the solution ensure the privacy of the communications?
4. **Security:** how vulnerable is the solution to be misused for other purposes than the fight against [online CSEA], including by companies, governments or individuals?
5. **Transparency:** to what extent can the use of the solution be documented and be publicly reported to facilitate accountability through ongoing evaluation and oversight by policymakers and the public?

A recurring theme in discussions of this issue is that proposed interventions will intrude on users' privacy. A complete discussion of the privacy aspects of the online CSEA problem is beyond the scope of this report, but two relevant points are worth considering when evaluating interventions. First, although a general user's privacy should be part of the analysis, the countervailing privacy interest of the child depicted in the CSAM must also be

⁵² European Commission, *Technical solutions to detect child sexual abuse in end-to-end encrypted communications*, https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. This document is a draft report that was leaked in September 2020.

Many civil society organizations took issue with the leaked draft's approach to encryption. See, e.g., Global Encryption Coalition, *Breaking encryption myths: What the European Commission's leaked report got wrong about online security*, <https://www.globalencryption.org/wp-content/uploads/2020/11/2020-Breaking-Encryption-Myths.pdf>. However, despite its limitations, the draft provides a helpful framework to consider when evaluating any proposed intervention.

taken into account. It is a grave invasion of their privacy when offenders post these images on the internet and redistribute them. Second, privacy itself requires a more nuanced examination—specifically, privacy vis-à-vis whom? The strictest interpretation as applied to communications would be privacy from all parties except the sender and intended recipient. But on a more nuanced level, people may be focused on privacy from unreasonable access by the government, from malicious actors such as criminal hackers, or from private companies. When balancing users’ privacy with combating online CSEA, policymakers must be transparent in how they define privacy.

* * *

I. Prevent

Given how end-to-end encryption shields particular activities from detection, the most effective ways to intervene involve moving further upstream in the patterns of harm—preventing the abuse, preventing the exploitation, preventing the dissemination of existing content. The interventions listed here often involve community-based and in-person strategies, but there are also opportunities for tech companies to engage.

A. COMMUNITY SPACES FOR MARGINALIZED YOUTH

Young people naturally seek connection. To prevent that void from being filled by a predator, communities should seek to provide safe alternatives—whether through enhanced contact with community elders or with peers. This is especially necessary for marginalized youth populations, such as transgender and queer children, who are at a heightened risk of exploitation.⁵³ This intervention would likely be most effective at curbing internet-enabled domestic child sex trafficking.

While many trusted organizations provide safe community spaces for children, it is worth noting that children are more likely to be abused by a trusted person rather than a stranger.⁵⁴ Policymakers seeking to implement this intervention in their community may want to impose structures to mitigate this risk, such as minimizing one-on-one time between children and adults and providing transparent screening and supervision guidelines for staff and volunteers.⁵⁵

B. COMPREHENSIVE SEX EDUCATION

Comprehensive sex education provides youth with the information necessary to make healthy decisions and responsible choices about their sexual and social relationships, and research suggests that it can help prevent sexual assault.⁵⁶ Children armed with knowledge about healthy relationships may be more likely to self-report when an offender approaches them (as discussed below) because the child has learned to recognize what qualifies as harmful contact. This intervention would be most effective at preventing the production and dissemination of PFP CSAM but may also be relevant in

⁵³ Interview by author, February 1, 2022.

⁵⁴ “Resources: Understand and Identify Child Sexual Abuse,” Canadian Centre for Child Protection, accessed September 24, 2022, <https://www.protectchildren.ca/en/resources-research/understanding-child-sexual-abuse/>.

⁵⁵ For more information, see Child Welfare Information Gateway at <https://www.childwelfare.gov/>.

⁵⁶ SIECUS: Sex Ed for Social Change, *If you care about sexual assault prevention... Then you should care about SEX ED*, 2020, <https://siecus.org/wp-content/uploads/2020/08/If-Then-Sexual-Assault-Final.pdf>.

See also Savannah Sly and Tarah Wheeler, “An education-based approach to curbing CSAM production,” Brookings Institution, March 17, 2022, <https://www.brookings.edu/techstream/an-education-based-approach-to-curbing-csam-production/>. (“Predators lurking on chat sites would have a harder time grooming and exploiting youth if young people were equipped with knowledge about boundaries, consent, and healthy/unhealthy relationships.”)

other situations where an offender seeks to gain the trust and ultimate control of the child, such as child sex trafficking.

One challenge with this approach is that people may believe it shifts the burden, transferring the responsibility from the adult offender to the child victim. Its values, however, exceed the narrow purpose of preventing online CSEA, and therefore, policymakers should envision it as just one part of a larger suite of intervention efforts.

C. ONLINE SAFETY EDUCATION FOR CHILDREN (EMBEDDED IN THE ICT)

ICTs, including social media platforms and messaging apps, can integrate program-specific education for children, such as how to report abuse, set personal limits, understand the risk of taking/sending explicit photos, and find support if needed. For example, the Roblox education team takes the approach that when it has a child's attention, it wants to instill technical skills alongside civility, digital literacy, and online safety.⁵⁷ Another opportunity for platforms is the implementation of in-app reminders with safety tips. For example, Facebook prompts children to accept "friend requests" only from people they know.⁵⁸

As with comprehensive sex education, decision-makers must be careful not to blur the line between supporting children with necessary knowledge and burdening children with excess responsibility. Safety education may be necessary, but it is not sufficient to combat the problem.

D. PERPETRATION PREVENTION PROGRAMS

There are a few examples of promising practices in this area. First, experts believe that providing peer support for people who are at risk of offending, but who have committed never to harm a child, can reduce child sexual abuse. An example of this model is the MAP (Minor-Attracted People) Support Club, which seeks to serve this goal while incorporating high standards of accountability for its members.⁵⁹ Similarly, there are hotlines available where people worried about sexual thoughts or behaviors involving children can receive confidential support.⁶⁰

Perpetration-prevention therapy has also been shown to help intervene with people at risk of sexually abusing children. In particular, Michael Seto, the director of forensic rehabilitation research in the Integrated Forensic Program of the Royal Ottawa Health Care Group, and Elizabeth Letourneau, director of the Moore Center for the Prevention of Child Sexual Abuse at Johns Hopkins University, along with their colleagues, have conducted extensive research in this area with promising results.⁶¹

⁵⁷ Interview by author, November 9, 2021. Although Roblox does not provide end-to-end encrypted communications, its education approach is relevant here, given that its target audience is children.

⁵⁸ "Safety Resources for Parents: What steps does Facebook take to keep minors safe?" Help Center, Facebook, accessed September 24, 2022, <https://www.facebook.com/help/1079477105456277>.

⁵⁹ "MAP Support Club," Prostagia Foundation, accessed September 24, 2022, <https://prostagia.org/project/map-support-club/>.

⁶⁰ "Resources for People Concerned About Their Own Sexual Thoughts and Behavior," Moore Center for the Prevention of Child Sexual Abuse, Johns Hopkins Bloomberg School of Public Health, accessed September 24, 2022, <https://www.jhsph.edu/research/centers-and-institutes/moore-center-for-the-prevention-of-child-sexual-abuse/resources/csa-prevention/>.

⁶¹ "Preventing Child Sexual Abuse Before It Begins," Moore Center for the Prevention of Child Sexual Abuse, Johns Hopkins Bloomberg School of Public Health, accessed September 24, 2022, <https://www.jhsph.edu/research/centers-and-institutes/moore-center-for-the-prevention-of-child-sexual-abuse/>; "Forensic Mental Health," The Royal's Institute of Mental Health Research, accessed September 24, 2022, <https://www.theroyal.ca/research/forensic-mental-health>.

Focusing on perpetrators and potential perpetrators may seem like a misuse of limited resources to combat this issue when so much support is needed to help victims of these crimes. However, effective demand-side interventions may have the broadest impact because by lowering the demand, there should be a concurrent reduction in the number of children exploited and abused as well.

E. SAFETY BY DESIGN

Safety by design “focuses on the ways technology companies can minimise online threats by anticipating, detecting and eliminating online harms before they occur.”⁶² Australia’s eSafety Commissioner has worked closely with industry, civil society, and product users (including children) to develop standards under this rubric.⁶³ The basic premise is that tech companies should implement safety controls and user protections from the outset of product development. As one interviewee said: “Safety by design is a good way to think. Security and privacy by design are important, but safety is critical.”⁶⁴

One challenge to widespread implementation of this ideal is cost—new start-ups face an uphill challenge to compete against well-established firms and may not have the necessary infrastructure to support effective content moderation and other key elements. The eSafety Commissioner’s office is working to level the playing field by engaging directly with the investment community to encourage the incorporation of safety by design into investment criteria and by providing practical tools and resources to the start-up community.

Examples of Safety by Design in Practice

Yubo instilled user safety as a company priority at inception and has continued to emphasize it as the company has grown.⁶⁵ The company recently launched a new age-verification procedure with the express purpose of keeping users safe.⁶⁶

- **WhatsApp**, one of the most widely used end-to-end encrypted messaging services, prohibits users from searching for people they do not already know and requires a user to have another user’s phone number in order to contact them.⁶⁷
- For **websites and platforms that primarily host explicit adult content**, a disconnect often exists between the incentives for people managing the site and the

62 “Safety by Design,” eSafety Commissioner, Australian government, accessed September 24, 2022 <https://www.esafety.gov.au/industry/safety-by-design>.

63 John Perrino, “Using ‘safety by design’ to address online harms,” Brookings Institution, July 26, 2022, <https://www.brookings.edu/techstream/using-safety-by-design-to-address-online-harms/>.

Australia’s eSafety Commissioner has promoted three core principles to safety by design: “**1. Service provider responsibility. The burden of safety should never fall solely upon the user.** Every attempt must be made to ensure that online harms are understood, assessed and addressed in the design and provision of online platforms and services. ... **2. User empowerment and autonomy. The dignity of users is of central importance.** Products and services should align with the best interests of users. ... **3. Transparency and accountability. Transparency and accountability are hallmarks of a robust approach to safety.** They not only provide assurances that platforms and services are operating according to their published safety objectives, but also assist in educating and empowering users about steps they can take to address safety concerns” (bold in original). “Safety by Design: Principles and background,” eSafety Commissioner, Australian government, accessed September 24, 2022, <https://www.esafety.gov.au/industry/safety-by-design/principles-and-background>.

64 Interview by author, September 23, 2021.

65 For a panel interview featuring Yubo’s Marc-Antoine Durand discussing the company’s safety commitment, see “Day 3: Global Tech Innovations // Global Resolve Against Online Sexual Exploitation of Children 2020,” International Justice Mission, <https://www.youtube.com/watch?v=x9zcEK36mys>.

66 “Yubo’s new age verification feature helps keep you safe,” Yubo, accessed September 24, 2022, <https://www.yubo.live/blog/yubos-new-age-verification-feature-helps-keep-you-safe>.

67 “How WhatsApp Helps Fight Child Exploitation,” Help Center, WhatsApp, accessed September 24, 2022, https://faq.whatsapp.com/154956905959033/?locale=en_US.

incentives for people creating content for the site. Content creators are often aligned with people seeking to combat child exploitation; by empowering creators to have more control over platform design features, less illicit content would be hosted on a given site. For instance, mandating preverified uploads—i.e., content cannot be uploaded without express consent and age verification of the individuals featured—protects both groups: content creators ensure their original work cannot be copied and uploaded without their permission, and CSAM would also be prevented from upload.⁶⁸

Other design decisions referenced by interviewees include age verification and age-gated content; contact limits, wherein a communications app prevents “cold contact” with an individual not already saved to the device’s contacts; and alignment of the incentive structures of the company, content creators, and users.

F. TECH SECURITY EDUCATION FOR CHILDREN (IN PERSON)

In addition to opportunities provided for tech companies to intervene within a given application or platform, children would benefit from tech safety literacy more broadly, and any trusted adult, including educators, parents, and caregivers, can bear this responsibility.⁶⁹

As with any prevention method targeted to children, decision-makers should be mindful in how they message these education opportunities, as it risks development of a victim-blaming mentality when something goes wrong and shifting the burden of prevention from the potential offender to the potential victim. That said, given the prominence of technology in everyday life, it would be negligent not to ensure children have the skills necessary to safely navigate the online world.

II. Detect

Even if policymakers implement a full suite of prevention methods, there will remain a need to detect CSEA online. There are six primary ways to detect this content—artificial intelligence image content analysis, hash scanning, metadata analysis, text-based analysis, undercover operations, and user reports.⁷⁰ Each of these methods has benefits and risks, discussed in greater detail below.

It is also necessary to understand *where* in the internet ecosystem to apply these methods. Some online CSEA activities occur outside end-to-end encrypted environments, such as an offender making initial contact with a potential victim or harvesting PFP content and posting it to public websites. In these non-end-to-end encrypted environments, more detection options are available.

Detection of online CSEA within end-to-end encrypted spaces is more complicated. The aforementioned European Commission draft report on technical solutions identified three basic components to end-to-end encrypted communication: device, server, and encryption type. These

⁶⁸ Interview by author, January 12, 2022.

⁶⁹ For examples of available resources, see “Guides to Online Safety,” ECPAT-USA, accessed September 24, 2022, <https://www.ecpatusa.org/online-safety-tips/>; “Resources: Online Safety,” Canadian Centre for Child Protection, accessed September 24, 2022, <https://www.protectchildren.ca/en/resources-research/online-safety/>; “Guides and resources: Children and young people,” UK Safer Internet Centre, accessed September 24, 2022, <https://saferinternet.org.uk/guide-and-resource/young-people/>; “NetSmartz,” NCMEC, accessed September 24, 2022, <https://www.missingkids.org/NetSmartz/>.

⁷⁰ In *Outside Looking In: Approaches to Content Moderation in End-to-End Encrypted Systems*, published by the Center for Democracy and Technology (CDT), the authors identify five techniques used to detect content—user reporting, traceability, metadata analysis, perceptual hashing, and predictive models. Although bucketed differently, these methods generally cover the same range of detection techniques discussed here, with the exception of undercover operations. The CDT report provides a more detailed privacy analysis in the context of end-to-end encryption for readers interested in learning more about that aspect. Seny Kamara, Mallory Knodel, Emma Llansó, Greg Nojeim, Lucy Qin, Dhanaraj Thakur, and Caitlin Vogus, *Outside Looking In: Approaches to Content Moderation in End-to-End Encrypted Systems*, Center for Democracy and Technology, August 2021, <https://cdt.org/insights/outside-looking-in-approaches-to-content-moderation-in-end-to-end-encrypted-systems/>.

components in turn determine the three possible types of technical solutions: device-related, server-related, and encryption-related solutions.⁷¹

“Device related” encompasses all solutions that operate on a user’s device; “client-side scanning” is an example of a device-related solution.⁷² This includes “attachment notices” in email platforms, in which the program cues on language in the email that suggests the user may have intended to attach a document and prompts accordingly, as well as commonly used spam filters. It also refers to an artificial intelligence or automated hash-scanning feature enabled to detect CSAM before the user uploads or shares the content.

“Server related” solutions cannot be implemented in end-to-end encrypted environments because the content of the communication must be decrypted on the server in order for the detection method to run. This would allow a third party access to the content.

Not Strictly End-to-End Encrypted⁷³

Some experts have proposed a few alternative solutions for detection that fall outside the narrow conception of end-to-end encryption, yet provide more privacy and security than many existing efforts.

1. **Secure enclave (“middle box” or “end-to-end-to-end-to-end encryption”).** This method relies on an intermediate secure enclave where messages traverse between the sender and recipient for detection of CSAM or other contraband.
2. **Single server matching.** This is where a full hash value is calculated on the device and that value is sent to a server—managed by either the ESP or a trusted third party—for matching.
3. **Multiple third-parties matching.** Based on multiparty computation, the device calculates the hash value for the image, breaks it into parts, encrypts them, and sends the parts to multiple third parties for partial matching through the ESP server, which does not have access to the encrypted partial hashes.

“Encryption related” solutions rely on a homomorphic encryption scheme.⁷⁴ The challenge is that although this is possible, it currently takes about 10 to 15 seconds per image (contrasting with the one-thousandth of a second that PhotoDNA currently takes), making it functionally prohibitive given the volume of images transmitted across communication platforms.⁷⁵ Absent technological improvements to this scheme, this is not a viable option.

⁷¹ European Commission, *Technical solutions to detect child sexual abuse in end-to-end encrypted communications*.

⁷² Although compatible with end-to-end encryption, client-side scanning has strong detractors who argue that it undermines security and results in bulk intercept. See Hal Abelson, Ross Anderson, Steven M. Bellovin, Josh Benaloh, Matt Blaze, Jon Callas, Whitfield Diffie, Susan Landau, Peter G. Neumann, Ronald L. Rivest, Jeffrey I. Schiller, Bruce Schneier, Vanessa Teague, and Carmela Troncoso, *Bugs in our Pockets: The Risks of Client-Side Scanning*, October 14, 2021, <https://arxiv.org/abs/2110.07450>.

⁷³ European Commission, *Technical solutions to detect child sexual abuse in end-to-end encrypted communications*; Hany Farid, “An Overview of Perceptual Hashing,” *Journal of Online Trust and Safety* 1, no. 1 (2021), <https://doi.org/10.54501/jots.v1i1.24>.

⁷⁴ Farid, “An Overview of Perceptual Hashing.”

⁷⁵ European Commission, *Technical solutions to detect child sexual abuse in end-to-end encrypted communications*, referencing Priyanka Singh and Hany Farid, *Robust Homomorphic Image Hashing*, 2019.

A. ARTIFICIAL INTELLIGENCE IMAGE CONTENT ANALYSIS⁷⁶

Artificial intelligence detection models rely on image classifiers to determine that the image is likely CSAM.⁷⁷ The primary advantage to this technique is that it can identify new material. Because images can circulate throughout the internet ecosystem for some time before they are reported, hashed, and added to a CSAM database, methods to detect new material are critical to cutting short that time frame and limiting victims' exposure to extensive dissemination.

Apple's iMessage child safety feature, which is a device-based intervention, is a prominent example of this intervention in practice. The device scans incoming and outgoing images using a classifier intended to identify inappropriate content. Positive matches result in a warning to the child user that they may be about to send or receive explicit content, and it prompts the user to affirmatively accept or send the image in question.⁷⁸ Google also uses machine learning classifiers to power its Content Safety API, which helps companies identify potential new CSAM.⁷⁹ However, as currently used, it requires the identified content to be reviewed by human analysts, making this practice incompatible with end-to-end encryption.

Artificial intelligence is often presented as a solution to any tech problem, but it is an evolving technology, and in combating online CSEA in particular, three key challenges must be addressed.

1. Currently, AI **lacks the necessary accuracy and reliability** to be deployed in environments that process a large volume of data.⁸⁰ To create an accurate model, the AI must be able to identify all the relevant elements of CSAM, which is challenging with the way image classifiers are currently constructed.
2. The accuracy and reliability of AI to detect CSAM is further complicated by the **lack of training data**. In the United States, there are laws that prohibit possession of CSAM, which includes possession by for-profit companies and nonprofit organizations that would like access only to train these models.⁸¹
3. Deploying an AI to detect CSAM in an encrypted environment is **unpredictable** because the AI will certainly encounter data and images it has not previously encountered, and without the ability to independently access the content, the validity of the model's findings cannot be confirmed.

⁷⁶ For purposes of this report, artificial intelligence (AI) refers to a computer system or program that simulates human reasoning. "Machine learning" uses algorithms and statistical models to analyze and draw inferences to generate an AI that is able to learn and adapt without explicit instructions.

⁷⁷ For purposes of this report, a classifier refers to an algorithm that orders data into a set or sets of classes.

⁷⁸ "Communication safety in Messages," Apple, accessed September 24, 2022, <https://www.apple.com/child-safety/>.

⁷⁹ "Developing and sharing tools to fight child sexual abuse," Fighting child sexual abuse online, Google, accessed September 24, 2022, <https://protectingchildren.google/#tools-to-fight-csam>.

⁸⁰ In discussing the artificial intelligence and machine learning techniques employed by a major social media company to identify abusive accounts (i.e., accounts that engage in spam, scams, fake accounts), Levy and Robinson note that this "system exhibits an AUC of 0.90 which is excellent (although limited in utility as a metric in this highly biased population context). The recall at precision 0.95 is 0.50, which is reasonable in general and useful in the context the system is used. We do not believe that similar performance can be achieved on the child sexual abuse harm types, but for the sake of argument, assume it can. **These best-case performance measures seem inappropriate for the harm caused by child sexual abuse cases.** The reality in this case would be that the system would be 95% sure accounts flagged as abusive really were abusive, but it would only be detecting 50% of the abuse on the platform. We expect that, in an operational context, no scaled platform would run a classifier at a precision of 0.95 due to the excessive error and so we would expect an operational precision around 0.99 with an attendant loss of recall. As in most machine learning problems at scale, precision is critical for real world use" (bold in original). Levy and Robinson, *Thoughts on Child Safety on Commodity Platforms*, 50.

⁸¹ At the time of writing, there are legislative proposals that may soften these restrictions, but whether those provisions are ultimately signed into law, as well as the impact they would have if they were, remains unknown.

4. The first two challenges—low reliability and lack of training data—feed one another. The best way to improve accuracy and reliability of the model is by improving the data used to train it. The Internet Watch Foundation is taking steps to make better data available to companies interested in developing this detection method. Analysts are reviewing known CSAM and layering additional metadata related to the content. The expectation is that using this kind of metadata alongside the images will result in better-trained models.⁸²

B. PERCEPTUAL HASHING

The most well-established method for detecting known CSAM is perceptual hashing, which is the technology used in PhotoDNA. “Hash values” can be computed for any image.⁸³ Although technically more complicated, a “hash value” can be thought of as a digital fingerprint.⁸⁴ And like a fingerprint, its utility is limited to those situations where a comparator set is available. In the United States, NCMEC manages several databases of hash values for confirmed CSAM in order to provide this matching service, and some tech companies maintain their own hash sets. ESPs use these comparator databases to automatically compare against the hash values of images uploaded and/or shared to their services, thereby detecting “known” CSAM.

The Value of Detecting Known Images

If an image is “known”—that is, already identified as CSAM and included in the hash database—it may seem to be “low priority” in terms of investigative resources. The victim may even have been identified and removed from direct physical harm. However, detection and removal of known imagery is still a necessary element of a comprehensive strategy to combat online CSEA.

- Many law enforcement representatives have reported that detection of only one known image has often led to a cache of new imagery and/or active production involving a child victim.
- As outlined in part 2, the redistribution of known imagery continues to traumatize victims, making detection and removal of even known images a valuable service.

This method has proved extremely accurate and fast, which is critical to making any technical solution scalable.⁸⁵ In a survey of online service providers regarding their trust and safety practices, automated content scanning (which includes hash matching) was identified as the most useful means of detecting CSAM.⁸⁶

⁸² “IntelliGrade,” Internet Watch Foundation, accessed September 24, 2022, <https://www.iwf.org.uk/our-technology/intelligrade/>.

⁸³ Hash values can be calculated not only for still images, using methods such as PhotoDNA or Apple’s NeuralHash system, but also for videos, for example using Google’s CSAI Match or Meta’s TMK+PDQF. See “PhotoDNA,” Microsoft, accessed September 24, 2022, <https://www.microsoft.com/en-us/photodna>; CSAM Detection: Technical Summary, Apple, August 2021, https://www.apple.com/child-safety/pdf/CSAM_Detection_Technical_Summary.pdf; Google, “Developing and sharing tools to fight child sexual abuse”; Antigone Davis and Guy Rosen, “Open-Sourcing Photo- and Video-Matching Technology to Make the Internet Safer,” Meta, August 1, 2019, <https://about.fb.com/news/2019/08/open-source-photo-video-matching/>.

⁸⁴ For a more technical discussion of the various permutations of automated hash scanning, see Levy and Robinson, *Thoughts on Child Safety on Commodity Platforms*.

⁸⁵ As Levy and Robinson state with respect to PhotoDNA, one of the perceptual hashes used to detect known CSAM, they “are aware of large-scale, private testing of the PhotoDNA algorithm which suggests that, in a non-adversarial model, PhotoDNA has a false positive rate of around 1 in 50 billion. PhotoDNA was designed to be robust in the face of standard image manipulations and transformations. Under these conditions, its true positive rate is greater than 99%.” Levy and Robinson, *Thoughts on Child Safety on Commodity Platforms*, 50–51.

⁸⁶ Riana Pfefferkorn, “Content-Oblivious Trust and Safety Techniques: Results from a Survey of Online Service Providers,” *Journal of Online Trust and Safety* 1, no. 2 (February 2022), <https://doi.org/10.54501/jots.v1i2.14>.

Typically, ESPs run the hash value comparison on their own servers, which by its very nature cannot occur with end-to-end encryption, as the hash value of the image is shared with a third party. As a result, the industry-wide move toward end-to-end encryption significantly affects this detection method. Apple introduced a privacy-preserving alternative with its proposed iCloud child safety feature, which was later rescinded. With this feature, the hashing and scanning would have occurred on the device rather than on the company's server.⁸⁷ Although Apple's version could not work in an end-to-end encrypted environment, on-device scanning of this nature *could* if it were coupled with a response that did not provide the content to a third party. Using homomorphic encryption could allow this detection method to operate within an end-to-end encryption scheme, but the technology is not currently advanced enough to support that method at scale.

Although current iterations of PhotoDNA and other programs based on perceptual hashing cannot be deployed in end-to-end encrypted environments, the technology itself can still be used to scan public surfaces. For example, Project Arachnid, spearheaded by the Canadian Centre for Child Protection, uses hash-matching technology to identify and remove CSAM from the open web. Although this effort does not operate within end-to-end encrypted environments, by removing this content from the open web, it is less likely to migrate into encrypted spaces.⁸⁸

The Slippery Slope⁸⁹

One argument that often arises in response to hash-matching methods is the “slippery slope”—what is to stop authoritarian governments from including hash values in the comparator database to identify dissenting propaganda, or malicious actors from hacking the database and including hash values of innocent images to undermine the system's legitimacy?

In response to this argument, it is helpful to evaluate the track record of CSAM detection in the United States. For more than a decade, ESPs have broadly used PhotoDNA and other hash-matching technologies without any reports of “sliding” out of the narrow scope of CSAM. This is relevant for two reasons.

First, CSAM—in particular, known and previously hashed imagery—is unique. Possession itself is illegal in the United States.⁹⁰ No other category of images is treated the same. The other most commonly referenced “universal bad” is terrorism, but possession of terrorist propaganda itself is not per se illegal. Given that the majority of large ESPs are based in

87 In the proposed plan, all images uploaded to iCloud would receive a “security voucher.” If a threshold number of positive matches to the known CSAM database were detected in an account, the vouchers would be unwrapped, and a human moderator at Apple could review information about the identified files, including a thumbnail image. Because some content would have been provided to a third party, this would have been inconsistent with an end-to-end encryption scheme. See Apple, *Expanded Protections for Children*, August 2021, https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Technology_Summary.pdf.

88 For more information, see “Project Arachnid,” Canadian Centre for Child Protection, accessed September 24, 2022, <https://www.projectarachnid.ca/en/>. To understand the approach and impact in greater detail, see Canadian Centre for Child Protection, *Project Arachnid: Online Availability of Child Sexual Abuse Material: An analysis of CSAM and harmful-abusive content linked to certain electronic service providers*, June 8, 2021, https://protectchildren.ca/pdfs/C3P_ProjectArachnidReport_en.pdf.

89 For additional discussion on this topic, see “CSINT Conversations: Stopping online abuse of children – Could Apple have the answer?” recorded January 21, 2022, Center for Security, Innovation, and New Technology, American University, <https://www.youtube.com/watch?v=aKISxDxb1kg>.

90 18 U.S.C.A. § 2252A(a)(5)(B) (any person who “knowingly possesses, or knowingly accesses with intent to view, any book, magazine, periodical, film, videotape, computer disk, or any other material that contains an image of child pornography that has been mailed, or shipped or transported using any means or facility of interstate or foreign commerce or in or affecting interstate or foreign commerce by any means, including by computer, or that was produced using materials that have been mailed, or shipped or transported in or affecting interstate or foreign commerce by any means, including by computer; [shall be punished according to this statute].”)

the United States, the First Amendment provides broad protection to all other categories of content.

Second, NCMEC triple-vets all images added to the database.⁹¹ This ensures no malicious “insider threat” to the database. When paired with databases maintained by other sovereign nations that impose similarly rigorous processes, this can help ensure that no “false positives” are intentionally added.

The main limitation to perceptual hashing is that it detects only known imagery. As the volume of newly created PFP content continues to grow, the impact of hash matching may decline, as there may come a point when it does not detect the majority of CSAM on the internet. The Internet Watch Foundation in the United Kingdom, in collaboration with the National Society for the Prevention of Cruelty to Children, has implemented “Report Remove,” a tool that allows youth to self-report self-generated content for inclusion in the CSAM databases.⁹² They are working with large social media companies to promote this option for children concerned that their images may be shared without authorization.

C. METADATA ANALYSIS

Metadata refers to information collected by ESPs about users outside of the users’ content (i.e., the images and text that users send to one another). This may include subscriber information (e.g., user-provided age, user-provided email address), information collected from the device itself (e.g., IP address, location data), and information about the user’s account (e.g., account username, account avatar). ESPs can use this information to identify abusive account behavior and may elect to suspend or deactivate the account based on this finding. For example, WhatsApp, an end-to-end encrypted messaging application, uses machine learning classifiers, relying on unencrypted data and other “signals,” to evaluate group behavior for suspected CSAM sharing.⁹³ Facebook uses a classifier to identify potentially harmful accounts, using signals from public profiles as well as predictive nonpublic signals, such as group membership and group behavior.⁹⁴

Metadata analysis is not the most effective technique to detect online CSEA. In a survey of service providers regarding trust and safety practices, none of the respondents stated that metadata was the “most useful” technique to identify child sexual abuse imagery, although some respondents felt metadata was useful in detecting child sexual exploitation, which researchers defined as “other child safety offenses such as grooming and enticement.”⁹⁵ Unlike hash matching, which has a proven track record, metadata analysis is conducted without the actual content of messages, making it difficult to calculate false positive and false negative rates outside testing environments.

Furthermore, law enforcement officials report that, absent huge improvements in accuracy and reliability, reports based on metadata alone will not be sufficient to establish probable cause to

91 For inclusion in the NCMEC hash database, an image must be triple-vetted by trained NCMEC staff members who confirm the image meets a series of criteria. See “NCMEC, Google and Image Hashing Technology,” Safety Center, Google, accessed September 24, 2022, <https://safety.google/stories/hash-matching-to-help-ncmec/>. The Canadian Centre for Child Protection has a similarly rigorous approach. Interview by author, May 12, 2022.

92 “Report Remove,” Internet Watch Foundation, accessed September 24, 2022, <https://www.iwf.org.uk/our-technology/report-remove/>.

93 WhatsApp, “How WhatsApp Helps Fight Child Exploitation”; interview by author, November 10, 2021.

94 Interview by author, July 13, 2021; Antigone Davis, “New Technology to Fight Child Exploitation,” Facebook, Meta, October 24, 2018, <https://about.fb.com/news/2018/10/fighting-child-exploitation/>.

95 Pfefferkorn, “Content-Oblivious Trust and Safety Techniques,” 16. In contrast, as discussed above, the majority of the survey respondents felt that automated content scanning, such as hash matching, was most useful in detecting child sexual abuse imagery.

support even a search warrant.⁹⁶ Metadata-based reports can be compared to receiving a tip from an unknown source—independent corroboration and substantiation would be required to pursue an investigation.⁹⁷ This limits the options (discussed below) of what companies can do once metadata analysis indicates a harmful account.

D. TEXT-BASED ANALYSIS

Rather than scanning imagery to detect CSAM, text-based analysis moves the detection effort upstream, seeking to identify patterns in text that suggest grooming or enticement. There have been some successful efforts throughout the tech industry to pinpoint indicators necessary to identify these categories of language. Microsoft spearheaded Project Artemis, initiated with a cross-industry hackathon in 2018, which led to the development of an anti-grooming tool built specifically to detect enticement and solicitation of a child for sexual purposes.⁹⁸ The accuracy of the tool is currently about 88%, meaning there is a 12% false positive rate, which may limit its utility in high-volume communications platforms.⁹⁹ Facebook uses text-based classifiers in messages where at least one party is a child in order to identify when the child may be getting uncomfortable; Facebook then directs messaging to the child to provide resources for support.¹⁰⁰

Whereas these text-based methods operate within a given platform or application, this technique can also be used on a device. Safety tech firm SafeToNet developed an AI-powered smart keyboard that can be downloaded to a child's device to work across ICTs. It identifies indicators of abuse and bullying, among other risky behaviors, and then prompts the child to help keep them safe.¹⁰¹ Content is not shared with parents or any other third parties, making it compatible with end-to-end encryption.

Text-Based Analysis Outside End-to-End Encryption

This detection method is also used outside end-to-end encrypted spaces in an effort to prevent abuse and exploitation before users move into encrypted environments.

- **Roblox** uses text-based filters to prevent users from sharing sensitive information with one another such as home addresses and other identifiers.¹⁰²
- **YouTube** regulates public comments that invite the content creator or other commenters to produce or share CSAM.¹⁰³

Any adult who has ever tried to keep up with language commonly used by children knows it can feel like an insurmountable task; words and phrases used in popular slang one day are likely to be dismissed as outdated the next. Therefore, it is imperative that any organization developing and

⁹⁶ See, e.g., Levy and Robinson: "We believe that **if implemented as suggested, AI and ML running on metadata alone will be unlikely to assist law enforcement in stopping offenders and safeguarding victims, but will likely lead to a chilling effect on free speech, and infringement on user privacy**" (bold in original). Levy and Robinson, *Thoughts on Child Safety on Commodity Platforms*, 54.

⁹⁷ Interview by author, August 18, 2021.

⁹⁸ Jacqueline Beauchere, "Microsoft hosts tech industry hackathon to combat child online grooming," *Microsoft on the Issues* (blog), Microsoft, November 9, 2018, <https://blogs.microsoft.com/on-the-issues/2018/11/09/microsoft-hosts-tech-industry-hackathon-to-combat-child-online-grooming/>; C. Fisher, "Microsoft releases a free tool to fight online child abuse," Engadget, January 9, 2020, <https://www.engadget.com/2020-01-09-microsoft-project-artemis-online-child-abuse.html>; interview by author, July 1, 2021.

⁹⁹ European Commission, *Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse*, May 11, 2022, 14, available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0209&from=EN>.

¹⁰⁰ Interview by author, July 13, 2021.

¹⁰¹ "Features," SafeToNet, accessed September 24, 2022, <https://safetonet.com/features/>. SafeToNet is continuing to develop the keyboard technology with a view to incorporating it into future versions of its solutions.

¹⁰² "Text Filtering," Guides, Roblox, accessed September 24, 2022, <https://developer.roblox.com/en-us/articles/Text-and-Chat-Filtering>.

¹⁰³ Interview by author, July 2, 2021.

training an AI to detect enticement conversations collaborates closely with survivors and victims' rights organizations to understand and stay updated on the nuances and coded uses of language.

The task of staying current is further complicated when acting in an end-to-end encrypted environment because classifiers will be constrained by existing training data and may lack sensitivity to evolutions in language patterns used by offenders. In less strict encryption schemes, these tools can be audited and updated to reflect both natural and intentional changes, but that access is unavailable in end-to-end encrypted environments.

Using text-based analysis as a detection method limits what companies can do next. Unlike possession of CSAM, most language is subject to First Amendment protection. Exchanges can cross the line into criminal enticement and extortion, but distinguishing between harmless trust-building dialogue and harmful grooming and enticement conversations can be difficult, but also necessary before reporting to law enforcement for investigation.

E. UNDERCOVER OPERATIONS

Law enforcement officials have relied on undercover operations to identify and ultimately prosecute many CSAM offenders.¹⁰⁴ Access to content (text or images) is necessary to establish probable cause for search warrants and other investigative steps. When *automated* detection methods (i.e., AI image analysis, hash matching, metadata analysis, and text-based analysis) operate within end-to-end encrypted environments, they cannot share detected content with third parties, which means law enforcement is not given the information needed to pursue an investigation. Therefore, in end-to-end encrypted environments, undercover operations are the only detection method that can identify offender-to-offender CSAM dissemination (i.e., pattern I) and live online child sexual abuse conducted by traffickers (i.e., pattern IV) in a format that can lead to criminal investigation and prosecution.¹⁰⁵

The primary limitation to undercover operations is how resource intensive they can be. Successful undercover agents must cultivate a great deal of expertise to understand the cadence of these encounters and the operations of the different platforms and apps. They must also devote a great deal of time and cannot work traditional hours, instead making themselves available any time the targeted offenders are meeting online. In the case of U.S. law enforcement, local and state officials report feeling overwhelmed by the sheer volume of reports received from NCMEC, such that dedicating the time necessary to engage in long-term proactive undercover operations is a challenge.¹⁰⁶

¹⁰⁴ See, e.g., Spencer S. Hsu, "U.S., South Korea dismantle secret online network that shared thousands of videos of child sexual abuse," *Washington Post*, October 16, 2019, https://www.washingtonpost.com/local/legal-issues/us-south-korea-dismantle-secret-online-network-that-shared-thousands-of-videos-of-child-sexual-abuse/2019/10/16/cdae13c2-eb63-11e9-9c6d-436a0df4f31d_story.html; Don Morgan, "San Antonio Man Sentenced to 28 Years in Prison for Producing Child Pornography," *KTSA*, June 24, 2022, <https://www.ksa.com/san-antonio-man-sentenced-to-28-years-in-prison-for-producing-child-pornography/>; Michael Krafcik, "Vicksburg couple arrested in FBI child pornography raid," *WWMT*, July 15, 2022, <https://www.wmt.com/news/local/fbi-vicksburg-sexual-explicit-children-federal-court-raid-chad-knowles-samantha-batts>.

¹⁰⁵ User reporting, discussed below, is the other detection method that can generate the evidence within end-to-end encrypted environments necessary for law enforcement to pursue an investigation. However, user reporting relies on one user identifying a harm and reporting it, which is why it is not an effective detection method in those situations where all parties knowingly engage in unlawful acts and would not want law enforcement involved.

¹⁰⁶ Interview by author, August 6, 2021; interview by author, August 11, 2021; interview by author, September 7, 2021; interview by author, September 15, 2021; interview by author, September 27, 2021.

F. USER REPORTS

In end-to-end encrypted environments, platforms or applications cannot access the content of users' communications without the consent of either the sender or the intended recipient(s). However, if a party to the conversation opts to report the content as abusive, that practice does not violate the encryption scheme, as the users themselves remain the locus of control.

A robust user-report ecosystem includes (1) sufficient tipline access globally, (2) easy-to-use in-app or on-platform reporting functions, and (3) safe and reliable in-person reporting options.

Tiplines are a key component to user-reporting systems. There are regions of the world that generate a large portion of CSAM but do not have the necessary reporting infrastructure. For example, the vast majority of content reported by Facebook to NCMEC originates from and is sent to regions of the world that lack the infrastructure to safely accept reporting, which instead results in nonmalicious sharing (discussed in pattern I).¹⁰⁷ Creating secure networks where users could report this content could yield a dramatic decrease in the amount of CSAM that is disseminated out of outrage.

Several concerted efforts are underway to address these gaps.

- Tech Matters has built an open-source platform that improves the efficiency and efficacy of helplines around the world.¹⁰⁸
- InHope is working to expand its network to countries without helplines.¹⁰⁹
- The Internet Watch Foundation has established reporting portals in some of the most needed countries, representing approximately 2.5 billion people.¹¹⁰

When building and enhancing tipline infrastructure, it is necessary to couple these efforts with community-specific education campaigns alerting people to these services. For example, Roblox funds local nongovernmental organizations to create region-specific parental guides for its platform.¹¹¹ A similar approach can be taken alongside the launch of a tipline. Similarly, Meta created a video for Facebook users explaining how sharing CSAM, rather than reporting it, compounds the abuse to the victim. The company used classifiers to identify which accounts are most likely to share this type of imagery and then directed the video message to them.¹¹²

All end-to-end encrypted services should have **easy-to-use reporting functions** that explicitly permit users to report online CSEA. The Canadian Centre for Child Protection has developed several recommendations to clarify and streamline the process for reporting CSAM. These include (1) creating reporting categories specific to CSAM and online CSEA, (2) including CSAM-specific options in easy-to-locate reporting menus, (3) ensuring reporting functions are consistent across the platform, (4) allowing users to report content that is visible without creating or logging in to an account, and (5) eliminating mandatory personal information fields in reporting forms.¹¹³

¹⁰⁷ Interview by author, July 12, 2021.

¹⁰⁸ "Aselo: Helping the World's Children," Tech Matters, accessed September 24, 2022, <https://techmatters.org/project/child-helplines/>.

¹⁰⁹ "Report Box," InHope, accessed September 24, 2022, <https://reportbox.inhope.org/EN/escape-project>.

¹¹⁰ "IWF Reporting Portals," Internet Watch Foundation, accessed September 24, 2022, <https://www.iwf.org.uk/about-us/our-international-work/reporting-portals/>.

¹¹¹ Interview by author, November 9, 2021.

¹¹² Interview by author, July 13, 2021.

¹¹³ Canadian Centre for Child Protection, *Reviewing Child Sexual Abuse Material Reporting Functions on Popular Platforms*, 2020, https://www.protectchildren.ca/pdfs/C3P_ReviewingCSAMMaterialReporting_en.pdf.

Reporting functions can help curb the nonmalicious sharing of CSAM and may provide off-ramps to children who are being enticed or extorted into producing PFP content. Given the ubiquity of technology, even hands-on abuse perpetrated by a relative or family friend is likely to involve technology as the offender seeks to gain the child's trust. This creates an opportunity for children to self-report during these initial stages, so long as the reporting mechanism is easy, straightforward, secure, and reliable.

Understanding Why Victims May Not Self-Report¹¹⁴

There are many reasons why child victims may not self-report abuse, especially when the pattern of harm is either PFP content or child sex trafficking. Understanding the barriers to a child's ability and willingness to self-report, either within an app or platform or to a trusted person, is absolutely essential for building effective tools that facilitate these reports.

Here are a few reasons discussed by interviewees; they should not be considered mutually exclusive or exhaustive.

1. The child may not consider the actions to constitute a "harm" that needs to be reported.
2. The child may realize the actions are harmful, but they may not want the offender to be punished.
3. The child may not feel safe enough to report, fearing retribution by the offender or others.
4. The child may not know of a secure reporting mechanism.
5. The child may feel as though they have no one to talk to about it.
6. The child may feel a sense of shame.
7. The child may be concerned about their own exposure to criminal liability.

Even if a child is unwilling or unable to report abuse online, they may **report the abuse in person**, either to a friend, a trusted adult, or law enforcement. One law enforcement official noted that while they had never received a tip via NCMEC in which the child self-reported the harm to the website or platform, they found it more likely that the child will tell a parent, sibling, or trusted adult, who will then file a report (either with a tipline such as NCMEC or directly to law enforcement).¹¹⁵ To foster a safe environment for children to make these reports, they need relationships with trusted adults who know how to respond. Training for first responders, such as police officers, and mandatory reporters, such as teachers and social workers, should be developed with input from survivors.¹¹⁶

* * *

¹¹⁴ Interview by author, February 1, 2022.

¹¹⁵ Interview by author, January 28, 2022.

¹¹⁶ Excellent references are available through NCMEC, ECPAT-USA, and the Canadian Centre for Child Protection. See "Training," NCMEC, accessed September 24, 2022, <https://www.missingkids.org/education/training>; ECPAT-USA, *A Educator Guide to Online Sexual Exploitation and Human Trafficking During COVID-19*, accessed October 12, 2022, <https://static1.squarespace.com/static/594970e91b631b3571be12e2/t/5eecec83d743f6598d9f0133/1592585348040/Educator+Guide+to+Online+Exploitation.pdf>; "Resources: Understand and Identify Child Sexual Abuse," Canadian Centre for Child Protection, accessed September 24, 2022, <https://www.protectchildren.ca/en/resources-research/understanding-child-sexual-abuse/>.

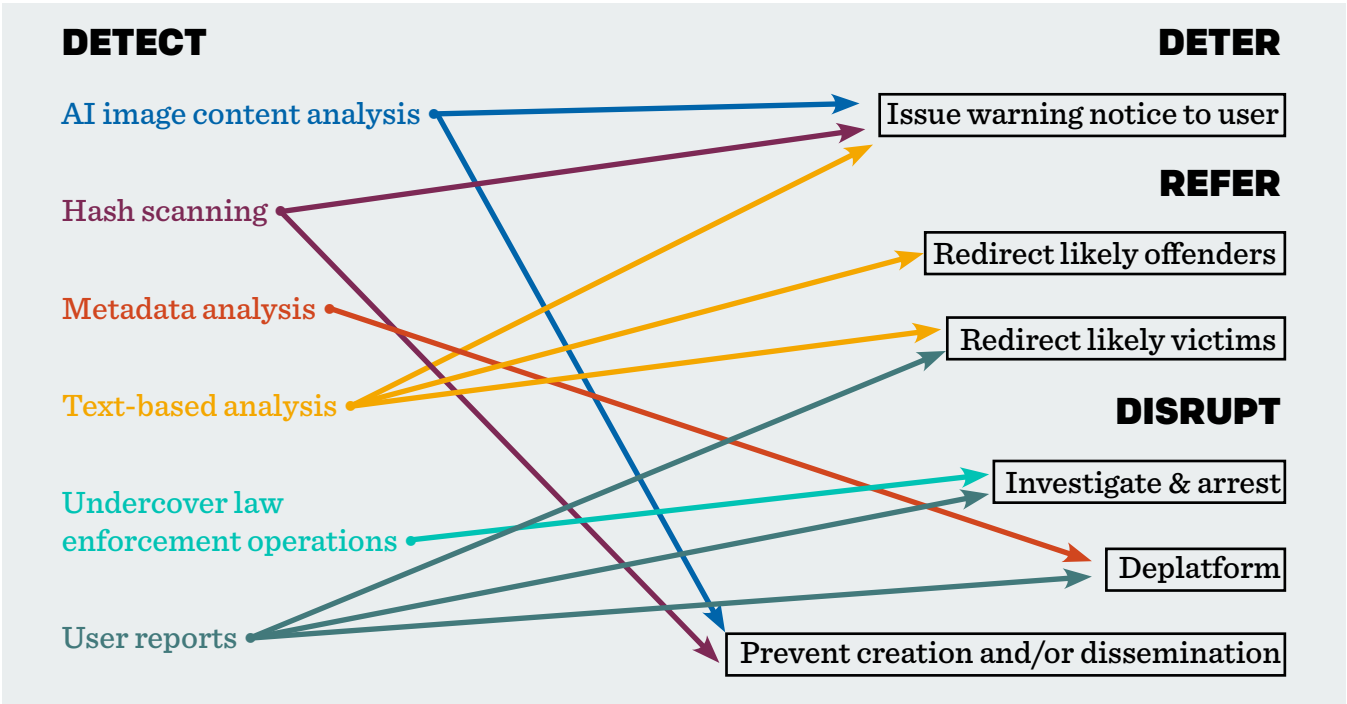
The proposed Jenna Quinn Law would work to provide education and resources to adults who work with children in order to prevent child sexual abuse and exploitation. To read more, see "The Jenna Quinn Law S. 924 Passed the U.S. Senate in 2020," Jenna Quinn, accessed September 24, 2022, <https://jennaquinn.net/the-jenna-quinn-law/>; "Support the Jenna Quinn Law," ProStasia Foundation, accessed September 24, 2022, <https://prostasia.org/campaign/jenna-quinn-law/>.

Upon detection of online child sexual abuse or exploitation, the question becomes, what next? Content—images and text—is necessary to pursue a law enforcement investigation, but there are many steps short of that which companies can take to combat this problem.

The first four detection methods (AI, perceptual hashing, metadata analysis, text-based analysis) are automated—the nature of the content is not assessed by a party to the communication but by some technical means. As a result, detection of online CSEA through these methods cannot be reported to a third party from an end-to-end encrypted environment. The latter two methods—undercover operations and user reports—involve authorized access to the content, and therefore detection can be reported without breaking the encryption scheme.

The following three categories of interventions—deter, refer, and disrupt—describe methods that can be coupled with detection efforts. However, not all detection methods can be paired with each subsequent option within end-to-end encrypted environments. Figure 14 (below) illustrates which detection methods can connect with which subsequent purpose.

FIGURE 14: MATCHING DETECTION METHODS WITH FEASIBLE NEXT STEPS IN END-TO-END ENCRYPTED ENVIRONMENTS



III. Deter

Deterrence may be a valid goal in combating online CSEA when it is directed at someone who does not want to cause harm—that is, people who share the content nonmaliciously, potential perpetrators, and/or victims (or potential victims) themselves. Within an end-to-end encryption scheme, companies can issue **warning notices** to users following detection of high-risk behaviors via artificial intelligence, hash matching, or text-based analysis.¹¹⁷ The language of these notices will need to be carefully tailored to the intended audience. Companies that elect to implement this type of deterrence effort should engage relevant stakeholders—including members of the target group, communications specialists, and psychological experts if the audience is either potential perpetrators or potential victims—during the drafting process.

SafeToNet’s smart keyboard provides an example of pairing text-based analysis with warning notices.¹¹⁸ Content is not shared with third parties, parents included, but prompts are included on the device for the child user. The program, which is compatible with all apps on the device, may recommend different responses or provide resources to get support or assistance.¹¹⁹

Apple’s iMessage child safety feature is an example of pairing AI image analysis with warning notices. As discussed above, this feature scans images sent or received by a device identified as belonging to a child, looking for signals suggesting the image may contain nudity.¹²⁰ Upon a positive match, the app warns the child user, letting them know the image may contain explicit material and asking them to affirmatively choose to either send or receive the image.¹²¹ The idea behind this is that when a “speed bump” is put in place, it provides children the opportunity to reconsider their actions.

Parental Reporting

Early public announcements for Apple’s iMessage child safety feature (described above) said it would send notices to users assigned as “parents” on the child’s family plan.¹²² These notices would not have included content but only a notification that a child under age 13 with this feature enabled overrode the “speed bump” and elected to view or send possibly explicit content. Many parents agreed with the premise underpinning this component, which has since been withdrawn, believing that such a provision properly places the onus on parents to educate their children about appropriate conduct.

However, as one interviewee pointed out, there is no evidence to support the premise that in the tech sphere, families are a safe space for children to discuss and/or report sexual

117 In theory, warning notices could also be provided if metadata analysis indicated that the account could be engaged in abusive behavior (e.g., by forwarding an image too many times). However, unlike the other three automated detection methods, metadata analysis is more likely to take into account a broader range of indicators and signals across the account—from account name and public-facing avatar to account behavioral indicators such as forwarding and sharing—and it is not as likely that any one indicator will be sufficient to determine the account is involved in online CSEA. As a result, timing and language of a warning notice could not be issued with any specificity.

118 As noted above, SafeToNet is continuing to develop the keyboard technology with a view to incorporating it into future versions of its solutions.

119 SafeToNet, “Features.”

120 Apple, “Communication safety in Messages.”

121 Apple collaborated with child psychologists to identify best practices in the messaging. However, there has been some criticism that the language used may actually induce risk-taking, which further illustrates how finely tuned the language in warning notices must be. See Bennett Bertenthal, Apu Kapadia, and Kurt Hugenberg, “Could Apple’s child safety feature backfire? New research shows warnings can increase risky sharing,” *The Conversation*, September 28, 2021, <https://theconversation.com/could-apples-child-safety-feature-backfire-new-research-shows-warnings-can-increase-risky-sharing-167035>.

122 Apple, “Communication safety in Messages.”

abuse.¹²³ Perpetrators of hands-on sexual abuse of children are often family members or other trusted adults, so it may not be correct to assume the plan-designated “parent” represents a safe haven to the child.

Companies should carefully consider whether it is appropriate to incorporate parental notices alongside these intervention techniques, as it may put the child at greater risk.

Warning notices may also be used on public-facing surfaces and unencrypted platforms to curb risk. For example, Roblox includes warning notices on its unencrypted platform, cautioning users from moving their conversations off Roblox to an encrypted communications application. These warning messages are intended to prevent off-platform abuse.¹²⁴ Facebook also issues warnings in its unencrypted spaces, particularly encouraging child users to accept friend requests only from people they already know. As Facebook moves toward fully encrypted messenger systems, these types of warnings on public-facing areas may play a heightened role in preventing abuse.¹²⁵

A similar approach could be paired with client-side hash matching. If Apple’s originally proposed iCloud feature¹²⁶ had warned the user that the content they sought to upload may be contraband and that possession itself is illegal in most countries, rather than trigger a notice to the company, the feature would not have raised as many privacy concerns. This sort of client-side scanning may be effective in curbing nonmalicious distribution, especially if paired with educational messaging directing the user to secure reporting functions, although it would not affect intentional offender-to-offender sharing.

However, unlike artificial intelligence analysis, the false positive rate of a hash match is negligible, which raises ethical concerns regarding what a company could or should know about a user’s actions by using this method but not reporting it. If a company *could* know that a user is in possession of known CSAM, for example, by triggering the creation of a security voucher upon a positive match like in Apple’s original proposal, what are the implications of the company’s choice not to? Currently, having the ability to implement a technical solution to identify CSAM-trading accounts does not correspond with legal liability, but it certainly raises the question about what ethical obligations a company owes its users, including the victims of these crimes.

IV. Refer

Companies with search functions—such as Google, Microsoft, Facebook, and Apple—can identify search terms related to child sexual exploitation and abuse used by either (potential) perpetrators, (potential) victims, or someone who wants to file a related report.¹²⁷ Relying on these signals, the company can populate the **search results with referrals to support services** to intervene and either prevent further offending, remove a child from an abusive situation, or direct the user to reporting platforms.

¹²³ Interview by author, January 12, 2021.

¹²⁴ Interview by author, November 19, 2021.

¹²⁵ Interview by author, July 13, 2021.

¹²⁶ This is the feature discussed above, which has since been rescinded, wherein images uploaded from the user’s device to their iCloud account would be scanned against a hash match database client-side, generating security vouchers accessible by the company only when a threshold number of positive matches was reached. See Apple, *Expanded Protections for Children*, August 2021, https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Technology_Summary.pdf.

¹²⁷ Interview by author, July 1, 2021; interview by author, July 2, 2021; interview by author, July 13, 2021; “Expanded Protections for Children,” Apple, accessed September 24, 2022; <https://www.apple.com/child-safety/>.

There are excellent resources online where either (potential) perpetrators or (potential) victims can receive support. However, in-person care may also be necessary, which requires investment in community-based supports.

V. Disrupt

“Disruption” in this context means to minimize the amount of CSAM circulating on the internet, even if only for a short time, by restricting actions of the demand-side offenders and the producers. There are three primary ways to pursue this goal: (1) investigate and arrest, (2) deplatform suspected offenders, and (3) prevent the production and/or dissemination of CSAM.

A. INVESTIGATE AND ARREST

In many circumstances, a criminal justice outcome may be the most appropriate for offenders involved in online CSEA, in particular for those offenders who produce CSAM, exploit or entice children to create PFP content, and/or traffic children, either domestically or in the context of travel and tourism. With end-to-end encryption, ESPs will not be able to provide the content of users’ communication to NCMEC as part of their ordinary reporting processes, which will greatly limit law enforcement’s ability to rely on these reports to generate probable cause to open investigations and obtain search warrants for abusive accounts. Because content is necessary for these aims, law enforcement engagement will follow only from undercover operations or user reports in which the user provides content or is willing to cooperate with the investigation.

This method of intervention in the patterns of harm is likely the most resource intensive,¹²⁸ but the commitment of law enforcement officials to pursue these investigations provides an important deterrence effect, which may have ripple effects beyond those who are ultimately arrested and prosecuted.

Arrest is not an appropriate outcome in every situation. Although the mere possession of CSAM is criminalized in the United States, there are circumstances when arrest would not yield a just outcome. For instance, children who create PFP content would be better served by referral to support services, or people who share CSAM out of outrage may be warned about the impact of their actions and directed to alternative ways to report the crimes.

B. DEPLATFORM ABUSIVE ACCOUNT

As private entities, ESPs can deplatform accounts that violate their terms of service, and they can do so without raising First Amendment questions, so long as they do so pursuant to their own policies and not as directed by the government.¹²⁹ For detection methods that cannot provide content, such as metadata analysis, this may be the most effective method of disruption, as it allows companies to remove an account from their platform so that the account can no longer engage in this harmful behavior.¹³⁰

128 Interview by author, August 6, 2021; interview by author, August 11, 2021; interview by author, September 7, 2021; interview by author, September 15, 2021; interview by author, September 27, 2021.

129 See, e.g., Paul Levinson, “I’m a First Amendment scholar — and I think Big Tech should be left alone,” *The Conversation*, January 20, 2021, <https://theconversation.com/im-a-first-amendment-scholar-and-i-think-big-tech-should-be-left-alone-153287>; Audie Cornish and Daphne Keller, “Deplatforming: Not a First Amendment Issue, but Still a Tough Call for Big Tech,” January 26, 2021, *Consider This from NPR*, National Public Radio, <https://www.npr.org/transcripts/959667930>.

130 WhatsApp reports that it bans more than 300,000 accounts per month on suspicion of sharing CSAM. WhatsApp, “How WhatsApp Helps Fight Child Exploitation.”

Although companies can deplatform a specific account based on suspicion of engaging in online CSEA, nothing stops that user from creating a new account to engage in this behavior, either on the original platform or on another. Complicating this is the fact that ESPs may be restricted from sharing user data, including information about a deplatformed account, with one another, which otherwise could prevent the user from creating an account on a different platform.¹³¹ For instance, if Company A deplatformed an account with the recovery email address abc123@xyz.com, Company A may not be able to share this information with Company XYZ, nor with any other ESP that may have an account registered using the same email address.¹³²

Ultimately, although assuming deplatformed users will jump to another platform or create a new account, law enforcement officials have said that even this minor disruption is a step in the right direction. This intervention can deter less tech-savvy offenders from continued engagement and keep persistent offenders off-balance.

C. PREVENT CREATION AND/OR DISSEMINATION OF CSAM

The last permutation for disruption is to use technical means to either prevent the creation of CSAM or prevent its dissemination.

To prevent the creation of CSAM, safety tech firm SafeToNet is developing prototypes of AI-enabled threat detection, which when coupled with a device's camera, could prevent the recording of CSAM, either videos or still images. To date, the firm has demonstrated the tool's potential to detect other categories of harm and has started training the solution on CSAM.¹³³ It expects that the tool will ultimately be capable of detecting and preventing the recording and distribution of CSAM in real time.

Project Arachnid & Removal Orders¹³⁴

As discussed above under detection methods, the Canadian Centre for Child Protection operates Project Arachnid, which has developed an API that companies can deploy to detect *and* remove CSAM from their systems. This cannot be deployed in an end-to-end encrypted environment, but by removing this content from the open web, it is less likely to migrate into encrypted channels.

Detection methods, especially those with known high rates of reliability such as perceptual hashing, could be deployed before the user uploads an image or shares the image via a messaging app. This could occur either on a device before a user uploaded the image from the device to a particular app or as part of a specific application's processing procedures prior to encryption. ESPs could then impose a gatekeeping function wherein they restricted users from sending or uploading photos that matched a hash value from the CSAM database.

¹³¹ Interviewees in the United States referenced the Stored Communications Act, 18 U.S.C. § 2701, and the Electronic Communications Privacy Act, 18 U.S.C. § 2511, as two laws that prohibit ESPs from engaging in more proactive efforts to share account information among one another. There may be liability carve-outs when it comes to sharing information in order to combat online CSEA, but absent explicit legislative approval to share account indicators in these instances, companies are not likely to take the risk.

¹³² There are good privacy-preserving reasons to support this. For example, the email address that the original account holder used to register the account may have been hacked, and a company sharing the information with other companies could further exacerbate the harm to the owner of the hacked account.

¹³³ Interview by author, August 11, 2021; interview by author, September 8, 2021; interview by author, October 12, 2022. For a panel interview that featured Sharon Pursey, co-founder of SafeToNet, discussing this method, see International Justice Mission, "Day 3: Global Tech Innovations // Global Resolve Against Online Sexual Exploitation of Children 2020."

¹³⁴ Interview by author, May 26, 2022. For more information, see Canadian Centre for Child Protection, "Project Arachnid."

Neither of these technical methods used to limit the supply of CSAM would directly support the law enforcement process, because there would be no third-party notice. However, with respect to the second technique, by restricting sharing, there would likely be a dramatic decrease in known CSAM circulating within the internet ecosystem and a resultant decrease in reports to NCMEC. Law enforcement officials have lamented the sheer volume of reports received via NCMEC, which often results in Internet Crimes Against Children squads focusing exclusively on reactive investigations, rather than proactive ones.¹³⁵ By decreasing the quantity of known CSAM, law enforcement's limited resources could be redirected to focus on investigations of new content where children may currently be in imminent physical danger.

¹³⁵ Interview by author, August 6, 2021; interview by author, August 11, 2021; interview by author, September 7, 2021; interview by author, September 15, 2021; interview by author, September 27, 2021.

RECOMMENDATIONS

Adoption of end-to-end encryption will drastically diminish ESPs' ability to detect and report CSEA on their services. Given the serious harms that will occur absent timely intervention, companies that elect to implement end-to-end encryption must consider what measures they will take to balance these risks. But tech companies cannot solve the problem in isolation; governments and civil society have important roles to play too, and they must work collectively to combat online CSEA.

These recommendations build on the experience and expertise of the interviewees who contributed to this project, and fall into three primary categories. Policymakers should (1) improve the report-to-prosecution pipeline, (2) engage relevant stakeholders, and (3) prioritize upstream efforts.

I. Improve the Report-to-Prosecution Pipeline

End-to-end encryption does not explicitly interrupt the complex system that ultimately leads to criminal justice intervention in online CSEA cases. However, because ESPs will no longer be able to access the content of users' communications, they will report far fewer cases to NCMEC, meaning law enforcement will investigate, and ultimately prosecute, fewer cases. It is therefore necessary to ensure that this system operates at peak performance to support these cases. Here are five specific actions that policymakers can take to achieve this.

A. SET INDUSTRY STANDARDS FOR DATA COLLECTION AND RETENTION BY TECH COMPANIES

Even companies with strong reporting practices may not collect useful account indicators, or if they do, they may not retain them for sufficient time. Interviewees said that even once a company files a report to NCMEC, it may still subject the reported account data to the company's standard retention period, which in some instances is only thirty days. As described above in part 1, many steps must occur before law enforcement officials are able to proactively investigate and seek additional information from the reporting provider. Often, by the time law enforcement has prepared a search warrant, the retention period has passed and the original account information has been purged from the company's system.

There are good reasons why the government should not impose specific data collection and retention rules on tech companies. However, tech companies could seek to establish industry standards as part of their work in various intra-industry collectives. They should take into account the law enforcement process above and collaborate with law enforcement representatives in order to ensure that whatever standards they identify will serve the intended purpose of better supporting law enforcement efforts. With respect to data collection, ESPs should establish what data should be collected for any account. This will vary by the type of service provided, but registration data and certain metadata indicators would be an appropriate starting point. As for data retention, ESPs may want to consider creating standards that apply only to those accounts that have triggered a report to NCMEC, wherein that account data is moved to an encrypted repository and subject to a different retention period. Companies could treat this data as they would data retained pursuant to a litigation hold, walling it off from periodic deletion.

B. CREATE UNIFORM CRITERIA FOR LAWFUL DATA REQUESTS TO TECH COMPANIES

Law enforcement officials often reported that to be successful in conducting this type of investigation, they needed to know what rules to follow for each ESP when serving legal process (e.g., subpoenas, search warrants). Even though the legal standard is the same, each company imposes different rules and processes on law enforcement, which leaves the burden on law enforcement to navigate the system and tailor each request for data based on the company.¹³⁶ Law enforcement reported that it relied heavily on personal connections with particular ESPs—an identified point of contact whom it could call.

While a single point of contact is useful in theory, if the efficacy of the process relies on that connection, then that person also serves as a single point of failure. Staff vacations or promotions, for example, can disrupt the entire apparatus if ESPs do not have in place official policies and practices for responding to law enforcement's requests. One interviewee noted that despite ensuring uniformity in search warrant language in order to obtain similar sets of data, they received differently scoped replies based on who at the given company received and processed the request. Creating industry standard practices would ensure consistency while still holding law enforcement to the necessary standard of proof.

C. STREAMLINE LAW ENFORCEMENT'S REPORTING METHOD TO NCMEC

When an investigation reveals a cache of CSAM, law enforcement is expected to submit new images to NCMEC for inclusion in the hash database. Unfortunately, at that stage in the investigation, law enforcement likely has other, more time-sensitive matters to attend to, such as reviewing and cataloging evidence and identifying victims. As a result, this report to NCMEC may get skipped. Given the heavy reliance on the NCMEC-maintained database, it is crucial that it remains as up-to-date as possible, which means creating a streamlined method to support this feedback loop.

D. ESTABLISH GUIDELINES FOR TRIAGING REPORTS

NCMEC sends reports to Internet Crimes Against Children (ICAC) task forces throughout the United States. These task forces typically include representatives from police departments and sheriff's offices throughout the region as well as federal investigators, including the FBI and Homeland Security Investigations. Each task force designates one or more people to review all the reports received by NCMEC and triage them to focus resources. Triage decisions are often based on perceived seriousness of injury, the estimated age of the victim, and whether the material is new, thereby suggesting ongoing hands-on harm, as well as more mundane considerations, such as appropriate jurisdiction or whether the report includes sufficient information to open an investigation. Law enforcement officials typically rely on their experience to conduct this triage. The development of more specific guidance or recommendations could provide a useful tool, especially for officers new to investigating this type of offense.

NCMEC could aid in this process. Interviewees reported that it would be helpful to know upon receipt of a report from NCMEC whether the content matches a known image in the database or is new. If there were standard criteria that ICAC intake officers considered when triaging the reports, those could also be included in the report from NCMEC to law enforcement.

¹³⁶ To be clear, whenever law enforcement executes a request for data, it must meet the legal threshold. By proposing legal process standards across the industry, that is only to streamline the workflow for law enforcement and not to diminish its burden of proof.

Identifying key factors about a given image may become easier following completion of a project managed by the Internet Watch Foundation. Funded by Thorn, the Internet Watch Foundation has employed a handful of analysts whose sole responsibility is to “grade” images of CSAM received from the repository held by the UK Home Office.¹³⁷ The analysts layer contextual metadata and generate unique hash values for each image. This information could be accessed by NCMEC analysts and passed along to law enforcement to streamline the triage process.

E. ADDRESS VICTIMS’ NEEDS THROUGHOUT THE PROCESS

This work is hard, emotionally draining, and often overwhelming. Everyone involved in this pipeline—from ESP content reviewers to NCMEC analysts to investigators to prosecutors to social workers—experiences burnout. Yet interviewees reported high levels of satisfaction with their work because they know they are making a difference in children’s lives. Centering the victims at every stage of this pipeline is not only necessary to stay engaged in the work, but it is ethically necessary given the inherent vulnerability of the victim population.

A few concrete actions can be taken to ensure victims remain centered. First, given the amount of digital evidence in online CSEA cases, prosecutors may be tempted to rely on texts and images rather than a victim’s testimony. For some victims, this might be preferable, as they may be reluctant to testify and would prefer not to relive the trauma of the original victimization. However, for others, this approach fails to take into account their need for reconciliation achieved through the confrontational and adversarial court process. Prosecutors should therefore engage directly with victims to understand what role they want to play in the process, while also connecting victims to necessary social support services.

In some instances, it may be appropriate to seek the assignment of an attorney, client advocate, or guardian ad litem who understands the system and can help the child navigate the process. Advocating Opportunity, a nongovernmental organization in Toledo, Ohio, works directly with trafficked and exploited persons, assigning both an attorney and client advocate to each client.¹³⁸ Using a trauma-responsive care model, this team works consistently with the client until the situation has been resolved or the client opts out of further services.

Victims often require ongoing support services that may outlast the criminal case. To the extent possible, victims should be connected to victim compensation funds, and when legally viable, financial restitution from the offenders should be sought. The Amy, Vicky, and Andy Child Pornography Victim Assistance Act of 2018 (P.L. No. 115-299), signed into law in 2018, allows these victims to receive a one-time payment of \$35,000 from the Child Pornography Victims Reserve within the Crime Victims Fund and also outlines restitution orders that can be placed upon convicted offenders. Although the funds for the reserve have been allocated, regulations outlining the process for victims to receive these one-time payments have not been finalized, blocking victims from receiving much needed financial support.

¹³⁷ Internet Watch Foundation, “IntelliGrade.”

¹³⁸ To learn more about Advocating Opportunity, see <http://www.advocatingopportunity.com/>.

II. Engage Relevant Stakeholders

Policymakers must think broadly when identifying and engaging stakeholders. Firms outside the typical “child safety” sphere can make meaningful efforts to combat this problem. A range of stakeholders, including the tech industry, civil society, law enforcement, and beyond, can contribute in myriad ways. There are four recommendations relating to assembling the right team.

A. IDENTIFY WHO CAN INTERVENE

For each intervention, it is critical to identify who has the power to implement the change. The nature of the intervention will dictate the decision-maker.

In many cases, the interventions described above require action by the ESP—online safety education for children, incorporation of safety-by-design principles, automated detection methods, warning notices, referrals in search returns, deplatforming, and technical prevention of creation and dissemination. However, many of these interventions will be most effective if they are developed in concert with victims’ rights groups, psychologists, and educators. These parties with relevant expertise should be brought in at the start of the development process.

Several intervention methods require action by civil society and local governments. Charitable organizations can partner with local governments to establish community spaces for marginalized youth. Public libraries and schools can support education efforts. Mental health departments can ensure that services are available for the prevention of CSEA and to support victims.

Whenever an intervention is targeted at a population in a specific region of the world, policymakers should partner with local experts to ensure regional specificity in language, tone, and messaging.

B. INVEST IN PARTNERSHIPS

One of the key themes that emerged from the research is that partnerships are essential in combating online CSEA. These may be intra-industry partnerships, such as the Tech Coalition¹³⁹ or ICAC task forces,¹⁴⁰ or interagency coalitions, such as the WeProtect Global Alliance¹⁴¹ and NCMEC-hosted roundtable events. These partnerships are necessary to keep the field nimble and support coordinated tracking of emerging patterns and behaviors. It also allows for information sharing and development of new and innovative solutions, such as those that have stemmed from child-safety hackathons and open-source projects to improve CSEA detection technology.

Examples of innovation and promising practices stemming from strong partnerships include the following:

- **InHope**, the International Association of Internet Hotlines (also known as tiplines), is a global network of 50 hotlines where the public can anonymously report CSAM online. It operates a data-exchange platform and an integrated report-management system that is hosted by Interpol. This

139 The Tech Coalition is “an alliance of global tech companies who are working together to combat child sexual exploitation and abuse online.” It is a membership organization composed of technology companies—large and small, old and new. To learn more, go to www.technologycoalition.org/about.

140 The Internet Crimes Against Children (ICAC) Task Force Program is “a national network of 61 coordinated task forces, representing over 5,400 federal, state, and local law enforcement, dedicated to investigating, prosecuting and developing effective responses to internet crimes against children.” To learn more, go to www.icactaskforce.org/.

141 The “WeProtect Global Alliance is a global movement of people and organisations” that “work together to transform the global response to child sexual exploitation and abuse online.” WeProtect is a “public-private partnership” with participation from civil society, international organizations, governments, and the private sector. To learn more, go to www.weprotect.org/alliance/.

allows hotlines from different countries to instantly and securely share CSAM-related reports. This is crucial for situations in which Country A receives a report that includes content hosted in Countries B, C, and D. InHope's system allows the tipline in Country A to seamlessly share the relevant information with its peers in Country B, Country C, and Country D to facilitate the rapid removal of CSAM from the internet.

- **Zoom** worked closely with **NCMEC** and the **FBI** to curb the troubling trend of CSAM meeting disruptions, wherein offenders disrupted Zoom meetings with short CSAM videos, something that cropped up at the start of the coronavirus pandemic. Zoom worked closely with an international task force, spearheaded by the FBI, to identify offending accounts and gather information necessary for legal process. Zoom also strengthened ties with NCMEC by building an API so that once a user filed a report to Zoom, an analyst could immediately review it and report it to NCMEC.¹⁴² Collectively, these efforts helped to effectively eliminate this problem.

C. LEVERAGE INCENTIVE STRUCTURES

For for-profit companies, a cost-benefit analysis will be conducted as part of any decision to implement a new program or initiative. Many of the interventions identified may be costly and resource intensive, especially for smaller or newer tech companies. There are external players who can press on the scale to incentivize these companies to make decisions that will aid in the fight against online CSEA. Even though the company leadership can decide whether or not to implement any given intervention, certain incentive structures may serve as a powerful lever.

Financial services companies, such as Mastercard or Visa, can include language in contracts or terms of service to tackle this issue. This nudge must be narrowly tailored to ensure it does not overextend and negatively impact lawful and consensual activities.¹⁴³

On-device app stores can play a similar function by requiring certain safety elements to be addressed before an app is made available through that store. This could induce app developers to make safety-driven design decisions.

Investment companies and venture capital firms can also play an important role in setting standards at the outset of a product's development. For example, the eSafety Commissioner's office in Australia is engaging directly with the investment community to encourage the incorporation of safety by design into investment criteria. By incentivizing the incorporation of these design principles at the outset, new apps and platforms are more likely to lead with safety.

Lawmakers can motivate ESPs to take certain precautionary measures and safety interventions through any number of statutory or regulatory changes. Tax benefits could be established for companies that implement best practices in this area. Explicit liability carve-outs could be created to permit ESPs to share account indicators with one another when an account is suspected of engaging in online CSEA. Or an exception to the prohibition on CSAM possession could be narrowly fashioned to permit tech safety researchers the opportunity to improve classifiers and other detection technology.

¹⁴² Interview by author, September 23, 2021. For more information about Zoom's efforts to combat meeting disruptions, go to blog.zoom.us/new-ways-to-combat-zoom-meeting-disruptions.

¹⁴³ The negative impacts of overly broad policies are widely documented. See, e.g., Issie Lapowsky, "OnlyFans shows Visa and Mastercard are 'choke-points' of online speech," Protocol, August 20, 2021, <https://www.protocol.com/policy/onlyfans-visa-mastercard>.

These levers for change can also be combined. For example, creating national or international standards for specific apps may leave regulators playing a virtual game of whack-a-mole as developers create a roving range of available apps. If the focus of regulation is instead placed at the point of sale or download—i.e., the app marketplace, such as Apple’s App Store or Google Play—then regulators may have an opportunity to yield a greater impact. For example, app stores have age ratings for different apps, but they are underenforced and inconsistent, and stores have promoted apps outside the user’s stated age range.¹⁴⁴ This is an area ripe for review by Apple and Google.

D. EXPLORE ALTERNATIVE SOURCES OF SIGNALS

Financial institutions can incentivize responsible corporate decision-making, but they are also well situated to flag early warning signs of behavior indicative of certain patterns of harm, specifically in the context of travel and tourism and live-streamed abuse. Banks and international money-transfer companies should work closely with law enforcement to identify behaviors indicative of CSEA. This would likely be most effective in combating live online child sexual abuse and the sexual exploitation of children in the context of travel and tourism. These offenses yield peculiar and specific money-transfer activities, which could be helpful evidence to identify these offenders.

Additionally, as noted in the *Luxembourg Guidelines*, “[s]pecific travel/tourism actors in the circuit of child sexual exploitation (such as hotels, travel agencies, tour operators, transportation companies, airlines, bars, and restaurants) become, knowingly or not, intermediaries in the commission of these offences, and can also play a role in their prevention.”¹⁴⁵ Policymakers should invite these stakeholders into the conversation and provide them with the information necessary to serve as good partners.

III. Prioritize Upstream Efforts

End-to-end encryption obscures the content of communications between offenders, so policymakers need to make a greater investment in interventions that can break the patterns of exploitation and abuse earlier in the process. These recommendations can be loosely grouped into community-based opportunities and tech-based opportunities.

A. INVEST IN COMMUNITY-BASED OPPORTUNITIES

Several of the interventions discussed above designed to prevent exploitation and abuse are entirely community-based efforts. These include building community spaces for marginalized youth, providing comprehensive sex education and tech safety education, and supporting perpetration prevention efforts. There are pockets of success that have built and sustained these programs, but they need to be more widely adopted and institutionalized for broader impact.

Law enforcement professionals consistently remarked on how time- and resource-intensive undercover operations are. To be effective in this type of investigation, officers require a great deal of training, support, and technical know-how. Governmental decision-makers should make targeted

¹⁴⁴ Canadian Centre for Child Protection, *Reviewing the Enforcement of App Age Ratings in Apple’s App Store and Google Play*, 2022, 3, https://www.protectchildren.ca/pdfs/C3P_AppAgeRatingReport_en.pdf.

¹⁴⁵ Interagency Working Group, *Luxembourg Guidelines*, 55.

ECPAT International called together an Interagency Working Group in 2014 with representatives from key stakeholders in the child safety space and chaired by Professor Jaap Doek, former chair of the UN Committee on the Rights of the Child. Representatives in the Interagency Working Group conducted an in-depth analysis of terminology and definitions for terms used to describe different forms of sexual exploitation and abuse of children. The *Terminology Guidelines for the Protection of Children from Sexual Exploitation and Sexual Abuse*, also known as the *Luxembourg Guidelines*, represents the outcome of this initiative. The terms discussed in the *Luxembourg Guidelines* are “meant to be ‘universal’ and applicable to work against these phenomena in all settings.” Interagency Working Group, *Luxembourg Guidelines*, 1.

investments in these efforts. That is not to suggest that lawmakers should offer a blank check to police to enhance broad surveillance efforts, but rather provide the units engaged in these investigations with sufficient staffing, training, and support to be effective.

B. RESEARCH TECH-BASED INTERVENTIONS

Some tech-based opportunities can be deemed “low-hanging fruit” because they present low-cost efforts that help address online CSEA. These include in-app online safety messaging targeted to children; public awareness campaigns directed at populations most likely to share CSAM without malicious intent; improved user-reporting functionality; and specific and directed warning notices to deter users from escalating behavior. These efforts build on existing frameworks and can be implemented without much additional investment.

Other opportunities require tech companies to make a larger investment. Specifically, because end-to-end encryption blocks ESPs from identifying and reporting CSAM on their platforms, they should build systems that will prevent it from entering and circulating in the internet ecosystem. For example, tech companies can work collaboratively with researchers to craft new detection programs, building on tech such as perceptual hashing and existing AI models, that could work as a filter, restricting a user’s ability to upload known CSAM to an app or program. This could occur on-device before an image is encrypted.¹⁴⁶

Although AI technical solutions for threat detection, such as the in-camera program in development by SafeToNet, remain in the testing and development stage, this sort of creative application is another opportunity for tech companies to invest in to help prevent the production of CSAM.

¹⁴⁶ ESPs would likely want to create an appeals process wherein users could submit the photo they were restricted from sending and a human reviewer, upon assessing no violation, could tag the content as approved so the issue would not crop up again for that particular image. Given that only users with true false positives would submit an image using this process, it should not overwhelm human moderation teams.

APPENDIX 1: TERMINOLOGY

Definitions and explanations for terms are included here and are widely drawn from the *Luxembourg Guidelines*, the U.S. Criminal Code, scholarly reports, and experts in the field. Disagreements between those sources are noted.

• Child Pornography

The U.S. Criminal Code defines this term as “any visual depiction, including any photograph, film, video, picture, or computer or computer-generated image or picture, whether made or produced by electronic, mechanical, or other means, of sexually explicit conduct, where—(A) the production of such visual depiction involves the use of a minor engaging in sexually explicit conduct; (B) such visual depiction is a digital image, computer image, or computer-generated image that is, or is indistinguishable from, that of a minor engaging in sexually explicit conduct; or (C) such visual depiction has been created, adapted, or modified to appear that an identifiable minor is engaging in sexually explicit conduct.”¹⁴⁷

The *Luxembourg Guidelines*, along with experts in this field, discourage the use of this term altogether. They suggest that “sexualised material that depicts or otherwise represents children is indeed a representation, and a form, of child sexual abuse, and should not be described as ‘pornography’. Pornography is a term primarily used for adults engaging in consensual sexual acts distributed (often legally) to the general public for their sexual pleasure. Criticism of this term in relation to children comes from the fact that ‘pornography’ is increasingly normalised and may (inadvertently or not) contribute to diminishing the gravity of, trivialising, or even legitimising what is actually sexual abuse and/or sexual exploitation of children.”¹⁴⁸

To the extent this term is used throughout this report, it is only in connection to the language used in the U.S. Criminal Code.

• Child Sex Trafficking

Under the U.S. Criminal Code, this offense is referred to as “sex trafficking of children or by force, fraud, or coercion” and is defined as knowingly “recruit[ing], entic[ing], harbor[ing], transport[ing], provid[ing], obtain[ing], advertis[ing], maintain[ing], patroniz[ing], or solicit[ing] by any means” a person who “has not attained the age of 18 years and will be caused to engage in a commercial sex act.”¹⁴⁹ As a federal offense, jurisdiction is established when the described actions occur in or affect interstate or foreign commerce, but this does not require that the child is physically moved, as the common understanding of “trafficking” would otherwise suggest. Instead, as explained in the *Luxembourg Guidelines*, “a consistent feature of ‘trafficking’ under international law is that its purpose is the exploitation of a human being (in this case the child).”¹⁵⁰

This term is used to describe the actions that may otherwise be known as “sexual exploitation of children in/for prostitution” or “child prostitution,” the former of which is actually the preferred term suggested by the *Luxembourg Guidelines*. “This form of exploitation consists of a child performing a

¹⁴⁷ 18 U.S.C. § 2256(8).

¹⁴⁸ Interagency Working Group, *Luxembourg Guidelines*, 38–39.

¹⁴⁹ 18 U.S.C. § 1591.

¹⁵⁰ Interagency Working Group, *Luxembourg Guidelines*, 60.

sexual act in exchange for (a promise of) something of value (money, objects, shelter, food, drugs, etc). It is not necessarily the child who receives the object of exchange, but often a third person. Moreover, it is not necessary that an object of exchange is actually given; the mere promise of an exchange suffices, even if it is never fulfilled.”¹⁵¹ The *Luxembourg Guidelines* go on to explain the concerns with the term “child prostitution,” as it “may arguably be interpreted in a manner to imply that the phenomenon represents a legitimate form of sex work or that the child has given her/his informed consent to prostitute her/himself.”¹⁵²

Interviewees argued that the inclusion of the word “prostitution” to describe this offense, even when used in conjunction with “exploitation,” should be avoided entirely, as it suggests the child could consent to the arrangement, which is legally dissonant.

• **Child Sexual Abuse Material (CSAM)**

This term is typically used to replace what has otherwise been known as “child pornography,” due to the widely accepted argument that “sexualised material that depicts or otherwise represents children is indeed a representation, and a form, of child sexual abuse.”¹⁵³ For the reasons discussed above, the latter term is no longer accepted among child safety advocates and law enforcement officials who work in this field.

There is some disagreement about whether CSAM is the most accurate term or whether “child sexual exploitation material” (CSEM) better encapsulates the range of material previously understood to be “child pornography.” In particular, the concern is that CSAM applies more narrowly to those images that depict an act of sexual abuse against a child and that it does not extend to include images that sexualize and exploit a child but do not explicitly depict abuse.¹⁵⁴

Despite the fact that CSEM may better describe the images at the core of this issue, CSAM is a more broadly accepted term. Therefore, when used throughout this report, it is intended to cover the broader category of harmful imagery.

• **Child Sexual Exploitation and Abuse (CSEA)**

As with CSAM and CSEM, there is nuance in the difference between child sexual abuse and child sexual exploitation. As the *Luxembourg Guidelines* state: “The [UN Convention on the Rights of the Child] does not make clear what the distinction is [between the two terms]. However, it is noteworthy that the sexual abuse of children requires no element of exchange and can occur for the mere purpose of the sexual gratification of the person committing the act, whereas the sexual exploitation of children can be distinguished by an underlying notion of exchange.”¹⁵⁵ It is also important to note that sexual abuse and/or exploitation requires neither physical contact nor explicit force.

Complicating the matter is the fact that the U.S. federal crime of “sexual exploitation of children” includes actions that do not require an element of exchange—the feature articulated as distinguishing

151 Interagency Working Group, *Luxembourg Guidelines*, 29.

152 Interagency Working Group, *Luxembourg Guidelines*, 30.

153 Interagency Working Group, *Luxembourg Guidelines*, 38.

154 For a discussion of this, see Interagency Working Group, *Luxembourg Guidelines*, 39, citing Danijela Frangež, Anton Toni Klančnik, Mojca Žagar Karer, Bjørn-Erik Ludvigsen, Jarosław Kończyk, Fernando Ruiz Perez, Mikko Veijalainen, and Maurine Lewin, “The Importance of Terminology Related to Child Sexual Exploitation,” *Journal of Criminal Investigation and Criminology* 66, no. 4 (2015): 296.

155 Interagency Working Group, *Luxembourg Guidelines*, 18, referencing Convention on the Rights of the Child, UN General Assembly, November 20, 1989.

the terms in the *Luxembourg Guidelines*. Under U.S. law, “[a]ny person who employs, uses, persuades, induces, entices, or coerces any minor to engage in ... any sexually explicit conduct for the purpose of producing any visual depiction of such conduct” is subject to criminal liability under this statute.¹⁵⁶

Given the discrepancies in definitions and the idea that the two terms are ultimately communicating similar ideas, this report treats them collectively unless otherwise delineated.

- **Demand-Side Offender**

This term was recommended by interviewees to describe the offender who engages in child sexual abuse and/or exploitation for the offender’s own sexual gratification. This is in contrast to someone who may produce such imagery for commercial gain.

- **Distribution**

Under U.S. law, CSAM-related offenses are often bucketed in the following way, with increasing penalties attached: (1) possession, (2) receipt and/or distribution, and (3) production. While possession and receipt have commonly understood definitions,¹⁵⁷ distribution is legally broader than one might expect. It includes actively sending the images to another but also may include providing someone access to them, for instance via a shared file system.¹⁵⁸

- **Grooming**

“In the context of child sexual exploitation and sexual abuse, ‘grooming’ is the short name for the solicitation of children for sexual purposes. ‘Grooming/online grooming’ refers to the process of establishing/building a relationship with a child either in person or through the use of the Internet or other digital technologies to facilitate *either online or offline* sexual contact with that person” (emphasis in original).¹⁵⁹ The *Luxembourg Guidelines* continue, stating that “[g]rooming” has sometimes also been defined as ‘*online enticement of children for sexual acts*’. ‘Enticement’ refers to something used to attract or tempt someone, which indeed reflects a common way of proceeding of the person ‘grooming’ a victim with the aim of sexually exploiting her/him” (emphasis in original).¹⁶⁰

Whereas “grooming” is not found in the U.S. Criminal Code, “enticement” is used throughout the sections related to child sexual abuse and/or exploitation. “Coercion and enticement” are explicitly listed as an offense, creating liability for “[w]hoever ... knowingly persuades, induces, entices, or coerces any [minor] to engage in prostitution or any sexual activity.”¹⁶¹

There are two primary issues to consider with the term “grooming.” First, it has been coopted by some to mean something much broader, including adults normalizing conversations around sexual orientation and/or gender identity.¹⁶² That conflation undermines its meaning in the context of child sexual abuse and exploitation. Second, “grooming” as defined here still encompasses a wide

¹⁵⁶ 18 U.S.C. § 2251(c).

¹⁵⁷ “Production” in this context is discussed below.

¹⁵⁸ See, e.g., *United States v. Stitz*, 877 F.3d 533, 538 (4th Cir. 2017) (“We now take this opportunity to join our sister circuits and hold that where files have been downloaded from a defendant’s shared folder, use of a peer-to-peer file-sharing program constitutes ‘distribution’ pursuant to 18 U.S.C. § 2252A.”)

¹⁵⁹ Interagency Working Group, *Luxembourg Guidelines*, 51.

¹⁶⁰ Interagency Working Group, *Luxembourg Guidelines*, 52.

¹⁶¹ 18 U.S.C. § 2422(b).

¹⁶² See, e.g., Hannah Natanson and Moriah Balingit, “Teachers who mention sexuality are ‘grooming’ kids, conservatives say,” *Washington Post*, April 5, 2022, <https://www.washingtonpost.com/education/2022/04/05/teachers-groomers-pedophiles-dont-say-gay/>.

range of behavior and may extend from compliments and rapport building, which would likely be constitutionally protected speech, to enticement or inducement to take sexually explicit images, which is a crime.

Given the complicated nature of the term, throughout this report it is coupled with clarifying language to ensure consistent understanding.

- **Hash Value**

“Hash values” are typically understood to be the digital equivalent of a fingerprint—believed to be unique—and their value resides in the ability to compare them against a known set of comparators. “Hashing is arguably the most widely used technology for limiting the spread of previously identified multimedia content. ... After an offending audio, image, or video is identified (either manually or automatically), a distinct digital signature is extracted from the content. ... The same hash is extracted from each future upload and compared against a database of offending hashes. ... Any matched content can, for example, be automatically blocked from upload or subjected to any policy a service provider initiates. ... In the digital realm, hashing refers to chopping a data file into small pieces and combining them to yield a concise numeric value that can be used to identify the original data file.”¹⁶³ This is the technology used in PhotoDNA and many other automated CSAM detection methods.

- **Live Online Child Sexual Abuse**

“[L]ive online child sexual abuse often represents a dual abuse of the child. She/he is coerced to participate in sexual activities, alone or with other persons—an act that already constitutes sexual abuse. The sexual activity is, at the same time, transmitted live through ICT and watched by others remotely. Often, the persons watching remotely are the persons who have requested and/or ordered the sexual abuse of the child, dictating how the act should be carried out ..., and those persons may be paying for the abuse to take place. Live online child sexual abuse has been observed to take on both commercial and non-commercial forms, and there are cases where it has been set up as a proper business with the only apparent objective being to make money out of the sexual exploitation of the children involved.”¹⁶⁴ The *Luxembourg Guidelines* make the important observation that the “fact that live child sexual abuse can now occur online through the use of ICTs does not mean the phenomenon as such is new. What is new, however, is the fact that such sexual abuse can now be carried out ‘remotely’ with the perpetrator viewing the abuse possibly in a different country than that of the victim.”¹⁶⁵

- **National Center for Missing & Exploited Children (NCMEC)**

NCMEC’s “mission is to help find missing children, reduce child sexual exploitation, and prevent child victimization,” and it serves as the United States’ clearinghouse and reporting center for all issues related to the sexual exploitation and abuse of children.¹⁶⁶ NCMEC holds a fairly unique legal status in that it is a private, nonprofit 501(c)(3) corporation, but it was established by statute. Furthermore, ESPs are mandated under U.S. law to “as soon as reasonably possible after obtaining actual knowledge of any facts or circumstances” from which there is an “apparent violation” of one

¹⁶³ Farid, “An Overview of Perceptual Hashing.”

¹⁶⁴ Interagency Working Group, *Luxembourg Guidelines*, 46–47.

¹⁶⁵ Interagency Working Group, *Luxembourg Guidelines*, 47.

¹⁶⁶ “About Us,” NCMEC, accessed October 12, 2022, <https://www.missingkids.org/footer/about>.

of the enumerated offenses involving CSAM, make a report to NCMEC.¹⁶⁷ NCMEC is then, in turn, mandated to make each report available to law enforcement.¹⁶⁸

Courts have issued mixed opinions on whether to classify NCMEC as a government agent or entity, which can have Fourth Amendment implications.¹⁶⁹ Although courts have uniformly held that ESPs qualify as “private actors” when they search for and detect CSAM on their networks, despite the mandatory reporting requirement, NCMEC falls into a grayer area.¹⁷⁰

• Perceived First Person (PFP) CSAM

In June 2021, the Tech Coalition suggested using the term “perceived first person CSAM,” as opposed to the widely adopted “self-generated CSAM,”¹⁷¹ arguing that “self-generated” may imply agency by the child, thereby inferring blame. For “PFP CSAM,” the Tech Coalition suggests the following definition: “Sexualized visual depictions of a child that are generated without the full knowledge, consent, and participation (for example, coercion, blackmail or grooming) of the child and without the physical presence of an instigator AND/OR that may have been originally voluntarily produced by the minor child, but then is distributed to or shared with others without the child’s full knowledge and consent.”¹⁷² It also noted that consent “should not be limited to legal capacity to consent to sexual activity in any particular jurisdiction, but should reflect the age-appropriated state of mind of the subject and should take into account if the generation was coerced.”¹⁷³

Although both terms are used by experts in the field, the default throughout this report will be perceived first person CSAM to better encapsulate the nuanced view articulated by the Tech Coalition.

• Production

As discussed above in the context of “distribution,” “production” in the context of online child sexual abuse and/or exploitation has a specific meaning. The U.S. Criminal Code defines “producing” as “producing, directing, manufacturing, issuing, publishing, or advertising.”¹⁷⁴ The term is generally

167 18 U.S.C. § 2258A(a). The report must include contact information for the provider, but all other content in the report is left to the sole discretion of the provider. 18 U.S.C. 2258A(b).

168 18 U.S.C. § 2258A(c).

169 See, e.g., *United States v. Ackerman*, 831 F.3d 1292, 1308 (10th Cir. 2016) (holding that NCMEC is a governmental entity or agent) (Gorsuch, J.), but see *United States v. Ackerman*, 804 F. App’x 900, 903 (10th Cir.), cert. denied, 141 S. Ct. 458 (2020) (affirming the district court’s decision to deny the motion to suppress because the court correctly determined that NCMEC searched defendant’s email in good faith, and noting that at the time of the search, no court had yet held that NCMEC was a government actor). See also *United States v. Meals*, 21 F.4th 903, 908 (5th Cir. 2021) (cert. pet. filed May 26, 2022) (“Contrary to [defendant’s] supposition, NCMEC is a private, nonprofit corporation, not a government entity. The government takes no position on this question, and like the district court, we need not do so either. But assuming arguendo that NCMEC is a government agent, NCMEC did not exceed the scope of Facebook’s search by merely reviewing the identical evidence that Facebook reviewed and placed in a cyber tip.”); *United States v. Ringland*, 966 F.3d 731, 737 (8th Cir. 2020), cert. denied, 141 S. Ct. 2797 (2021) (“[W]e need not decide whether NCMEC is a government agency or whether it expanded its search beyond Google’s search.”); *United States v. Powell*, 925 F.3d 1, 5 (1st Cir. 2018) (noting that the parties do not dispute that NCMEC was acting as a governmental entity or agent for all relevant purposes).

170 See, e.g., *United States v. Rosenow*, 33 F.4th 529, 541 (9th Cir. 2022); *United States v. Bebris*, 4 F.4th 551, 562 (7th Cir.), cert. denied, 142 S. Ct. 489 (2021) (affirming district’s finding that Facebook did not act as a government agent in sending a CyberTip to NCMEC upon finding that defendant had sent child pornography on its platform); *United States v. Stevenson*, 727 F.3d 826, 830 (8th Cir. 2013) (“A reporting requirement, standing alone, does not transform an Internet service provider into a government agent whenever it chooses to scan files sent on its network for child pornography.”); *United States v. Cameron*, 699 F.3d 621, 638 (1st Cir. 2012).

171 Self-generated CSAM is defined as “[e]xplicit imagery of a child that appears to have been taken by the child in the image. This imagery can result from both consensual or coercive experiences. Kids often refer to consensual experiences as ‘sexting’ or ‘sharing nudes.’” Thorn, *Self-Generated Child Sexual Abuse Material: Attitudes and Experiences: Complete findings from 2019 qualitative and quantitative research among 9–17 year olds and caregivers*, August 2020, <https://www.thorn.org/resources-and-research/>.

172 Tech Coalition, *Multi-Stakeholder Forum: Charting a Collective Path Forward*, June 2021, www.technologycoalition.org/knowledge-hub/this-is-test-knowledge-3.

173 Tech Coalition, *Multi-Stakeholder Forum: Charting a Collective Path Forward*.

174 18 U.S.C. § 2256(3).

used in this context to refer to the actual recording of the child sexual abuse and/or exploitation as videos or photos.

- **Sexual Exploitation of Children in the Context of Travel and Tourism**

“The term ‘sexual exploitation of children in (the context of) travel and tourism’ refers to sexual exploitation of children that is embedded in a context of travel, tourism, or both. The offence can be committed by either foreign or domestic tourists and travellers and longer-term visitors. ... The term ... is used as an alternative to the broadly used term ‘child sex tourism’. It focuses on the fact that the child is being sexually exploited, and that such exploitation occurs within a specific context.”¹⁷⁵

- **Trafficker**

In contrast to “demand-side offender,” defined above, interviewees recommended that “trafficker” be used to describe an individual who aids, facilitates, and/or coordinates the hands-on abuse or CSAM production, often with commercial motivation. That is not to say that the trafficker may not also sexually abuse the child, but that their role is typically distinct.

¹⁷⁵ Interagency Working Group, *Luxembourg Guidelines*, 55.

APPENDIX 2: RELEVANT U.S. STATUTES

I. CSAM Production and Distribution

18 U.S.C. § 2252. Certain activities relating to material involving the sexual exploitation of minors

- a. Any person who—
 - 1. knowingly transports or ships using any means or facility of interstate or foreign commerce ... by any means including by computer or mails, any visual depiction,¹⁷⁶ if—
 - A. the producing of such visual depiction involves the use of a minor¹⁷⁷ engaging in sexually explicit conduct; and
 - B. such visual depiction is of such conduct;
 - 2. knowingly receives, or distributes, any visual depiction [using any means affecting interstate or foreign commerce], if—
 - A. the producing of such visual depiction involves the use of a minor engaging in sexually explicit conduct; and
 - B. such visual depiction is of such conduct;
 - 3. either—
 - A. [within U.S. jurisdiction], knowingly sells or possesses with intent to sell any visual depiction; or
 - B. knowingly sells or possesses with intent to sell any visual depiction that has been mailed, shipped, or transported [by any means affecting interstate or foreign commerce], if—
 - i. the producing of such visual depiction involves the use of a minor engaging in sexually explicit conduct; and
 - ii. such visual depiction is of such conduct; or
 - 4. either—
 - A. [within U.S. jurisdiction], knowingly possesses, or knowingly accesses with intent to view, [materials] which contain any visual depiction; or
 - B. knowingly possesses, or knowingly accesses with intent to view, [material] which contain any visual depiction that has been mailed, or has been shipped or transported [by any means affecting interstate or foreign commerce], if—

¹⁷⁶ “[V]isual depiction’ includes undeveloped film and videotape, data stored on computer disk or by electronic means which is capable of conversion into a visual image, and data which is capable of conversion into a visual image that has been transmitted by any means, whether or not stored in a permanent format.” 18 U.S.C. § 2256(5).

¹⁷⁷ “[M]inor’ means any person under the age of eighteen years.” 18 U.S.C. § 2256(1).

- i. the producing of such visual depiction involves the use of a minor engaging in sexually explicit conduct; and
 - ii. such visual depiction is of such conduct;
 shall be punished as provided in subsection (b) of this section.
- b. (1) Whoever violates, or attempts or conspires to violate, paragraph (1), (2), or (3) of subsection (a) shall be fined under this title and imprisoned not less than 5 years [and not more than 40 years].
- (2) Whoever violates, or attempts or conspires to violate, paragraph (4) of subsection (a) shall be fined under this title or imprisoned not more than [20 years].

18 U.S.C. § 2252A. Certain activities relating to material constituting or containing child pornography

- a. Any person who—
 - 1. knowingly mails, or transports or ships using any means or facility of interstate or foreign commerce or in or affecting interstate or foreign commerce by any means, including by computer, any child pornography;
 - 2. knowingly receives or distributes—
 - A. any child pornography using any means [affecting interstate or foreign commerce]; or
 - B. any material that contains child pornography using any means [affecting interstate or foreign commerce];
 - 3. knowingly—
 - A. reproduces any child pornography for distribution [by means affecting interstate or foreign commerce]; or
 - B. advertises, promotes, presents, distributes, or solicits through the mails, or using any means [affecting interstate or foreign commerce] any material or purported material in a manner that reflects the belief, or that is intended to cause another to believe, that the material or purported material is, or contains—
 - i. an obscene visual depiction of a minor engaging in sexually explicit conduct; or
 - ii. a visual depiction of an actual minor engaging in sexually explicit conduct;
 - 4. either—
 - A. [within U.S. jurisdiction], knowingly sells or possesses with the intent to sell any child pornography; or
 - B. knowingly sells or possesses with the intent to sell any child pornography that has been mailed, or shipped or transported [in a manner affecting interstate or foreign commerce];

5. either—

- A. [within U.S. jurisdiction], knowingly possesses, or knowingly accesses with intent to view, [any material] that contains an image of child pornography; or
- B. knowingly possesses, or knowingly accesses with intent to view, [any material] that contains an image of child pornography that has been mailed, or shipped or transported [in a manner affecting interstate or foreign commerce]; [or]

...

7. knowingly produces with intent to distribute, or distributes, by any means [affecting interstate or foreign commerce], child pornography that is an adapted or modified depiction of an identifiable minor.
shall be punished as provided in subsection (b).

- b. (1) Whoever violates, or attempts or conspires to violate, paragraph (1), (2), (3), (4) ... of subsection (a) shall be fined under this title and imprisoned not less than 5 years [and not more than 40 years].
- (2) Whoever violates, or attempts or conspires to violate, subsection (a)(5) shall be fined under this title or imprisoned not more than [20 years].
- (3) Whoever violates, or attempts or conspires to violate, subsection (a)(7) shall be fined under this title or imprisoned not more than 15 years, or both.

II. Perceived First Person Material Production and Distribution

18 U.S.C. § 2251. Sexual exploitation of children

- a. Any person who employs, uses, persuades, induces, entices, or coerces any minor to engage in, or who has a minor assist any other person to engage in, or who transports any minor in or affecting interstate or foreign commerce, or in any Territory or Possession of the United States, with the intent that such minor engage in, any sexually explicit conduct for the purpose of producing any visual depiction of such conduct or for the purpose of transmitting a live visual depiction of such conduct, shall be punished as provided under subsection (e), if such person knows or has reason to know that such visual depiction will be transported or transmitted [in a manner affecting interstate or foreign commerce].
- b. Any parent, legal guardian, or person having custody or control¹⁷⁸ of a minor who knowingly permits such minor to engage in, or to assist any other person to engage in, sexually explicit conduct for the purpose of producing any visual depiction of such conduct or for the purpose of transmitting a live visual depiction of such conduct shall be punished as provided under subsection (e) of this section, if such parent, legal guardian, or person knows or has reason to know that such visual depiction will be transported or transmitted [using any means affecting

¹⁷⁸ “[C]ustody or control’ includes temporary supervision over or responsibility for a minor whether legally or illegally obtained.” 18 U.S.C. § 2256(7).

interstate or foreign commerce].

- c. (1) Any person who, in a circumstance described in paragraph (2), employs, uses, persuades, induces, entices, or coerces any minor to engage in, or who has a minor assist any other person to engage in, any sexually explicit conduct outside of the United States, its territories or possessions, for the purpose of producing any visual depiction of such conduct, shall be punished as provided under subsection (e).

(2) The circumstance referred to in paragraph (1) is that—

- A. the person intends such visual depiction to be transported to the United States, ... by any means ...; or
- B. the person transports such visual depiction to the United States, ... by any means

...

- e. Any individual who violates, or attempts or conspires to violate, this section shall be fined under this title and imprisoned not less than 15 years [nor more than life].

18 U.S.C. § 2252A. Certain activities relating to material constituting or containing child pornography

- a. Any person who—

...

- 6. knowingly distributes, offers, sends, or provides to a minor any visual depiction ... where such visual depiction is, or appears to be, of a minor engaging in sexually explicit conduct—

- A. that has been mailed, shipped, or transported [by means affecting interstate or foreign commerce];
- B. that was produced using materials that have been mailed, shipped, or transported [by means affecting interstate or foreign commerce]; or
- C. which distribution, offer, sending, or provision is accomplished using the mails or any means or facility of interstate or foreign commerce, for purposes of inducing or persuading a minor to participate in any activity that is illegal;

...

shall be punished as provided in subsection (b).

- b. (1) Whoever violates, or attempts or conspires to violate, paragraph ... (6) of subsection (a) shall be fined under this title and imprisoned not less than 5 years [and not more than 40 years].

U.S.C. § 2422. Coercion and enticement

...

- b. Whoever, using the mail or any facility or means of interstate or foreign commerce, ... knowingly persuades, induces, entices, or coerces any individual who has not attained the age of 18 years, to engage in prostitution or any sexual activity for which any person can be charged with a criminal offense, or attempts to do so, shall be fined under this title and imprisoned not less than 10 years or for life.

III. Internet-Enabled Domestic Child Sex Trafficking

18 U.S.C. § 1591. Sex trafficking of children or by force, fraud, or coercion

- a. Whoever knowingly—
 - 1. [in or affecting interstate or foreign commerce], recruits, entices, harbors, transports, provides, obtains, advertises, maintains, patronizes, or solicits by any means a person; or
 - 2. benefits, financially or by receiving anything of value, from participation in a venture which has engaged in an act described in violation of paragraph (1), knowing, or, except where the act constituting the violation of paragraph (1) is advertising, in reckless disregard of the fact, ... that the person has not attained the age of 18 years and will be caused to engage in a commercial sex act, shall be punished as provided in subsection (b).
- b. The punishment for an offense under subsection (a) is—
 - 1. ... if the person recruited, enticed, harbored, transported, provided, obtained, advertised, patronized, or solicited had not attained the age of 14 years at the time of such offense, by a fine under this title and imprisonment for any term of years not less than 15 or for life; or
 - 2. if ... the person recruited, enticed, harbored, transported, provided, obtained, advertised, patronized, or solicited had attained the age of 14 years but had not attained the age of 18 years at the time of such offense, by a fine under this title and imprisonment for not less than 10 years or for life.

18 U.S.C. § 2251. Sexual exploitation of children

...

d.

1. Any person who, in a circumstance described in paragraph (2), knowingly makes, prints, or publishes, or causes to be made, printed, or published, any notice or advertisement seeking or offering—
 - A. to receive, exchange, buy, produce, display, distribute, or reproduce, any visual depiction, if the production of such visual depiction involves the use of a minor engaging in sexually explicit conduct and such visual depiction is of such conduct; or
 - B. participation in any act of sexually explicit conduct by or with any minor for the purpose of producing a visual depiction of such conduct; shall be punished as provided under subsection (e).
 2. The circumstance referred to in paragraph (1) is that—
 - A. such person knows or has reason to know that such notice or advertisement will be transported using any means [affecting interstate or foreign commerce]; or
 - B. such notice or advertisement is transported using any means [affecting interstate or foreign commerce].
- e. Any individual who violates, or attempts or conspires to violate, this section shall be fined under this title and imprisoned not less than 15 years [nor more than life].

18 U.S.C. § 2251A. Selling or buying of children

- a. Any parent, legal guardian, or other person having custody or control of a minor who sells or otherwise transfers custody or control of such minor, or offers to sell or otherwise transfer custody of such minor either—
 1. with knowledge that, as a consequence of the sale or transfer, the minor will be portrayed in a visual depiction engaging in, or assisting another person to engage in, sexually explicit conduct; or
 2. with intent to promote either—
 - A. the engaging in of sexually explicit conduct by such minor for the purpose of producing any visual depiction of such conduct; or
 - B. the rendering of assistance by the minor to any other person to engage in sexually explicit conduct for the purpose of producing any visual depiction of such conduct; shall be punished by imprisonment for not less than 30 years or for life and by a fine under this title, if any of the circumstances described in subsection (c) of this section exist.

- b. Whoever purchases or otherwise obtains custody or control of a minor, or offers to purchase or otherwise obtain custody or control of a minor either—
 - 1. with knowledge that, as a consequence of the purchase or obtaining of custody, the minor will be portrayed in a visual depiction engaging in, or assisting another person to engage in, sexually explicit conduct; or
 - 2. with intent to promote either—
 - A. the engaging in of sexually explicit conduct by such minor for the purpose of producing any visual depiction of such conduct; or
 - B. the rendering of assistance by the minor to any other person to engage in sexually explicit conduct for the purpose of producing any visual depiction of such conduct; shall be punished by imprisonment for not less than 30 years or for life and by a fine under this title, if any of the circumstances described in subsection (c) of this section exist.
- c. The circumstances referred to in subsections (a) and (b) are that—
 - 1. in the course of the conduct described in such subsections the minor or the actor traveled in or was transported in or affecting interstate or foreign commerce;
 - 2. any offer described in such subsections was communicated or transported using any means or facility of interstate or foreign commerce or in or affecting interstate or foreign commerce by any means including by computer or mail; or
 - 3. the conduct described in such subsections took place in any territory or possession of the United States.

18 U.S.C. § 2421. Transportation generally

- a. In General.—Whoever knowingly transports any individual in interstate or foreign commerce ... with intent that such individual engage in ... any sexual activity for which any person can be charged with a criminal offense, or attempts to do so, shall be fined under this title or imprisoned not more than 10 years, or both.

18 U.S.C. § 2421A. Promotion or facilitation of prostitution and reckless disregard of sex trafficking

- a. In General.—Whoever, using a facility or means of interstate or foreign commerce or in or affecting interstate or foreign commerce, owns, manages, or operates an interactive computer service ... with the intent to promote or facilitate the prostitution of another person shall be fined under this title, imprisoned for not more than 10 years, or both.
- b. Aggravated Violation.—Whoever, using a facility or means of interstate or foreign commerce or in or affecting interstate or foreign commerce, owns, manages, or operates an interactive computer service ... with the intent to promote or facilitate the prostitution of another person and—

1. promotes or facilitates the prostitution of 5 or more persons; or
2. acts in reckless disregard of the fact that such conduct contributed to sex trafficking, in violation of 1591(a),
shall be fined under this title, imprisoned for not more than 25 years, or both.

18 U.S.C. § 2423. Transportation of minors

- a. Transportation With Intent to Engage in Criminal Sexual Activity.—A person who knowingly transports an individual who has not attained the age of 18 years in interstate or foreign commerce ... with intent that the individual engage in prostitution, or in any sexual activity for which any person can be charged with a criminal offense, shall be fined under this title and imprisoned not less than 10 years or for life.

18 U.S.C. § 2425. Use of interstate facilities to transmit information about a minor

Whoever, using the mail or any facility or means of interstate or foreign commerce, or within the special maritime and territorial jurisdiction of the United States, knowingly initiates the transmission of the name, address, telephone number, social security number, or electronic mail address of another individual, knowing that such other individual has not attained the age of 16 years, with the intent to entice, encourage, offer, or solicit any person to engage in any sexual activity for which any person can be charged with a criminal offense, or attempts to do so, shall be fined under this title, imprisoned not more than 5 years, or both.

IV. Live Online Child Sexual Abuse & Sexual Exploitation of Children in the Context of Travel and Tourism

18 U.S.C. § 2260. Production of sexually explicit depictions of a minor for importation into the United States

- a. Use of Minor.—A person who, outside the United States, employs, uses, persuades, induces, entices, or coerces any minor to engage in, or who has a minor assist any other person to engage in, or who transports any minor with the intent that the minor engage in any sexually explicit conduct for the purpose of producing any visual depiction of such conduct or for the purpose of transmitting a live visual depiction of such conduct, intending that the visual depiction will be imported or transmitted into the United States ... shall be punished as provided in subsection (c).
- b. Use of Visual Depiction.—A person who, outside the United States, knowingly receives, transports, ships, distributes, sells, or possesses with intent to transport, ship, sell, or distribute any visual depiction of a minor engaging in sexually explicit conduct (if the production of the visual depiction involved the use of a minor engaging in sexually explicit conduct), intending that the visual depiction will be imported into the United States ... shall be punished as provided in subsection (c).

c. Penalties.—

1. A person who violates subsection (a) ... shall be [fined under this title and imprisoned not less than 15 years nor more than life].
2. A person who violates subsection (b) ... shall be [fined under this title and imprisoned not less than 5 years nor more than 40 years].

18 U.S.C. § 2423. Transportation of minors

...

- b. Travel With Intent To Engage in Illicit Sexual Conduct.—A person who travels in interstate commerce or travels into the United States, or a United States citizen or an alien admitted for permanent residence in the United States who travels in foreign commerce, with a motivating purpose of engaging in any illicit sexual conduct with another person shall be fined under this title or imprisoned not more than 30 years, or both.
- c. Engaging in Illicit Sexual Conduct in Foreign Places.—Any United States citizen or alien admitted for permanent residence who travels in foreign commerce or resides, either temporarily or permanently, in a foreign country, and engages in any illicit sexual conduct with another person shall be fined under this title or imprisoned not more than 30 years, or both.
- d. Ancillary Offenses.—Whoever, for the purpose of commercial advantage or private financial gain, arranges, induces, procures, or facilitates the travel of a person knowing that such a person is traveling in interstate commerce or foreign commerce with a motivating purpose of engaging in illicit sexual conduct shall be fined under this title, imprisoned not more than 30 years, or both.

APPENDIX 3: SENTENCING FACTORS RECOMMENDED BY U.S. SENTENCING COMMISSION¹⁷⁹

The U.S. Sentencing Commission identified three primary factors as recommended areas of focus in sentencing “non-production child pornography offenders” (e.g., receipt, distribution, or possession).¹⁸⁰ These include:

1. “the **content** of the offender’s [CSAM] collection and nature of the offender’s collecting behavior”;
2. “the offender’s degree of involvement with other offenders, particularly in an internet **community** devoted to [CSAM] and child sexual exploitation”; and
3. “the offender’s engagement in sexually abusive or exploitative **conduct** in addition to the [CSAM] offense, either during the instant offense or in prior history.”

Beyond these factors, the Sentencing Guidelines include enhancements based on aggravating circumstances. Relevant in particular to this report is a two-level increase for “use of a computer”¹⁸¹ and a two-to-five-level increase based on the number of images.¹⁸² In fiscal year 2019, the Sentencing Commission found that over 95% of nonproduction offenders received a sentencing enhancement for use of a computer, and these offenses “involved a median number of 4,265 images, with some offenders possessing and distributing millions of images and videos.”¹⁸³ This further underscores the scope of the problem as well as how CSEA has morphed with technology and the internet.

For people convicted of production offenses, the U.S. Sentencing Commission has identified three separate, albeit similar, factors as relevant to sentencing.¹⁸⁴ These include:

1. **Proximity**—“the physical proximity and relationship between offenders and victims, methods of communication used to induce victims’ participation in the offense, and whether offenders and victims lived in the same household”;
2. **Participation**—“the offender’s level of involvement with victims during the production offense, such as the method used to produce the child pornography, whether the offender engaged in sexual contact with victims, or whether the offender manipulated victims through incapacitation, coercion, enticement, or misrepresentation”; and
3. **Propensity**—“the offender’s level of engagement in child pornography or exploitive conduct in addition to the production offense, such as whether the offender distributed or collected child pornography in addition to the production offense or engaged in unrelated exploitation or physical sexual abuse of a child.”

¹⁷⁹ For the complete reports, which also include data on sentencing patterns for the different types of offenses, see www.ussc.gov/topic/child-pornography.

¹⁸⁰ U.S. Sentencing Commission, *Federal Sentencing of Child Pornography: Non-Production Offenses*, 2 and 28.

¹⁸¹ U.S. Sentencing Commission, *Guidelines Manual*, 2021, § 2G2.2(b)(6).

¹⁸² U.S. Sentencing Commission, *Guidelines Manual*, § 2G2.2(b)(7).

¹⁸³ U.S. Sentencing Commission, *Federal Sentencing of Child Pornography: Non-Production Offenses*, 4.

¹⁸⁴ U.S. Sentencing Commission. *Federal Sentencing of Child Pornography: Production Offenses*.

As with nonproduction offenses, the Sentencing Guidelines also include enhancements for aggravating circumstances in production offenses, including “whether the offense involved the knowing misrepresentation of a participant’s identity or use of a computer to persuade or entice a minor to participate in sexually explicit conduct.”¹⁸⁵ Far more production offenders received this enhancement in fiscal year 2019 (45.7%) than those sentenced in fiscal year 2010 (19.5%), further demonstrating the changes in offending behavior.¹⁸⁶

¹⁸⁵ U.S. Sentencing Commission, *Federal Sentencing of Child Pornography: Production Offenses*, 13. U.S. Sentencing Commission, *Guidelines Manual*, § 2G2.1(b)(6) (A two-level enhancement applies “if, for the purpose of producing sexually explicit material or for the purpose of transmitting such material live, the offense involved (A) the knowing misrepresentation of a participant’s identity to persuade, induce, entice, coerce, or facilitate the travel of, a minor to engage [in] sexually explicit conduct; or (B) the use of a computer or an interactive computer service to (i) persuade, induce, entice, coerce, or facilitate the travel of, a minor to engage in sexually explicit conduct, or to otherwise solicit participation by a minor in such conduct; or (ii) solicit participation with a minor in sexually explicit conduct.”)

¹⁸⁶ U.S. Sentencing Commission, *Federal Sentencing of Child Pornography: Production Offenses*, 20.



TECH, LAW & SECURITY
PROGRAM